

Posters

Researchers' Trends, Intentions, and Awareness, Towards Publications in Open Access	1
Aharony, Noa; Hadad, Shlomit	
An Information-Resilient Big-Data Workbench with PDP-DREAM Software.....	4
Anand, Aniruddh; Taswell, Carl	
Examining Transparency, Due Diligence, and Digital Inclusion as Resilient Values for Provenance Research Work	6
Buchanan, Sarah; Bartley, Jane; Bennett, Danielle; Campbell, Elizabeth; Niu, Jinfang; Reed, Victoria	
How Do Adult Digital Literacy Curricula Address Problematic Information?	9
Bugre, Charles; Wedlake, Stacey	
Game Development as Speculative Design: Teaching Data Science Ethics Using Decentralized Research Groups	12
Byun, Mimi; Evans, Sarah; Herman, Bernease; Aragon, Cecilia; Rojas-Ponce, Diana	
Equipping Students and Beyond with Sound COVID-19 Knowledge to Survive and Thrive Despite the Pandemic	15
Chan, Helen; Chu, Sam; Gala, Bhakti; Islam, Md. Anwarul; Batool, Syeda Hina; Bhardwaj, Raj Kumar; Lamba, Manika; Jayasekara, Prasadi Kanchana	
"Prove It!" A User-centric Client for the Blockchain-Based Research Lifecycle Transparency Framework	18
Chang, Yun-Chi; Kung, Li-Fei; Hsiao, Hsu-Chun; Jeng, Wei	
Identifying Immediate Impact Disciplines and Journals by Combining Three Citation-Based Indicators	21
Chang, Yu-Wei	
A Preliminary Analysis of Geography of Collaboration in Data Papers by S&T Capacity Index	24
Chen, Pei-Ying; Li, Kai; Jiao, Chenyue	
Engagement for Good or Ill: Comparing Characteristics of Co-Creative and Co-Destructive Online Communities	27
Chen, Yimin; Cornwell, Sarah, E.; Delellis, Nicole, S.; Kelly, Dominique; Liu, Yifan; Mayhew, Alex; Rubin, Victoria, L.	
Census Racial Metadata: Categories Used for Different Racial Groups Across Countries and Time	30
Cheng, Siyao; Twidale, Michael	
Reading Between the Lines in Discussing Multicultural Books for Children: Preliminary Results	32
Choi, Yunseon; Drouillard, Colette	
A Plug and Play Approach to Vocabulary Mapping and Document Annotation	35
Clunis, Julaine	
Screen Media Modeling & Mentoring: Parent's Views	38
Cox, Evelyn D.; Abbas, June M.; Munyao, Moses M.; Agosto, Denise E.	
The Incessancy of #gobacktothekitchen and Responses to Normalized Online Misogyny	41
Daly, Diana; Smith, Maddie Raeann; Bao, Duo	
Applying Positive Psychology's Subjective Well-Being to Online Interactions	44
Delellis, Nicole, S.; Kelly, Dominique; Liu, Yifan; Mayhew, Alex; Chen, Yimin; Cornwell, Sarah, E.; Rubin, Victoria, L.	
Towards Data-driven GIM tools: Two Prototypes	48
Dinneen, Jesse David; Donner, Sascha; Bubinger, Helen; Low, Jwen Fai; Krtalić, Maja	

A Survey of Undergraduate Engineering Students' Personal Information Management Practices.....	51
Dodson, Samuel	
Exploring News Tagging on Spanish Digital Press	54
Domínguez-Delgado, Rubén; López-Hernández, María-Ángeles; Muñoz-García, Francisco-Javier	
Transitioning Towards Sustainability: The Information Practices of Sustainable Food Consumers.....	57
Du, Xiaotong	
A Preliminary Analysis of Outreach Librarian's Knowledge and Skills.....	60
Du, Yunfei; Byrne, Sephra	
Who Talked about What Regarding Derek Chauvin's Trial: A Work-In-Progress Analysis.....	63
Duan, Yiran; Hemsley, Jeff; Smith, Alex; Gray, LaVerne; Gandhi, Dhvani	
Implementing Virtual Reality Training for Librarians: Supporting Patrons in Crisis	66
Dumas, Catherine; Williams, Rachel D.; Flanagan, Joanna; Porwol, Lukasz	
Collection Development Policies for Special Collections at Selected Academic Libraries	69
Faulkner, James; Kim, Jeonghyun	
Predicting Search Task Difficulty through a Discrete-Time Action Log Representation on Spectrum Kernel.....	72
Gacitúa, Daniel; González-Ibáñez, Roberto	
COVID-19 Tweet Links: A Preliminary Investigation of Type and Relevance	75
Goh, Dion Hoe-Lian; Lee, Chei Sian; Zheng, Han; Theng, Yin Leng	
Sensitively Describing Objects Made to Represent Humans: A Case Study of Dolls at the Strong National Museum of Play.....	78
Graf, Ann	
Uncover Marginalized Narratives of Japanese American Incarceration: An Annotation Scheme for Natural Language Processing and Data Analytics.....	80
Han, Yingying; Maganti, Roopesh; Chen, Jiangping; Chen, Haihua	
Measuring Author Similarity by Topic Networks.....	83
Hsiao, Tsung-Ming; Chen, Kuang-hua	
An Exploratory Study on Chinese Parents' Concerns about Preschoolers' Use of Digital Devices and Expectations for Digital Literacy Education	86
Huang, Ruhua; Shi, Leyi	
Analysis Of Acknowledgments of Libraries in the Journal Literature Using Machine Learning.....	89
Hubbard, David; Laddusaw, Sierra; Tan, Qiaoyu; Hu, Xia	
Google Scholar Metrics: Assessing LIS Researchers' Scholarship in a Developing Country	92
Islam, Md. Anwarul; Howlader, Ahmadul Islam; Roy, Prodip	
Sustainable Development Goals and Public Libraries: An Exploratory Study in a Developing Country	95
Islam, Md. Anwarul; Sultana, Rafia; Widen, Gunilla	
Student Internships in the Time of Crisis: How IT Companies Cope with New Challenges and Opportunities	98
Jakopiec, Tomislav; Aparac Jelušić, Tatjana; Szombathelyi, Donata	
Children's Interest, Search, and Knowledge: A Pilot Analysis of a STEM Maker Workshop.....	100
Jung, Yong Ju; Liu, Jiqun	

Supporting Prosocial Behaviour in Online Communities through Social Media Affordances	103
Kelly, Dominique; Liu, Yifan; Mayhew, Alex; Chen, Yimin; Cornwell, Sarah E.; Delellis, Nicole S.; Rubin, Victoria L.	
A Preliminary Analysis between Available and Shared Resources for the Lebanese Job Crisis ..	106
Khoury, Christy	
Evaluating Visitor Engagement through Instagram: A Case Study of <i>One with Eternity: Yayoi Kusama in the Hirshhorn Collection</i>.....	109
Kim, Jinyi; Lee, Myeong	
The Transitioning of the Term “Digital Library” in Response to Information Resilience: A Thematic Mapping of Literature From 2000-2021	112
Kumari, Madhuri; Gala, Bhakti	
Differences in User Information Behavior between Official Media and Private Media during the COVID-19 Pandemic.....	115
Lei, Li; Xuyan, Wang; Shujun, Liu; Kun, Huang	
Intertwining Search and Non-Search Activities during Cross-Session Search Tasks	118
Li, Yuan; Capra, Rob	
A Preliminary Study on App-Switching during the Mobile Search	121
Liang, Shaobo; Wei, Ziyi; Zou, Ning	
Nudging Learning Behaviour: A Systematic Review	124
Lim, Kok Kiang; Lee, Chei Sian	
Redeeming by Unlearning: A Critical Discourse Analysis of COVID-19 Vaccine Hesitancy on Facebook	127
Ma, Jinxuan; Bonnici, Laurie	
Revisiting Connotations of Digital Humanists: Exploratory Interviews	130
Ma, Rongqian	
Civic Data Education Series: Developing Civic Data Literacies to Serve Communities	133
Mattern, Eleanor; Rapchak, Marcia	
Envisioning Ethical Mass Influence Systems	136
Mayhew, Alex; Chen, Yimin; Cornwell, Sarah, E.; Delellis, Nicole, S.; Kelly, Dominique; Liu, Yifan; Rubin, Victoria, L.	
A Multi-Methodological Assessment to Further the Role of Alabama’s Public Libraries in Addressing the Opioid Crises	139
Mehra, Bharat; Viehouser, Misha; Jaber, Baheya	
Quantifying Individual Research’s Distance from the Trends based on Dynamic Topic Modeling.....	142
Meng, Jie; Lou, Wen; He, Jianguo	
Professional Development: Perception among School Librarians	144
Munyao, Moses M.	
The Effects of Collaborative Country and Discipline on an International Research Collaboration Indicator: The Case of the United States	147
Nguyen, Ba Xuan	
Is This a Good Idea? A Case Study on Civil Perception of Smart City Projects	150
Nguyen, Bich Chau; Zimmer, Franziska	
Cross-Analysis of Researchers’ Different Shared File Management Activities in Cloud Storage	153
Oh, Kyong Eun	

Impacts of Information Source and Perceived Health on Information Behavior	156
Park, Min Sook; Lee, Hyun Seung	
Seeing Resilience in Students: Information Literacy as a Social Practice for Academic Instruction Librarians	159
Rath, Logan	
Give or Take?: Conceptualizations of Data and Its Creation in Information Science	162
Ridenour, Laura; Moulaison-Sandy, Heather	
A User Study in a Pandemic: Some Lessons Learned	165
Rogers, Kristine M.; Trubey, Emily; Oard, Douglas W.	
Self-disclosure of Mentally Ill individuals in Private Facebook Groups: The Means-End Chain Model	168
Rothschild, Nava; Aharony, Noa	
Data Care as Conversation: A Resource to Support Dialogue, Decision Making, and Design Activities that Respect Newcomer Communities and their Data	171
Shankar, Saguna; Nathan, Lisa P.	
Factors Influencing Criminologists' Open Research Data Practices: Trust, Contract, and Value	174
Shih, Yi-Ru; Yang, Yu-Ju; Jeng, Wei	
Organizing Common Bean (<i>Phaseolus vulgaris</i>) Research: A Model for Information Resources Integration	177
Shobhaneer, Mahbub; Ram, Shri	
Visual Semantics of Memes: (Re)Interpreting Memetic Content and Form for Information Studies	180
Smith, Alexander O.; Tacheva, Jasmina; Hemsley, Jeff	
Influence of Interdisciplinarity on Scientific Impact: Case of Climate Change Field	183
Song, Yijia; Fang, Zhichao; Fu, Hui-zhen	
On the Reliability of Funding Acknowledgements as Research Data: Evidence from Astronomy	186
Stahlman, Gretchen	
United for Justice? A Critical Review of Social Justice Research in Information Science and Technology	189
Tacheva, Zhasmina; Namvarrad, Sepideh	
Towards a Diversity, Equity, and Inclusion (DEI) Competency Framework: A Content Analysis of Visible DEI Efforts on LIS Associations' Websites	192
Tang, Rong; Hu, Zhan; Pallo, Bianca	
Rethinking Algorithmic Fairness in the Context of Information Access	195
Udoh, Emmanuel Sebastian; Yuan, Xiaojun; Rorissa, Abebe	
How Does Awe Fuel Information Seeking? A Mixed-methods, Virtual Reality Study	198
Urban, Alex	
Are Virtual Reference Services Color and Gender Blind? Service Equality Revisited	201
Vladoiu, Megan; Fichman, Pnina	
Making a Cocoon: The Social Factors of Pandemic Misinformation Evaluation.....	204
Wan, Yi; Thompson, Kim M.	
Investigating the Relationship between In-Situ User Expectations and Web Search Behavior	207
Wang, Ben; Liu, Jiqun	

Using Textual Analysis Tools to Identify Trends in Privacy Concern Research	210
Weinberger, Maor; Bouhnik, Dan	
Academic Casualization, Precarity, and Information Practices: A Scoping Review	213
Willson, Rebekah; Stewart-Robertson, Owen; Julien, Heidi; Given, Lisa	
Parasitic Knowledge Infrastructures: Data Reuse by Anthropogenic Climate Change Skeptics...217	
Wofford, Morgan	
Persuasion in Audio-based Social Media: A Clubhouse Interview Study	220
Xiao, Lu; Wang, Yueyue; Wei, Xinyi	
Performance, Usability, and User Experience of Rayyan for Systematic Reviews	223
Yu, Fei; Liu, Chuanrui; Sharmin, Selina	
Perception and Use of COVID Contact Tracing Mobile Applications in New York State (NYS)	225
Yuan, Xiaojun; Bennett Gayle, DeeDee; Dadson, Yvonne; Jung, Ellie	
Creating an Undergraduate Data Science Program in a Disciplinary and Institutional Context in an iSchool.....	228
Zhang, Yin; Zeng, Marcia Lei	
Is it Feasible to Use BP Neural Network for Academic Journal Evaluation: an Attempt Based on 6 Single Objective Weighting Methods	230
Zhao, Rongying; Zhu, Weijie; Chen, Wenxin	
The Influence of Open Access on the Academic Purification Effect of Retracted Publications: A Causal Inference Analysis.....	233
Zheng, Er-Te; Fu, Hui-Zhen	

Researchers' Trends, Intentions, and Awareness, Towards Publications in Open Access

Aharony, Noa
Hadad, Shlomit

Bar-Ilan University, Israel | Noa.Aharony@biu.ac.il
The Open University of Israel, Israel | shlomith@openu.ac.il

ABSTRACT

This study examined Israeli researchers' trends, intentions and awareness regarding scientific publications in open access (OA) journals and repositories. A survey was distributed to 202 Israeli researchers. According to the findings, most researchers used a combination of closed (traditional/toll access) journals with OA gold/green routes. Researchers showed low awareness of OA key concepts, which was significantly associated with the tendency not to publish in OA. In addition, researchers who reported that there was a formal policy concerning publishing articles in high-impact journals in their department tended to publish less in OA. The most common reasons that caused researchers not to publish in OA routes were lack of funding and lack of awareness. Conclusions present a need for an OA's declared policy as well as financial and informational support.

KEYWORDS

Open-Access (OA) publication, green OA, gold OA, blended publishing models

INTRODUCTION

Open Access (OA) is a term that is used to address unrestricted online access to scientific articles, as part of an effort to "open up" scientific outputs to the public. Further, OA may improve the rigor, replicability, and availability of research (Clayson et al., 2021; Piwowar et al., 2018). Countries and research institutions across the globe have committed to providing OA for their research output while developing a clear OA policy (Moskovkin et al., 2021). In addition, various grants require that the research outputs will be available in OA journals (gold OA) and repositories (green OA). Both models have considerable challenges. Gold OA includes high processing charges for OA journals and the risk of encountering or using "predatory" journals (those that are driven by financial interest and do not perform quality examinations). Green OA, includes misunderstandings concerning publishers' copyright policies that prevent researchers from self-archiving their articles (Narayan & Luca, 2017). Due to the lack of OA policy and research regarding OA publication in Israel, the current study sought to examine Israeli researchers' trends, intentions, and awareness, regarding scientific publications in OA journals and repositories.

METHOD

A quantitative research design based on two questionnaires was used (Barrett et al., 2017; Woszczyński & Whitman, 2016). The questionnaire was distributed through researchers' email list on the faculties websites and/or through the faculty secretariat. 202 Israeli researchers from universities, colleges and research institutes in Israel took part in the study. Among them, 106 (52.5%) were from the Social Sciences and the Humanities, and 96 (47.5%) were from the Exact Sciences. The structured questionnaire contained closed end questions regarding OA awareness, engagement and intention to engage with OA, the department / university position, and reasons why they do not publish or intend to publish in OA journals and/or repositories.

RESULTS

Engagement with Open Access Publishing

Researchers were asked to choose the models they used in the last 3 years to disseminate their publication (e.g. article, book, book chapter, conference paper, working paper). Publication models are presented in Table 1.

Publishing model	No of researchers (%)	Description
Blended: Open-Access +Traditional subscription access journals	157 (77.7%)	<i>Blended publishing models</i> (the researcher could choose more than one model, therefore there are more models than the total results count): <ul style="list-style-type: none">○ Published in closed (toll access) journals: 149 researchers (95%)○ OA journal: 81 researchers (52%)○ Discipline or subject repository (e.g. ArXiv, SSRN): 63 (40%)○ Academic social networks (e.g. ResearchGate): 60 (38%)○ Personal/University website: 52 researchers (33%)
Non-OA only (i.e. traditional journal)	35 (17.3%)	
Open-Access (Gold/Green only)	10 (5%)	

Table 1. Researchers' Publishing Models

Intention to Engage with Open Access

Researchers were asked if they intend to publish in OA journals and/or repositories in the coming year. Answers were coded on a three-level ascending order scale as presented in Figure 1. 30.7% of the researchers reported that they do not intend to publish in OA models.

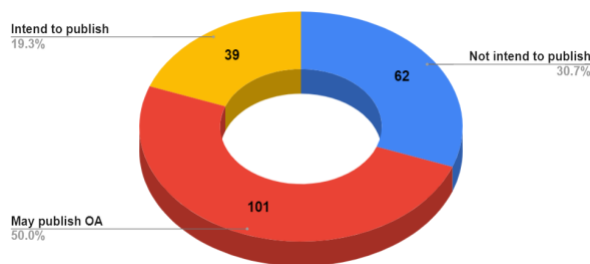


Figure 1. Researchers' Intention to Publish OA

Awareness of Open Access Publishing

Researchers were asked to rate their awareness regarding the following forms of OA publishing: Open Access: Gold model, Open Access: Green model, Creative Commons license, hybrid OA, predatory journals. The mean score (2.24 (n=202, SD=1.05)) on a five-point scale (where 5=Very familiar) shows that respondents had a low average score about OA key concepts, with a midpoint of 3.0. Using Spearman's rank-order correlation test indicated researchers' awareness has a weak, positive, and significant relationship ($r(202) = .184, p = .043$) concerning their intentions to publish in OA journals or repositories.

Publishing an article – Formal/non-formal policy

Researchers were asked if their department had a formal policy concerning publishing articles in journals. Answers were coded on a three-level ascending order scale (1-3) as detailed below:

1. I do not know of any formal or non-formal policy—57 researchers (28%).
2. There is a *non-formal* policy that requires publishing in a high impact factor journal, indexed in *Scopus* or the *Web of Science*, and intending to use a Q1 journal—63 researchers (31%).
3. There is a *formal* policy position that requires publishing in a high impact factor journal, indexed in *Scopus* or in the *Web of Science*, and intend to use a Q1 journal—82 researchers (41%).

A chi-square test of independence showed significant association between intention to publish in OA and a publication policy ($\chi^2(2, N = 202) = 4.1, p = .035$). Those who reported a formal policy expressed a lower intention to publish in OA.

Reasons for not publishing / lack of intention to publish in OA journal/repository

Researchers were also asked to rate the following statements according to their importance, to explain why they have *not* published (35, 17%)/do not intend to publish in OA journals (62, 30.7%). Results range from the most common reason (1) to the least common (4):

1. I was unable to fund/I am not willing to pay an article processing charge (43%).
2. I am not aware of Gold Open Access as a publishing model (28%).
3. I am not familiar enough with quality and peer reviewed OA journals in my field that are required for promotion (17%).
4. I am concerned about “predatory journals” (12%).

Reasons why researchers do not intend to self-archive articles in a OA repository (i.e., Green OA). Results are presented ranging from the most common reason (1) to the least common (3):

1. I am not aware of any repository in my field (47%).
2. I am concerned that I will violate copyright if I put my article in a repository (42%).
3. I do not think it worth the time and effort involved (11%).

CONCLUSION

Based on this study's findings, the overarching recommendation is for the urgent need to regulate policies at the national and institutional levels regarding scientific OA publications. Authors assume that by following a declared policy, researchers' awareness towards OA models would increase. Moreover, as findings indicate, academic researchers need financial and informational support. Therefore, dedicated budgets should be provided to researchers who choose to publish in OA, as well as more information regarding OA models and concerning copyright and appropriate repositories. These facilitating conditions will allow researchers flexibility in choosing a journal in which to publish and encourage them to use OA journals. Further, it may raise their awareness concerning self-archiving repositories. Further studies may crosscheck researchers' perspectives with a bibliometric analysis.

REFERENCES

- Barrett, J., Dalton, M., Greene, J., Harper, C., O'Neill, J. & Schoen, R. (2017). Open access publishing survey. University College Dublin Library. <http://hdl.handle.net/10197/8396>
- Clayson, P. E., Baldwin, S. A., & Larson, M. J. (2021). The open access advantage for studies of human electrophysiology: Impact on citations and altmetrics. *International Journal of Psychophysiology*, 164, 103-111.
- Moskovkin, V. M., Saprykina, T. V., Sadovski, M. V., & Serkina, O. V. (2021). International movement of open access to scientific knowledge: A quantitative analysis of country involvement. *The Journal of Academic Librarianship*, 47(1), 102296. <https://www.sciencedirect.com/science/article/pii/S0099133320301877>
- Narayan, B., & Luca, E. (2017). Issues and challenges in researchers' adoption of open access and institutional repositories: a contextual study of a university repository. *Information Research: An International Electronic Journal*.
- Piowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., & Haustein, S. (2018). The state of OA: A large-scale analysis of the prevalence and impact of open access articles. *Peer Journal*, 6, e4375. https://peerj.com/articles/4375/?utm_source=TrendMD&utm_campaign
- Woszczyński, A. B., & Whitman, M. E. (2016). Perspectives on open access opportunities for is research publication: Potential benefits for researchers, educators, and students. *Journal of Information Systems Education*, 27(4), 259-276.

An Information-Resilient Big-Data Workbench with PDP-DREAM Software

Anand, Aniruddh

Brain Health Alliance, USA | aanand@bhavi.us

Taswell, Carl

Brain Health Alliance, USA | ctaswell@bhavi.us

ABSTRACT

PORTAL-DOORS Project DREAM Software, available as an open-source C#-centric codebase from a Github public repository at PDP-DREAM, implements the PDP-DREAM principles and PDP-FAIR metrics with web-enabled workbench software for distributed data repositories in the Nexus-PORTAL-DOORS-Scribe Cyberinfrastructure. PDP-DREAM Software has been developed for Microsoft platform technologies with ASP.NET Core, SQL Server, and Internet Information Server. As a web-enabled workbench, PDP-DREAM provides many features for big data management with tools and services to support information resilience in defense of truth in science and integrity in research.

KEYWORDS

PORTAL-DOORS Project, NPDS Cyberinfrastructure, PDP-DREAM software

INTRODUCTION

The PORTAL-DOORS Project (PDP) began in 2006 with the mission of developing software to serve as a bridge between the lexical web and the semantic web. The original acronyms PORTAL and DOORS represented respectively the phrases “*Problem Oriented Registry of Tags and Labels*” and “*Domain Ontology Oriented Resource System*”. The 2006 architectural blueprint design for the infrastructure system was conceived with PORTAL as an analogue inspired by the IRIS registry system and DOORS as an analogue inspired by the DNS directory system (Taswell 2007). PDP now maintains this distributed data management system as the Nexus-PORTAL-DOORS-Scribe (NPDS) Cyberinfrastructure for big data and metadata repositories. NPDS can be described as a ‘who what where’ diristry-registry-directory-registrar system with Nexus diristries, PORTAL registries, DOORS directories, and Scribe registrars for identifying, describing, locating, and linking things on the internet, web and grid.

PDP-DREAM SOFTWARE

PDP-DREAM Software for NPDS from PDP has been available since its initial public release as open-source software in 2021 at a Github repository PDP-DREAM (Taswell 2021) with the acronym DREAM for the phrase “*Discoverable Data with Reproducible Results for Equivalent Entities with Accessible Attributes and Manageable Metadata*”. PDP-DREAM Software has been coded with the C# programming language for Microsoft platform technologies with ASP.NET Core, SQL Server, Internet Information Server, and a Visual Studio Solution which provides a collection of Visual Studio Projects that manage data repositories in the NPDS Cyberinfrastructure. Nexus servers built with PDP-DREAM Software provide “anonymous user read access to resource metadata in the Nexus diristries, PORTAL registries, and DOORS directories”—see the website <https://www.portaldoors.org> for “information on the continuing design, development, and implementation of NPDS as a cyberinfrastructure for semantic computing on the internet, web, and grid.” Scribe registrar servers, which require secure https connection for authorized agent write access, provide registrar services for resource metadata and data management at Nexus diristries, PORTAL registries, and DOORS directories. Resource metadata record agents with Author and/or Editor roles may access write privileges at the BHA, GTG, and PDP Scribe Registrars, but not for management of NPDS components where write access is restricted to agents with Admin roles. In an extensible and flexible manner, other independent organizations may instantiate and manage their own problem-oriented domain-specialty NPDS servers to interoperate with those already created by Brain Health Alliance to prototype, develop, and demonstrate the NPDS Cyberinfrastructure distributed network of data repositories.

Branches and Releases

Over the course of the next year, the PDP-DREAM public repository at <https://github.com/BHAVIUS/PDP-DREAM> will follow a software development roadmap with 3 named branches Aoraki, Cervin, and Gangkhar which will be maintained respectively in Microsoft’s current .NET 6, preview .NET 7, and planned .NET 8. Thus, the branches will be named Aoraki-Net6 and Cervin-Net7 with releases in 2022, and Gangkhar-Net8 with releases in 2023. Dated releases will append an 8-digit code for the date of each software release, for example, Cervin-Net7-20221108 anticipated for same date as Microsoft’s release of .NET 7 later this year. All branches and releases will be available from the PDP-DREAM public repository.

Aoraki-Net6

Aoraki will focus on .NET 6 web apps built with ASP.Net Core MVC Views and Telerik UI for ASP.Net Core.

85th Annual Meeting of the Association for Information Science & Technology | Oct. 29 – Nov. 1, 2022 | Pittsburgh, PA. Author(s) retain copyright, but ASIS&T receives an exclusive publication license.

Cervin-Net7

Cervin will focus on .NET 7 web apps built with ASP.Net Core Razor Pages and Telerik UI for ASP.Net Core.

Gangkhar-Net8

Gangkhar will focus on .NET 8 web apps built with ASP.NET Core Blazor and Telerik UI for Blazor.

Relevance to Information Resilience

Software from PDP for the NPDS Cyberinfrastructure has been in continuous development since 2007. Discussions in the early papers from PDP (such as Taswell 2007 and Taswell 2010) addressed concerns about the importance of distributed and democratized information systems in supporting the free flow of information without monopolization by a single person, organization, or government. Providing support with open-source software for technology platforms that help to counter the spread of false, deceptive, manipulative and/or wrongful information has since become increasingly important as a means of fighting back against the worsening information cyberwars that threaten truth and integrity when preserving the published record of information in libraries. Solutions to some of these problems with the tools available in PDP-DREAM Software will be demonstrated at the ASIS&T 2022 Online Workshop entitled “Who Are the Guardians of Truth and Integrity?” (Craig et al 2022).

CONCLUSION

PDP-DREAM Software, available as an open-source C#-centric codebase from a Github public repository at PDP-DREAM, provides a Microsoft-platform implementation of the PDP-DREAM principles, PDP-FAIR metrics, and web-enabled workbench software for data repositories in the NPDS Cyberinfrastructure. Named branches of the codebase respectively for Microsoft .NET 6, 7, and 8 will be maintained with Aoraki-Net6 and Cervin-Net7 available in 2022 and Gangkhar-Net8 available in 2023. PDP-DREAM Software and the NPDS Cyberinfrastructure will continue to serve as a bridge connecting the semantic web with the lexical web while supporting the PDP mission not only to advance information technology, but also to promote truth in science and integrity in research with information resilience.

REFERENCES

- A Craig et al. 2019, "DREAM Principles and FAIR Metrics from the PORTAL-DOORS Project for the Semantic Web," IEEE 11th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), 2019, pp. 1-10, <https://doi.org/10.1109/ECAI46879.2019.9042003>.
- A Craig et al. 2022, “Who are the Guardians of Truth and Integrity”, ASIS&T 2022 Annual Meeting Online Workshop, <https://www.portaldoors.org/NPDS/Site/Papers#Craig2022WAGTI>
- PORTAL-DOORS Project (n.d.). “NPDS Cyberinfrastructure,” Retrieved 5 June 2022 from <https://www.portaldoors.org/NPDS/Site/Info>.
- C Taswell 2007, “DOORS to the semantic web and grid with a PORTAL for biomedical computing,” IEEE Trans Inf Technol Biomed, 12(2):191-204, <https://doi.org/10.1109/TITB.2007.905861>.
- C Taswell 2010, “ A Distributed Infrastructure for Metadata about Metadata: The HDMM Architectural Style and PORTAL-DOORS System”, Future Internet 2(2):156-189, <https://doi.org/10.3390/fi2020156>
- C Taswell 2021, “The NPDS Cyberinfrastructure,” <https://doi.org/10.5281/zenodo.5585404>.
- S Dutta et al. 2019, "Managing Scientific Literature with Software from the PORTAL-DOORS Project," IEEE 15th Intl Conf on eScience (eScience), pp. 588-593, <https://doi.org/10.1109/eScience.2019.00081>.
- S Dutta et al. 2020, "DREAM Principles from the PORTAL-DOORS Project and NPDS Cyberinfrastructure," IEEE 14th Intl Conf on Semantic Computing (ICSC), pp. 211-216, <https://doi.org/10.1109/ICSC.2020.00044>.
- SK Taswell et al. 2020, “The hitchhiker's guide to scholarly research integrity,” Proc Assoc Inf Sci Technol, <https://doi.org/10.1002/pr2.223>.

Examining Transparency, Due Diligence, and Digital Inclusion as Resilient Values for Provenance Research Work

Buchanan, Sarah	University of Missouri, USA buchans@missouri.edu
Bartley, Jane	University of Missouri, USA jebbz6@missouri.edu
Bennett, Danielle	The Menil Collection, USA dbennett@menil.org
Campbell, Elizabeth	University of Denver, USA elizabeth.a.campbell@du.edu
Niu, Jinfang	University of South Florida, USA jinfang@usf.edu
Reed, Victoria	Museum of Fine Arts Boston, USA vreed@mfa.org

ABSTRACT

In response to the rising calls for transparency in displaying collections information, an ongoing research project is developing and studying provenance research methods in four domains, which cross museum and archive settings and their legal oversight in the United States: art collections (furthermore colonial era and Nazi-looted era), special collections and manuscripts, Native and Indigenous research, and local or rural history collections. We illustrate three resilient values underpinning successful or at least informative provenance research: transparency, due diligence, and digital inclusion. Together the values, when practiced further, will guide information professionals towards researching collections provenance more effectively according to internal and public-facing goals.

KEYWORDS

Provenance research; Information management; Ethics of cultural heritage; Public access; Descriptive metadata.

INTRODUCTION: FROM DIVERSE DEFINITIONS TO PERSISTENT VALUES

Provenance research aims to ascertain, through use of primary and secondary sources, the immediate and preceding ownership history for a cultural object since its creation. Documentation of an item's provenance may involve accessing a wide number of far-flung and unexpected source materials: from sales receipts to personal property listings, and from photographic geolocation data to comprehensive lists (catalogue raisonnés) of an artist's body of work. While provenance research in practice has evolved and expanded largely in positive ways over the past twenty years—in part motivated by the adoption of the 1998 Washington Conference Principles on Nazi-Confiscated Art by 44 countries and 13 NGOs (non-governmental organizations) (Campbell, 2019)—it is still characterized by isolated expertise from one domain to another, few freely accessible instructional and training opportunities, and dynamic internal uses and external outcomes of the work.

Provenance research in the archives management field, for example, differs from that in art history and archaeology. Archaeologists often study the provenance of individual objects, whereas archivists seek out the origin and history of groups of records, i.e. series, collections or fonds. Due to the very long temporal distance between (ancient) creation and discovery, the provenance of archaeological objects is usually unknown, whereas the provenance of archival materials is usually known when (museum) archivists acquire records directly from the creators. Therefore unlike archaeologist teams who need to *investigate*, to first discover or infer the provenance of objects, archivists are focusing more on how the provenance information should inform the appraising, organizing, describing and using of archival materials. Archival provenance description includes creator history, records history, and custodial history (Millar, 2002). Some scholars suggest expanding archival provenance further to include the sociopolitical context in which the records were created and evolved (Nordland, 2004; Wurl, 2005; Beattie, 2009). Niu (2013) also suggests technical provenance information be included for digital archives produced in databases, workflows or published on the Web. Technical provenance information refers to technical details of the creation, transformation, derivation and access processes of digital archives—the software programs, hardware, workflow instruments, and information about how web servers retrieve the digital image and deliver it to the browser which renders it.

The remainder of this poster analyzes active professional work vignettes to illustrate three resilient values underpinning successful or at least informative provenance research: transparency (the open sharing of known and missing information), due diligence (the time-based exhaustivity of effort put towards researching), and digital inclusion (the use of accessible resources for sharing expert contributions). The three values emerge from the above evolving definition of provenance research in service to current practices at American museums and archives.

TRANSPARENCY

Archivists and museum curators' long-standing efforts to describe objects accurately have renewed conversations with Native American communities regarding the ethical stewardship of cultural and sacred items. Recognizing that

museums need to build relationships with tribes—beyond the legal provisions of NAGPRA—the Center for Art Collection Ethics (ACE) at the University of Denver (DU), CO, USA has developed a postgraduate certificate program on the ethical stewardship of Native items. It was developed in consultation with members of displaced tribes of Colorado, and with Native DU faculty and staff members. The program highlights object-based research and related archival study, in partnership with three museums, while underscoring the value of transparent, unambiguous face-to-face dialogues. Though safety concerns about the COVID-19 pandemic forced the program cancellation in June 2020, ACE continues to promote ongoing public discussions about stewardship issues and plans to offer the training program when the pandemic eases further and Native representatives, including tribal elders, can be safely welcomed to campus: their ancestral lands. As observed by Denver Art Museum curator Dakota Hoska (Oglala Lakhóta), a collection of Native items, when approached ethically and transparently, can be promoted “as a rebuilding tool; as a source of pride, reference and power” (Pierce, 2021).

DUE DILIGENCE

Since January of this year, the Museum of Fine Arts, Boston, MA, USA (MFA) has announced the restitution, or physical return, of no fewer than four individual objects from its collection: a painting by Salomon van Ruysdael looted from its Jewish owner in Budapest in 1945, two Malian terracottas illegally excavated in the late 1980s, and an ancient Roman head that was pillaged from Italy during the upheaval of World War II. Even if those resolutions were the correct course of action, it is in many ways the worst-case scenario for a museum to have to deaccession, or use acquisition funds to re-purchase, a work of art in light of the fact that it was stolen or smuggled. How did these illicit works of art make their way into the Museum to begin with? No matter what the curatorial department or the precise rationale for restitution, each of the MFA’s past acquisition mistakes can be faulted for insufficient due diligence at the time the object was acquired. Due diligence—in this context, the steps necessary for museums to make responsible and judicious decisions—includes requesting information about the work of art’s ownership, export and importation history, understanding and acknowledging the acquisition’s inherent legal and financial risk, and conducting original research on the object’s provenance. Implementing a system of due diligence therefore requires ongoing education about cultural property law, ethics, and museum policies, access to legal resources, and access to research resources such as libraries, archives, and databases. With such tools, and a method of checks and balances built into the staff organization, collecting institutions will be able to establish consistency and accountability in the acquisition process. Only with greater due diligence can museums reduce their own legal and reputational liability and help diminish the market demand for illicit materials.

DIGITAL INCLUSION

Access to sales catalogs, as well as archives and publications about historical dealers, are substantial resources for provenance researchers. They often can be difficult to find, since many are not cataloged in library systems, or easily accessible. Digitization (see Arizona State Library, Archives and Public Records, 2018: p. 3) provides more ways for such materials to become reachable by provenance scholars, since the resources can dramatically increase knowledge about the works in museum collections. For example, a white-ground oil vessel (alabastron) depicting an archer in the Menil Collection, TX, USA demonstrates how such resources can reveal longer object biographies. The vase was purchased on behalf of the Menil Foundation by Galerie Koller in Zurich, Switzerland in 1979 from the posthumous sale of the collection of Ernest Brummer (The Ernest Brummer Collection, 1979: no. 694). Tracing the vase through catalog sales and digitized dealer archives has identified it as previously being in the collections of the art dealers Josef (1883-1947) and Ernest Brummer (1895-1964) between 1927 and 1979 (Brummer Gallery Records, n.d.: no. N1147). Their records indicated that prior to the vase entering the Brummer Collection, it belonged to Alphonse Kann, another dealer in Paris, France, who had sold it in 1927, and mentioned a previous “Lambros” collection. The Kann sales catalog (Kann and Canessa, 1927: no. 5) designates that the vase came from the collection of Jean P. Lambros (1843-1909), an archaeologist, collector, and antiquities dealer in Athens, Greece whose collection was sold in 1912. Luckily, there is a published catalog of the 1912 sale (Collections de feu..., 1912: no. 39) with an image of the vase. That recently digitized catalog is what allows the institution now to trace the modern biography to the early 20th century and to identify it with a probable provenience of Attica, Greece. As the result of this collection research, the provenance history has greatly expanded, and the vase is now on view for the first time in its history in the Museum. It will be highlighted in both public programs and digital features, including being added to the online collection page with a short public-facing text about its history and iconography.

CONCLUSION

The persistent values of transparency, due diligence, and digital inclusion are at the heart of provenance research, as they should be. This poster provided examples of each of these values in practice and implementation into day-to-day operations—achieving internal goals and increasing accessibility to institutions’ collections. The three values, though powerful on their own, truly reach their full potential when practiced together; explicitly training information professionals about them, accordingly, will facilitate more effective researching of and storytelling with provenance.

ACKNOWLEDGMENTS

Buchanan (as Project Director), Bartley (as Graduate Research Assistant), Bennett, Campbell, Niu, and Reed (as Project Advisors) collaborate on the “Developing Extensible Methods for Provenance Research” project that is generously funded by the Institute of Museum and Library Services (IMLS), grant RE-246339-OLS-20.

REFERENCES

- Arizona State Library, Archives and Public Records (2018). Arizona Tribal Libraries Digital Inclusion Summit 2018. Phoenix: Arizona State Library, Archives and Public Records, Library Development Division. Retrieved from (<https://azmemory.azlibrary.gov/digital/collection/statepubs/id/32418>) and (<https://www.digitalinclusion.org/blog/2017/10/10/arizona-tribal-libraries-digital-inclusion-summit/>).
- Beattie, H. (2009). Where narratives meet: Archival description, provenance, and women’s diaries. *Libraries & the Cultural Record*, 44(1), 82-100.
- Brummer Gallery Records. (n.d.). N1147. Retrieved from (<https://libmma.contentdm.oclc.org/digital/iiif/p16028coll9/60597/full/full/0/default.jpg>).
- Campbell, E. (2019). *The Washington Principles 20 years later: Some progress but not enough*. Retrieved from (<https://liberalarts.du.edu/art-collection-ethics/news-events/all-articles/washington-principles-20-years-later-some-progress-not-enough>).
- Collections de feu M. Jean P. Lambros d’Athene et de M. Giovanni Dattari du Caire. [1912]. Auction 17th-19th June 1912 at Hotel Drouot, Paris. No. 39 (pl. VIII). Retrieved from (<https://archive.org/details/collectionsdefeu00htel/page/8/mode/2up>).
- Kann, Alphonse, and Ercole Canessa. (1927). The Alphonse Kann collection, Sold by his order. Vol. 1, No. 5. New York: American Art Association.
- Millar, L. (2002). The death of the fonds and the resurrection of provenance: Archival context in space and time. *Archivaria*, 53, 1-15.
- Niu, J. (2013). Provenance: Crossing boundaries. *Archives and Manuscripts*, 41(2), 105-115.
- Nordland, L.P. (2004). The concept of “secondary provenance”: Re-interpreting Ac ko mok ki’s map as evolving text. *Archivaria*, 58, 147-159.
- Pierce, C. (2021). Interview with Dakota Hoska (Oglala Lakhóta), Assistant Curator of Native Arts, Denver Art Museum. Retrieved from (<https://liberalarts.du.edu/art-collection-ethics/news-events/all-articles/interview-dakota-hoska-oglala-lakhota-assistant-curator-native-arts-denver-art-museum>).
- The Ernest Brummer Collection. (1979). Auction sale 16th-19th October 1979 at Grand Hotel Dolder, Zurich. Vol. II, No. 694 (pp. 330-331). Zurich: Galerie Koller and Spink & Son.
- Wurl, J. (2005). Ethnicity as provenance: In search of values and principles for documenting the immigrant experience. *Archival Issues*, 29(1), 65-76.

How Do Adult Digital Literacy Curricula Address Problematic Information?

Bugre, Charles

University of Washington, United States | cbugre@uw.edu

Wedlake, Stacey

University of Washington, United States | staceyaw@uw.edu

ABSTRACT

Public libraries, nonprofits, and community organizations are key providers of digital literacy and technology education in their communities. Given that they play a crucial role in helping persons outside formal education to navigate the digital world, these organizations have the potential to be key players in addressing problematic information. These institutions could provide critical support in this area since they teach people to use information and communication technologies and how to find, evaluate, create, and communicate information effectively and efficiently. This poster explored how seven digital literacy curricula for adults address problematic information. We found problematic information related lessons were siloed from other lessons on social media or online searching, and these curricula do not use current best practices for evaluating information but rely on older information literacy models.

KEYWORDS

Digital literacy, problematic information, misinformation, disinformation, information literacy

INTRODUCTION

Previous research has shown that adult education and digital literacy training can help adults learn to navigate problematic information (Seo et al., 2020; Sirlin et al., 2021). Problematic information is an umbrella term that includes information that is “inaccurate, misleading, inappropriately attributed, or all together fabricated” (Jack, 2017). Jack recommends the term “problematic information” as a broader term for concepts like misinformation, disinformation, fake news, information disorder, conspiracy theories or any information that is likely to lead to physical or emotional harm or interfere with the proper functioning of public institutions (2017). However, traditional information literacy lessons typically taught by librarians do not take into consideration that problematic information can exploit personal beliefs and socio-political divides leading to emotional and psychological responses (Sullivan, 2018; Young, 2021). Additionally, these approaches do not sufficiently prepare students to assess information in real, online settings (Caulfield, 2016; Wineburg et al., 2020). Recent research has shown teaching lateral reading, evaluating the credibility of a source by conducting simple searches, can improve information assessment (Pavlounis et al., 2021; Wineburg & McGrew, 2017). Digital literacy curricula cannot treat problematic information assessment as simply an issue of determining truth.

After exiting formal education, adults learn additional digital literacy skills on their own but also from institutions such as public libraries, nonprofits, and community organizations (Rhinesmith & Kennedy, 2020). Some research has explored public libraries’ responsibility and potential to help educate the public on problematic information (Young et al., 2020), but there has been little evaluation of the impact of this programming (Sullivan, 2019). And a gap exists when examining training in other organizations. These digital literacy providers teach people how to use information and communication tools, but the types of skills covered in these programs and the approaches to teaching can widely vary from teaching common workplace software in a classroom setting to ad hoc help during a case management meeting on how to check emails (Dahya et al., 2020; Wedlake et al., 2019). Thus, we need to understand how these providers approach problematic information, if at all. To start our research, we wanted to know, how does current adult digital literacy curricula incorporate teaching about problematic information?

METHODS

In May 2022, we conducted three background interviews with digital literacy instructors to understand their approach to and understanding of problematic information. We then conducted a document analysis (Bowen, 2009) of online digital literacy curriculum resources recommended by the National Digital Inclusion Alliance (NDIA) in their Digital Inclusion Startup Manual (Siefer et al., n.d.). The interviewed instructors also referenced using many of them. These resources are developed by government agencies, nongovernmental organizations, and corporations in the United States for adult digital literacy providers. Although thirteen resources are listed in the manual, we limited our analysis to seven of them that had instructor materials or curricula. We excluded self-paced-only learning resources or pages that hosted mainly external links or resources targeted at youth. For the analysis itself, we looked for whether topics related to problematic information were covered and what techniques were taught. To do this, we located topics by searching the resources’ menus for keywords “information literacy,” “media literacy,” “misinformation,” “disinformation,” and “fake news.” We also searched for the terms “evaluate” and “evaluating”

since these could be common terms used to describe information assessment and examined social media and web search related lessons to look for embedded topics.

RESULTS AND DISCUSSION

Our preliminary analysis found that six out of the seven resources included topics related to problematic information. See Table 1 for the list of digital literacy curricula reviewed and the topic headings where lessons concerning problematic information were found.

Organization	Resource	Problematic information related topics
Goodwill Community Foundation (GCF)	Learn Free Tutorials	Digital Media Literacy, Critical Thinking, Being a Good Digital Citizen
Mozilla Foundation	Web Literacy	Web Detective
TechBoomers	TechBoomers	Fake News Resource list, How to Know if News Website or Story is Credible or Not
Google	Applied Digital Skills	Write Using Online Research
Denver Public Library	Technology Classes & Workshops	How to Spot Fake News
Northstar	Digital Literacy Assessment & curricula	Information Literacy

Table 1. Digital literacy instructor curricula with problematic information topics

Upon initial review, we noted that problematic information topics were not found under any uniform heading, and each site took a different approach to categorizing the topics. All the resources except from Google and Mozilla had clear headings that would help an instructor understand the lesson covers problematic information. However, the variation in headings requires instructor familiarity with topics such as “information literacy” and “digital media literacy.” The types of actual skills covered by the resources varied in complexity. The most common approaches the resources covered included examining the URL for .edu, .gov etc., and looking for the author or source of the material. A few resources went into more in-depth approaches. Denver Public Library’s “How to Spot Fake News” lesson plan does introduce some basic lateral reading methods such as checking other websites to find out author credentials and encourages reflection about “emotional responses” versus “logical analysis of information” (Denver Public Library, n.d.)” GCF’s tutorials also incorporate lessons on personal confirmation bias, echo chambers, critical thinking and logical fallacies, but the skills focus on evaluating the webpage itself such as looking at the about page and determining if the website shows bias (GCFGlobal.org, n.d., n.d., n.d.) instead of using the lateral reading technique of searching of the source on the open web or via Wikipedia (Wineburg et al., 2020).

Our keyword searches and examination of social media and web search lessons found almost no discussion of problematic information or source evaluation within these topics. Even though all the curricula had these lessons, only Google incorporated evaluating information/sources into a non-problematic information specific lesson about writing a research paper (*Lesson Plan*, n.d.). However, other Google lessons related to web search only noted the existence of ads and not further source evaluation (*Google Search for Beginners*, n.d.). GCF web search lessons covered using advanced search features and issues around copyright, but source evaluation was found in a different, disconnected lesson. Many social media lessons did cover issues around the learner’s digital footprint, urging the user to use discretion about posting personal information but left out how to navigate the information a person may see in social media (Northstar Digital Literacy, n.d.).

CONCLUSION

In summary, this poster shows that the analyzed digital literacy instructor resources and curricula do have some lessons about problematic information but do not align with recent best practices for assessing problematic information (Pavlounis et al., 2021). For example, lateral reading, an emerging best practice that emphasizes evaluating credibility by conducting simple web searches does not appear whereas, vertical reading, evaluating the source by its own website through the URL and “About” page appears often. These methods and approaches are not integrated into topics such as online search or social media. Instead, instructors must specifically seek out these lessons. Additionally, most of the skills that are taught tend to focus on information found while conducting online searches and not how to handle the deluge of information seen on social media. Of course, what these resources cover and do not cover may not reflect what instructors teach. Future research could explore how instructors address navigating problematic information in real-life settings and work with instructors and communities to co-design lessons that better address problematic information digital literacy needs for adults.

REFERENCES

- Bowen, G. A. (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal*, 9(2), 27–40. <https://doi.org/10.3316/QRJ0902027>
- Caulfield, M. (2016, December 19). Yes, Digital Literacy. But Which One? *Hapgood*. <https://hapgood.us/2016/12/19/yes-digital-literacy-but-which-one/>
- Dahya, N., Garrido, M., Yefimova, K., & Wedlake, S. (2020). *Technology access & education for refugee women in Seattle & King County* [Technical Report]. Technology & Social Change Group. <https://digital.lib.washington.edu:443/researchworks/handle/1773/46221>
- Denver Public Library. (n.d.). *How to Spot Fake News*. Denver Public Library. Retrieved June 6, 2022, from <https://www.denverlibrary.org/ctc/how-spot-fake-news>
- GCFGlobal.org. (n.d.-a). *Communication Skills: Being a Good Digital Citizen*. GCFGlobal.Org. Retrieved June 6, 2022, from <https://edu.gcfglobal.org/en/communicationskills/being-a-good-digital-citizen/1/>
- GCFGlobal.org. (n.d.-b). *Critical Thinking and Decision-Making*. GCFGlobal.Org. Retrieved June 6, 2022, from <https://edu.gcfglobal.org/en/problem-solving-and-decision-making/>
- GCFGlobal.org. (n.d.-c). *Digital Media Literacy Tutorial*. GCFGlobal.Org. Retrieved June 6, 2022, from <https://edu.gcfglobal.org/en/digital-media-literacy/>
- Google Search for Beginners*. (n.d.). Applied Digital Skills. Retrieved June 8, 2022, from <https://applieddigitalskills.withgoogle.com/c/college-and-continuing-education/en/google-search-for-beginners/overview.html>
- Jack, C. (2017). *Lexicon of Lies: Terms for Problematic Information* (p. 22). Data & Society. https://datasociety.net/wp-content/uploads/2017/08/DataAndSociety_LexiconofLies.pdf
- Lesson Plan: Write Using Online Research*. (n.d.). Applied Digital Skills. Retrieved June 8, 2022, from https://docs.google.com/document/d/129CfX5eQpjn4PXkktv1QTjtT6MedwGkWD4tx3J13_SA/edit?usp=sharing&usp=embed_facebook
- Northstar Digital Literacy. (n.d.). *Social Media Standards*. Retrieved June 8, 2022, from <https://www.digitalliteracyassessment.org/features>
- Pavlounis, D., Johnston, J., Brodsky, J., & Brooks, P. (2021). *The Digital Media Literacy Gap: How to build widespread resilience to false and misleading information using evidence-based classroom tools*. CIVIX Canada. <https://ctrl-f.ca/en/wp-content/uploads/2022/01/The-Digital-Media-Literacy-Gap.pdf>
- Rhinesmith, C., & Kennedy, S. (2020, November 18). *The Impacts of COVID-19 on Digital Equity Ecosystems*. Benton Foundation. <https://www.benton.org/blog/impacts-covid-19-digital-equity-ecosystems>
- Seo, H., Blomberg, M., & Vu, H. T. (2020). Vulnerable populations and misinformation: A mixed-methods approach to underserved older adults' online information assessment. *New Media & Society*, 23(7), 2012–2033.
- Siefer, A., Callahan, B., & Balboa, P. (n.d.). *Chapter 4: Digital Literacy Training*. The Digital Inclusion Startup Manual. Retrieved June 3, 2022, from <https://startup.digitalinclusion.org/ch4.html>
- Sirlin, N., Epstein, Z., Arechar, A. A., & Rand, D. G. (2021). Digital literacy is associated with more discerning accuracy judgments but not sharing intentions. *Harvard Kennedy School Misinformation Review*. <https://doi.org/10.37016/mr-2020-83>
- Sullivan, M. C. (2018). Why librarians can't fight fake news: *Journal of Librarianship and Information Science*. <https://doi.org/10.1177/0961000618764258>
- Sullivan, M. C. (2019). Leveraging library trust to combat misinformation on social media. *Library & Information Science Research*, 41(1), 2–10. <https://doi.org/10.1016/j.lisr.2019.02.004>
- Wedlake, S., Lothian, K., Keyes, D., & Coward, C. (2019). *Digital Skill Sets for Diverse Users: A Comparison Framework for Curriculum and Competencies*. Technology & Social Change Group, University of Washington Information School. <https://digital.lib.washington.edu/researchworks/bitstream/handle/1773/43467/digital%20skills%20recommendations.pdf?sequence=1&isAllowed=y>
- Wineburg, S., Breakstone, J., Ziv, N., & Smith, M. (2020). *Educating for Misunderstanding: How Approaches to Teaching Digital Literacy Make Students Susceptible to Scammers, Rogues, Bad Actors, and Hate Mongers*. Stanford University.
- Wineburg, S., & McGrew, S. (2017). *Lateral Reading: Reading Less and Learning More When Evaluating Digital Information* (SSRN Scholarly Paper ID 3048994). Social Science Research Network. <https://papers.ssrn.com/abstract=3048994>
- Young, J. C. (2021). Disinformation as the weaponization of cruel optimism: A critical intervention in misinformation studies. *Emotion, Space and Society*, 38, 100757. <https://doi.org/10.1016/j.emospa.2020.100757>
- Young, J. C., Boyd, B., Yefimova, K., Wedlake, S., Coward, C., & Hapel, R. (2020). The role of libraries in misinformation programming: A research agenda. *Journal of Librarianship and Information Science*, 0961000620966650. <https://doi.org/10.1177/0961000620966650>

Game Development as Speculative Design: Teaching Data Science Ethics Using Decentralized Research Groups

Byun, Mimi	University of North Texas, USA mimibyun@my.unt.edu
Evans, Sarah	University of North Texas, USA sarah.evans@unt.edu
Herman, Bernease	University of Washington, USA bernease@uw.edu
Aragon, Cecilia	University of Washington, USA aragon@uw.edu
Rojas-Ponce, Diana	University of North Texas, USA dianarojas-ponce@my.unt.edu

ABSTRACT

This exploratory study investigates the use of game development as a speculative activity to teach data science ethics incorporating the Directed Research Groups (DRG) format, that decentralizes classroom dynamics, emulates real-life working environments, and offers students creative choices driven by their own interests. This DRG focuses on creating a video game addressing an ethical issue in data science. Working in groups of peers from diverse backgrounds and education allows for multiple meanings to occur in a team-based learning environment. This enriches the students' engagement with the material and by creating video games that have an impact on these social issues, it also grounds their role as an active change agent for a future they envision.

KEYWORDS

Speculative education; data science ethics; video game development; project-based learning; higher education

INTRODUCTION

With the modern prevalence of data and the insights extracted, it is often overlooked that data science as a field is frequently layered on top of biased structures. A more nuanced understanding of this and its human impact is needed in computer science and data science education beyond technical skills (Aragon et al., 2022; Herman et al., 2020). Our project consists of University of Washington (UW) and University of North Texas (UNT) students working together in groups to develop video games that address ethical issues in data science. A decentralized model of education called a directed research group (DRG) is used. Finding solutions to complex problems involves applied knowledge and collaborating with colleagues spanning a variety of disciplines.

For this poster, we address the design of this course as a model of student engagement in the emerging field of human-centered data science and look at game development as a speculative learning process, imagining ways to investigate ethical issues. The research questions we seek to answer are:

- How does the directed research group model facilitate deeper student engagement with class content?
- What role does game development as a speculative exercise play in student engagement with class content?

BACKGROUND

The major components of the course are summarized below.

DRG Format

Ethical Games was developed to teach data science ethics in the DRG format to undergraduate and graduate students typically underrepresented in the field of data science. The DRG is a unique education model established at the University of Washington to reconfigure traditional roles in education through the formation of a shared community of practice that centers around scholarly research (Turns & Ramey, 2006; Larson et al., 2009). It addresses the urgent need for undergraduate research opportunities that contribute new intellectual knowledge into their field, a standard of education set by the Boyer Commission on Educating Undergraduates, the Council on Undergraduate Research, and the National Conference for Undergraduate Research (Hu et al., 2008). The format is similar to project-based learning (PBL) pedagogy, which takes the emphasis off of lectures and content transmission onto the interests and motivation of students as they navigate through self-selected projects. Applying knowledge to solve real-world problems enlists deeper learning compared to receiving information passively (Miller & Krajcik, 2019). While already understood and implemented in K-12 educational settings, PBL is not as researched, applied or supported in higher education, and especially not as a means to address diversity issues.

UNT, a Hispanic-serving institution seeking to solidify its role as a research university, partnered with UW in order to replicate this model. Starting in January 2022, students participated in multiple levels of engagement: a) individual assignments and readings, b) meeting with a group of peers to work on the group project of developing a video game addressing an ethical issue in data science or AI/machine learning and c) a weekly class time, where

students, teachers, and student researchers meet together over Zoom and offered comments and feedback through a shared Google Docs. The working groups provided built-in support from peers whose experience and education levels varied. More importantly, they served as the setting for group learning approach. Because students came from multiple disciplines in addition to spanning two universities, they worked through cultural and epistemological impasse(s) in order to produce a viable prototype of a videogame. Working through any impasse or disorientation is another vital component of deep learning that occurs in PBL (Webster et al., 2022). Throughout the course of group meetings, students could decide whether they wanted to pursue their individual project for game development or to work as a team developing one of their projects together. They were given the freedom to change to different groups and develop multiple projects until a decision had to be made to go forward. It was important to initially combine UW and UNT students because several of the UW students have participated in a DRG previously and understood the dynamics of working in self-directed groups.

It was important to provide students with opportunities for further work and research. This decentralized power in the classroom helps to develop student autonomy. Students chose to be in the class through a screening process and if they wanted to receive credit, they could choose the number of credit hours they wanted to register for. It was important to set the classroom culture to establish there was no set class hierarchy and that the students were expected to take ownership of the course and initiate any change. One student created a Slack channel for participants when they felt there was a lack of communication between students and another created an attendance spreadsheet to make the class sign in process more efficient. The second feature is the student as colleague model, which is most evident during the research phase of the course that we are currently in. Students work alongside the instructors and researchers to co-develop the hypothesis, research design and experimental procedure, so that they take a significant role in every phase of research (Hu et al., 2008). They were also encouraged and given opportunities to take the lead in research based on their interest and comfort level. Ultimately, this novel course modality and DRG format takes advantage of the shift toward online learning with its limitations and opens the door for cross-discipline and multiple-university collaboration for research, equipping the students with the ability to adapt and flourish in an ever-changing environment, a crucial skill in both academia and industry.

Teaching Using Game Development

Game development as a mode of learning blends well with the DRG format because it builds upon the students' already high motivation when they are able to decide on the ethical topic to address and the narrative elements of the game. Kafai notes that the greatest learning benefit is from the process of designing instructional games, not playing instructional games (2006). By changing the role of the student from consumer to producer (Kafai, 2006), it forces a series of decisions related to game mechanics and a deeper understanding of the ethical topic in order to produce optimal outcomes in players, such as a more critical attitude in using AI in mortgage loan approvals. Working through these ideas as a group also gives opportunities for collective meaning-making as each student contributes their own ideas and opinions. These activities reinforce the constructivist model of learning Piaget proposed where knowledge is determined by the act of "creation of novelty" and knowledge is determined by students and their creativity in learning about the ethical issues through game development (Mozelius, et al., 2013).

Game Development as Speculative Design

Higher education is ripe for change as it seeks to justify its importance in modern civilization. Staley takes this call to consider multiple possibilities in reimagining the purpose of the university through speculative design and how it might take on future forms as, "feasible utopias (p. 14, 2019)". Game development and the DRG model both seek to "unsettle the present (Staley, 2019)" as the first step of change. It invites us into the process of reimagining relationships between teaching and learning, and questions foundational concepts taken for granted, including systems of bias represented in both data and education (Bang et al., 2013). While the initial course consisted of lectures with case studies of the harms AI can cause, game development provides learning moments for students by allowing them to collectively address social issues such as digital privacy or misinformation in order to create a more just future. As Garcia et al. (2020) indicates, a speculative approach does more than identify and critique inequalities, but "invites in collective imagining and action (p. 21)".

METHODOLOGY

The instructors and student researchers are currently gathering research data from students using a qualitative approach to pre-course reflections, post-course reflections, and periodic surveys. A content analysis of the interviews will also be performed to find emerging themes and insights about the learning process. We will also look at the artifacts produced by the groups to take a deeper look at the level of engagement from students.

CONCLUSION

Preliminary results consistently indicate that students had a meaningful experience working in groups. We are currently in the second stage, where the students can participate in research, and there are several journal articles and conference presentations being developed with students. We intend to pursue funding to have the students design research around the student-developed games being played in the community they were intended for.

REFERENCES

- Aragon, C., Guha, S., Kogan, M., Muller, M. & Neff, G. (2022). *Human-Centered Data Science: An Introduction*. The MIT Press.
- Bang, M., Warren, B., Rosebery, A. S., & Medin, D. (2013). Desettling expectations in science education. *Human Development*, 55(5–6), 302–318. <https://doi.org/10.1159/000345322>
- Garcia, A., Mirra, N., & Teacher Community, the D. D. D. (3d). (2022). Futures bound: Re-designing literacy research as a conduit for healing and civic dreaming. *International Studies in Sociology of Education*, 31(1–2), 5–26. <https://doi.org/10.1080/09620214.2021.1945481>
- Herman, B., Aragon, C., Evans, S., & Shanley, L. (2020). Advancing diversity in human centered data science education through games. *Proceedings of the Association for Information Science and Technology*, 57(1), e314. <https://doi.org/10.1002/pra2.314>
- Hu, S., Scheuch, K., Schwartz, R. A., Gayles, J. G., & Li, S. (2008) Reinventing undergraduate education: Engaging college students in research and creative activities. ASHE Higher Education Report, 33(4). <https://doi.org/10.1002/aehe.3304>
- Kafai, Y. B. (2006). Playing and Making Games for Learning: Instructionist and Constructionist Perspectives for Game Studies. *Games and Culture*, 1(1), 36–40. <https://doi.org/10.1177/1555412005281767>
- Larson, J., Birge, C., Huang, Y.-M., Sattler, B., Turns, J., & Yellin, J. M. H. (2009). Directed Research Groups as a Means of Training Students to Become Technical Communication Researchers. *Technical Communication*, 56(2), 172–177.
- Miller, E. C., & Krajcik, J. S. (2019). Promoting deep learning through project-based learning: A design problem. *Disciplinary and Interdisciplinary Science Education Research*, 1(1), 7. <https://doi.org/10.1186/s43031-019-0009-6>
- Mozelius, P., Shabalina, O., Malliarakis, C., Tomos, F., Miller, C., & Turner, D. (2013). Let the Students Construct Their own fun And Knowledge—Learning to Program by Building Computer Games. *Proceedings of the European Conference on Games Based Learning*, 1, 418–426.
- Staley, D. J. (2019). *Alternative Universities: Speculative Design for Innovation in Higher Education*. Johns Hopkins University Press.
- Turns, J., & Ramey, J. (2006). Active and Collaborative Learning in the Practice of Research: Credit-based Directed Research Groups. *Technical Communication*, 53(3), 296–307.
- Webster, A., Metcalf, A., Kelly, L., Bisesi, A., Marnik-Said, M., Colbeck, C., Marine, R., Vinces, M., Campbell, A., & Allen, T. (2022). Undergraduates' lived experience of project-/problem-based learning in introductory biology. *Advances in Physiology Education*, 46(1), 162–178. <https://doi.org/10.1152/advan.00042.2021>

Equipping Students and Beyond with Sound COVID-19 Knowledge to Survive and Thrive Despite the Pandemic

Chan, Helen

University of Hong Kong, Hong Kong | helenc98_1998@yahoo.com

Chu, Sam

University of Hong Kong, Hong Kong | samchu@hku.hk

Gala, Bhakti

Central University of Gujarat, India | bhakti.gala@cug.ac.in

Islam, Md. Anwarul

University of Dhaka, Bangladesh | anwar@du.ac.bd

Batool, Syeda Hina

University of the Punjab, Pakistan | hina.im@pu.edu.pk

Bhardwaj, Raj Kumar

St. Stephen's College Delhi, India | raajchd@gmail.com

Lamba, Manika

University of Delhi, India | lambamanika07@gmail.com

Jayasekara, Prasadi Kanchana

University of Ruhuna, Sri Lanka | kanchana@lib.ruh.ac.lk

ABSTRACT

The COVID-19 pandemic necessitated the understanding of the infodemic for making informed decisions. Limiting the spread of health misinformation and disinformation was the primary goal of the health informatics project. The project became the recipient of the ASIS&T "Chapter Innovation of the Year Award 2021". A repertoire of online and offline initiatives was carried out with 9 well-researched videos for promoting health informatics. Since August 2021, thousands of academics, librarians, teachers, parents, and students from 16 countries and regions were invited to be science communicators to create and disseminate accurate health information in their areas through an international digital story writing competition. In this paper, we will discuss the strategies of responding to the information crisis, including employing interventions that protect against the infodemic and mitigate its harmful effects, to strengthen the resilience of individuals and communities in dealing with it in an information-resilient society.

KEYWORDS

Health informatics, COVID-19 infodemic, misinformation, disinformation, ASIS&T South Asia, Health Information Literacy

INTRODUCTION

Since the outbreak of COVID-19 in December 2019, the challenge of an overabundance of misinformation, and disinformation about the pandemic has heightened the importance of health information literacy. Health information literacy deals with the "ability of individuals to obtain, process, and understand basic health information and services needed to make appropriate health decisions (MLA, 2011). The authors expand the scope to include health informatics as an important component of the current study. Tedros Adhanom Ghebreyesus, the Director-General of the World Health Organisation (WHO) pointed out at the 2020 Munich Security Conference that "We're not just fighting an epidemic, we're fighting an infodemic" (Editorial, The LANCET Infectious Disease, 2020). Based on WHO's Public Health Research Agenda for Managing Infodemic as a reference, we seek the strategies for responding and deploying interventions that protect against the infodemic and mitigate its harmful effects, evaluating infodemic interventions, and strengthening the resilience of individuals and communities to infodemic, and promoting the development, adaptation, and application of tools for managing infodemic. The urgency of the action is especially important to the Asian countries with a population equivalent to 59.76% of the world's total population (Worldometer, 2022). Core members from 5 Asian countries/economies started the "Dealing with COVID-19 and saving people's lives in South Asia (SA) areas & beyond - Health Informatics Promotion Project" project in June 2021.

PROJECT GOALS AND METHODOLOGY

The Health Informatics Promotion Project was implemented under the collaborative effort of the core team members from Bangladesh, Hong Kong (China), India, Pakistan, and Sri Lanka. It was operated with funding from the Knowledge Exchange Project Fund of The University of Hong Kong and the ASIS&T Special Project Fund for the year 2021-2022. The major project goals included, (1). Spreading awareness about COVID-19 issues and behavior (2). Sharing the experience of countries with experience of dealing with similar viruses; (3). Helping to inform people by identifying pertinent topics to deal with COVID-19 and highlighting the importance of reliable and accurate information; (4). Identifying and sharing accurate and reliable information from authoritative information to fight misinformation and disinformation regarding COVID-19 and (5). Enhancing the accessibility of health information. The impact of the project was measured by collecting quantitative and qualitative data from the mixed methods approach. The survey covers people in rural areas and the underprivileged. Different exciting accomplishments have been achieved towards the goals of the Project by the team, such as the ASIS&T Chapter Innovation Award for the South Asia Chapter from ASIS&T after running the project for 4 months in 2021,

individual core members received the Letter of Recognition from the Executive Director of ASIS&T, Lydia Middleton, complement from participated academics, parents and students from different countries.

Usage of YouTube for COVID Health Information Advocacy

9 well-researched videos in five languages, including English, Bengali, Hindi, Urdu, and Sinhala, made up the total of 45 videos, which were ultimately uploaded to the 'ASIS&T South Asia' YouTube Channel (<https://www.youtube.com/channel/UC-pt07aaW07tWg7InS2dSDg/videos>). The videos' content was prepared using trustworthy sources, such as WHO, the Center for Disease Control and Prevention, authoritative websites from each country, and research-based journal articles.

The intervention design for actions from local to international Levels

To disseminate the videos to people in Asia and beyond, a repertoire of online and offline initiatives was carried out with the videos as the intervention design for action from local to international levels for mitigating the infodemic from June 2021 to May 15, 2022. To disseminate the video contains a series of initiatives such as the 'International Digital Story Writing Competition', 'Compete with Socrates', and 'Covid-19 Knowledge Challenge' was planned and executed.

International Digital Story Writing Competition

Based on the experiences of holding local and Asian-wide Digital Story Writing Competition in Hong Kong (China), the International Digital Story Writing Competition was launched in October 2021. Under the broader theme, "AI, Robot and Health" of the competition, students could write fiction or nonfiction specifying the use of technology in handling different health issues in the pandemic. Under the narrower theme, "COVID-19: Imagine, Create & Share", students could focus on one recent health issue that is affecting the entire world, including stopping the spread of health misinformation and disinformation. Ten free online workshops were held from Jan 1, 2022, to Mar 5, 2022, to prepare students for the Competition. Judges from 7 countries joined the International Digital Story Writing Competition judges' team. 1058 primary and secondary school students from 16 countries/regions registered for this exciting event. Lots of positive responses and compliments from school educators, parents, and students were received starting from the event kick-off. In total, 63 eBook submissions were received from the primary and secondary school students of 9 countries/regions, including Hong Kong (China), India, Qatar, Pakistan, Romania, Sri Lanka, Singapore, Malaysia, and Nepal.

Compete With Socrates

To enhance the critical thinking skills of students in combating the COVID-19 infodemic, the questioning activity "Compete with Socrates" was introduced to all International Digital Story Writing registrants after the seventh workshop on How to Learn the Art of Creating Questions. Four Processes of Comprehension were introduced in the workshop: (1) Retrieving explicitly stated information; (2) Inferences; (3) Interpret and integrating information; (4) Examine and evaluate information (Mullis & Martin (Eds.), 2015). To join this "really difficult challenge, students are requested to get the "secret link" to "unlock" the folder of the 9 videos about dealing with COVID-19 and create one question for each Process of Comprehension.

COVID-19 Knowledge Challenge

To help people stay safe and deal with COVID-19 with sound knowledge and inform people of the scientific facts and strategies of handling pandemics and misinformation from trustworthy sources in addition to accelerating their healthy information literacy, messages of "COVID-19 Knowledge Challenge!" have been widely disseminated to people of different areas by the project team starting from February 2022. After watching each of the 9 informative videos on combating COVID-19 and the infodemic, participants will ask to answer 3 questions. Those who pass this challenge and get at least 80% correct answers will get an electronic certificate to acknowledge their excellent knowledge in dealing with COVID-19. 98 children and adults, including Kindergartens, submitted their answers after one month.

A Series of Cross-boundary Webinars

To disseminate accurate COVID-19 knowledge, webinars with related topics were held in India, Bangladesh, Sri Lanka, Nepal, and Pakistan in November 2021. Parents, schoolteachers, and librarians of different areas benefited from professionals' sharing in the webinars.

CONCLUSION

Information resilience is a precondition to building a resilient community (Chaudhuri, 2022). In this paper, the strategies of responding to the information crisis, and employing interventions that protect against the infodemic and mitigate its harmful effects from local to international levels, were discussed to strengthen the resilience of individuals and communities to the infodemic for an information resilient Society. The provision of the right information at the right time in the right format with the right activities can facilitate individuals, including people of rural areas and the underprivileged, to take the right actions to protect themselves, their families, and communities against the COVID-19 pandemic. Right actions for handling crisis, transition, and resilience for an information-resilient society are not limited to researchers, scientists, educators, and politicians. We're all in this together.

REFERENCES

- Chaudhuri, Sabuj Kumar (2022). Information Resilience as a Precondition to Build a Resilient Community: A New Research Paradigm [Recording]. Retrieved from <https://www.asist.org/information-resilience-a-precondition-to-build-a-resilient-community-a-new-research-paradigm/>
- Editorial, The LANCET Infectious Disease (2020). "The COVID-19 infodemic". *The LANCET Discovery Science*. Retrieved from <https://www.thelancet.com/action/showPdf?pii=S1473-3099%2820%2930565-X>
- Medical Library Association (2011). What is health information literacy? Retrieved from <https://web.archive.org/web/20150906122613/https://www.mlanet.org/resources/healthlit/define.html>
- Hucíková, Anežka & Babic, Ankica (2017). "Medical Informatics Idle YouTube Potential". *National Library of Medicine: National Centre for Biotechnology Information*. Retrieved from <https://pubmed.ncbi.nlm.nih.gov/28679905/>
- McLachlan, Stacey (2022). "YouTube Stats That Matter to Marketers in 2022". *Hootsuite*. Retrieved from <https://blog.hootsuite.com/youtube-stats-marketers/>
- Michigan Engineering, University of Michigan (2022). The Six Type of Socrate Questions. Retrieved from <http://problemsolving.engin.umich.edu/strategy/cthinking.htm>
- Mullis, Ina V.S., & Martin, Michael O. (Eds.). (2015). *PIRLS 2016 Assessment Framework (2nd ed.)*. Chestnut Hill, MA, USA: TIMSS & PIRLS International Study Center.
- Niemiec, Emilia (2020). "COVID-19 and Misinformation: Is Censorship of Social Media a Remedy to the Spread of Medical Misinformation?" *EMBO Reports*, 21.11: E51420.
- Wikipedia (2022). *Socrates*. Retrieved from <https://en.wikipedia.org/wiki/Socrates>.
- World Health Organization (2022). Coronavirus disease (COVID-19) pandemic. Retrieved from https://www.who.int/emergencies/diseases/novel-coronavirus-2019?adgroupsurvey={adgroupsurvey}&gclid=CjwKCAjw3cSSBhBGEiwAVII0Z8cN208atggBDIdE2BouNa4gAo4In5JucjEaelfWImtdBnQnIIPWFxoCjeQQA_vD_BwE
- Worldometer (2022). *Asia Population*. Retrieved from <https://www.worldometers.info/world-population/asia-population/#:~:text=Asia%20population%20is%20equivalent%20to%2059.76%25%20of%20the%20total%20world%20population>

"Prove It!" A User-centric Client for the Blockchain-Based Research Lifecycle Transparency Framework

Chang, Yun-Chi National Taiwan University, Taiwan | r09126026@ntu.edu.tw
Kung, Li-Fei National Taiwan University, Taiwan | r10126022@ntu.edu.tw
Hsiao, Hsu-Chun National Taiwan University, Taiwan | hchsiao@csie.ntu.edu.tw
Jeng, Wei National Taiwan University, Taiwan | wjeng@ntu.edu.tw

ABSTRACT

This work develops a user-centric interface for an existing blockchain-based research transparency framework. As researchers in different disciplines may have different research data management and reproducibility needs, we first conducted an exploratory study to comprehend their motivations and expectations when managing research records. We analyzed data from a design thinking workshop with 17 participants and conducted six interviews. We have three critical components in a user-centric interface design: 1) user-decided autonomy, 2) a progress dashboard, and 3) research-team management. In the future, we would like to build an scholarly information ecosystem that fosters safe research lifecycle transparency with the assistance of technology. Thus, we plan to accumulate more interview data from different fields to improve the design of the user-centric client.

KEYWORDS

blockchain, research lifecycle transparency, user-centered design, open science, research management tool, reproducibility

INTRODUCTION

In 2021 UNESCO promoted that openness and transparency in scientific practice are essential for research advancement (UNESCO, 2021). This paper extends the study of Blockchain for Lifecycle Transparency Framework (hereafter: BLT or the BLT Framework), deliberated by Jeng et al. (2020), which introduced a decentralized framework by recording individuals' scholarly activities and promoting scientific reproducibility. This poster describes the efforts on user-centric design for developing a client tool for such a decentralized framework.

There are two fundamental design considerations in designing BLT's client. One report stated that only 39% of psychological studies published in Nature are reproducible (Baker, 2015). To avoid this, reproducibility can be greatly enhanced by recording every detail and exception condition during the experiment. In addition, misconduct during the research process could also lead to QRPs, which affect the reliability of research. Our design can enhance the notion of research transparency and eliminate questionable research practices (QRPs) and reproducibility crisis. In addition, an ideal BLT client tool should take advantage of current technology, using time- and effort-saving methods to elevate the trustworthiness of published scientific research and reduce the cost of reproducing others' work.

Secondly, our client tool should ensure the integrity of user records. Researchers should be able to present evidence to demonstrate the genuineness of research by keeping meticulous documentation during the course of their research. Even if there are some considerations that such documentation does not disclose to the public, it can function when it's necessary as a "surveillance video" of its own, demonstrating that, in general, the researcher's work is original and ensuring no one else can copy or plagiarize the ideas.

The current study conducted follow-up research based on Jeng et al. (2020): a user-centric platform was designed and implemented accordingly to promote the concept to the broader scientific community. During the design process, we interviewed, prototyped, and then iterated through a series of data gathering, to enrich the understanding of researchers' daily research activities in natural science and life science disciplines, as well as the interaction between researchers and blockchain-based design. In brief, our research questions are as follows:

RQ1: What are the motivations for researchers to record their research with blockchain?

RQ2: What are researchers' expectations of a blockchain-based research record management tool?

METHODS

First, we implemented content analysis to understand researchers' needs and barriers of sharing research data, and designed UI accordingly. Subsequently, evaluation via additional interviews was conducted to ensure we were progressing toward the goal of our research and UI would be improved after receiving feedback from the interviewees.

Content analysis: a design thinking workshop

The data from a design thinking workshop was used in the content analysis, including 17 principal investigators (PIs) in Taiwan in the life science field and the obstacles they experienced in research data management. Life science is chosen as the first field to study for the following rationales: 1) it adopted open science measures earlier than other fields (Tenopir et al., 2011) and 2) life science research contains complicated steps and numerous stakeholders, such as co-authors, clinicians, drug suppliers, etc., which highlights the importance and complexity of research data management.

Interview: user study and interface evaluation

We conducted interviews in order to better understand the users' feelings about the overall service, as well as the shortcomings of the existing design and the barriers users met. As of April 2022, we have conducted six interviews in the fields such as biomedicine, basic science, applied science, and social science. The follow-up interviews covered three aspects of the interviewees' day-to-day research practices: record keeping, storage of research records, and academic collaboration. Additionally, we asked questions about their academic ethics and UX of the prototype.

PRELIMINARY FINDINGS

In accordance with our RQs, we discovered the researchers' rationales for keeping research records, and gained understanding of their practices and expected functions of a blockchain-based research record management tool. We found that "improving academic impact" is the greatest consideration among scholars when carrying out research data management-related activities. From this point of view, the motivation can be further divided into three sub-goals: better research output and publication, managing research results properly, and maximizing the benefits of data sharing. Moreover, because of the large number of students and co-researchers, it was challenging to establish a unified and systematic data recording standard, resulting in difficulties and complexities of utilizing data, data sharing and authority management. In brief, tools we have nowadays don't meet with researchers' need of recording their research, and this situation sheds light on the necessity of designing a new and user-centric tool with blockchain.

Therefore, that leads to our RQ2: What are researchers' expectations of a blockchain-based research record management tool? Based on the data obtained, we identified three core functions of the interface, as described below.

User-decided autonomy

BLT UI's design corresponds to our aforementioned goal of protecting researchers' reputations by giving them a safe, verifiable, and immutable mechanism for recording their research data. Thus, when adding a new record to the system, the researcher can decide whether to send the added object to the blockchain.

Progress dashboard

The system's homepage is designed as a dashboard, which presents recent projects, latest records, visualized recent activities, group dynamics and account balances. Researchers can shift between projects, view each one's progress and make corresponding adjustments to ensure that the research is being conducted as planned.

Research-team management

To facilitate the complete process of publication, technology transfer, and patent application, we designed the new UI in the spirit of project management. Specifically, there are three user classes, Owner, Editors and Viewers, each with its own level of access control. To better display BLT UI, the pictures of different editions of prototype is presented on OSF: <https://osf.io/5g2jr/>

FUTURE RESEARCH PLAN

The ultimate goal of this current work is to build an ecosystem that fosters secure research lifecycle transparency with the assistance of technology. Firstly, we plan to conduct more interviews from researchers in different fields, and conduct an iterative process to improve our UI design accordingly. Meanwhile, we expect to conduct more studies regarding user's mental model of perceiving technical value in a blockchain-based data record platform. Through future studies, we can enhance the delivery efficiency of decentralized technical value, reducing their learning barrier, to provide better user experience for our users. Both directions of next steps are in the hopes of increasing the usability of the system, providing users with better research-data recording tools, improving the use of information-recording systems by researchers in different disciplines, and better understanding their interactions with the scholarly information system. Ultimately, we can provide a concrete guideline to related blockchain-based research management tools.

ACKNOWLEDGEMENT

This work was financially supported by the Ministry of Science and Technology (MOST) in Taiwan, under MOST 111-2636-H-002-004- and MOST 111-2634-F-002-018-, and the Center for Research in Econometric Theory and Applications (Grant no. 111L900204) from The Featured Areas Research Center Program within the framework of

the Higher Education Sprout Project, and the Universities and Colleges Humanities and Social Sciences Benchmarking Project (Grant no. 111L9A002) by the Ministry of Education (MOE) in Taiwan. Please address all correspondence to Wei Jeng (wjeng@ntu.edu.tw).

REFERENCES

- Baker, M. (2015). First results from psychology's largest reproducibility test. *Nature*, 30(10.1038).
- Bausell, R. B. (2021). *The Problem with Science: The Reproducibility Crisis and What to Do about It*. Oxford University Press.
- Jeng, W., Wang, S. H., Chen, H. W., Huang, P. W., Chen, Y. J., & Hsiao, H. C. (2020). A decentralized framework for cultivating research lifecycle transparency. *Plos one*, 15(11), e0241496.
- Mai, A., Pfeffer, K., Gusenbauer, M., Weippl, E., & Krombholz, K. (2020). User mental models of cryptocurrency systems-a grounded theory approach. *Proceedings of the Sixteenth USENIX Conference on Usable Privacy and Security*, 341-358. <https://www.usenix.org/system/files/soups2020-mai.pdf>
- Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Percie du Sert, N., ... & Ioannidis, J. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1(1), 1-9.
- Norman, D.A.(1983). Some Observations on Mental Models. In Gentner, D., & Stevens, A.L. (Eds.), *Mental Models* (1st ed.)(Chap. 1, pp.1-8). Psychology Press. <https://doi.org/10.4324/9781315802725>
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., ... & Frame, M. (2011). Data sharing by scientists: practices and perceptions. *Plos one*, 6(6), e21101.
- UNESCO. (2021). *UNESCO Recommendation on Open Science*. [Unesdoc.unesco.org](https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en). Retrieved April 13, 2022, from <https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en>

Identifying Immediate Impact Disciplines and Journals by Combining Three Citation-Based Indicators

Chang, Yu-Wei

National Taiwan University, Taiwan | yuweichang2013@ntu.edu.tw

ABSTRACT

This study identifies immediate impact disciplines and journals by using a new indicator, which combines the journal impact factor (JIF), five-year impact factor (5-year IF), and immediacy index (II). Although the numbers of immediate impact journals in both the natural sciences (NS) and social sciences (SS) have increased, NS disciplines have a higher average level of immediate impact than SS disciplines. Among 12,272 NS and SS journals indexed by the 2020 version of *Journal Citation Reports*, most (62.7%) exhibited the highest value of journal impact in 5-year IF and lowest value in II, followed by journals with a JIF value higher than that of 5-year IF (34.3%). Although only 113 journals (0.9%), including SS journals, had the highest value in II and the lowest value in 5-year IF, the finding rejects the stereotype regarding the nonimmediate impact of SS journals.

KEYWORDS

Immediate impact; journal impact factor; 5-year impact factor; immediacy index.

INTRODUCTION

Journal Citation Reports (JCR) has long used citation-based indicators to annually measure the impact of leading journals in a wide range of disciplines, including both the social sciences and natural sciences. The higher the indicator value, the greater the journal impact. The study focused on three longest-established indicators: the immediacy index (II), journal impact factor (JIF), and five-year impact factor (5-year IF). JCR assigns each journal one or more subject categories. The categories are widely used as a proxy for disciplines (Porter & Rafols, 2009). For example, journals in the category of “mathematics” are defined as mathematics journals. Journals belonging to the same category can be ranked on the basis of II, JIF, and 5-year JIF.

II, JIF, and 5-year IF present the average number of citations per paper. The indicators comprise two elements: the number of citations received by any paper in a given journal in a given period as the numerator, and the number of substantive papers (articles, reviews, and proceeding papers) published during the same period as the denominator. II is calculated by dividing the number of citations to papers published in a given year by the number of scholarly papers published in the same year. JIF measures the citation rate by dividing the number of citations received by papers published in the previous 2 years by the number of scholarly papers published in the previous 2 years; The 5-year IF is identical to the JIF with the only difference being that it applies to a 5-year period.

The values of JIF, 5-year IF, and II can reflect the speed of journal information dissemination and vary by discipline. It is common that journals' highest impact value is found for 5-year IF, followed by JIF and II. Natural sciences disciplines have a higher immediate impact than social sciences disciplines, in terms of speed of information dissemination. For instance, newly published papers in natural sciences journals tend to obtain more citations than those in social sciences journals (Liang et al., 2019). However, differences in immediate impact in terms of information dissemination also exist among disciplines within the natural sciences and social sciences and among journals in same discipline. Numerous factors contribute to the number of citations received by a paper (Tahamtan et al., 2016). In addition, previous studies related to citation-based indicators only focused on the correlation between two indicators (Chang et al., 2011; Midorikawa et al., 1984; Okagbue & Teixeira da Silva, 2020; Smart & Elton, 1982; Yue et al., 2004). The relationships among the II, JIF, and 5-year IF is neglected.

Journals in the same discipline are assumed to exhibit similar relationships among II, JIF, and 5-year IF. Therefore, immediate impact journals without large 5-year IF values and low II values are considered anomalous. This present study proposes a new indicator, combining three target indicators to identify the immediate impact journals. Moreover, disciplines with immediate impact journals are identified as the immediate impact disciplines. Two main research questions were addressed in this study: Do disciplines in natural sciences have higher proportion of immediate impact journals than those in social sciences? What are the characteristics of immediate impact journals by discipline?

METHODOLOGY

Data collection

The sample comprised journals in the 2020 and 2010 versions of the JCR database from SCIE and SSCI. Each journal was assigned one or more categories. In addition, the JIF, 5-year IF, and II values for each journal were collected.

Data processing

Three figures were calculated: the II value minus the JIF value (A), the JIF value minus the 5-year IF value (B), and the II value minus the 5-year IF value (C). Immediate impact journals were defined as those for which at least A, B, or C was greater than 0. Immediate impact disciplines were those associated with at least one immediate impact journal. The immediate impact level (III) for each discipline was calculated by summing the proportion of journals with $A > 0$, $B > 0$, and $C > 0$. For example, if a discipline had 0.2% of journals with $A > 0$, 30% with $B > 0$, and 0.8% with $C > 0$, then the discipline's III is 31.0%.

RESULTS

Differences and changes in immediate impact level by discipline

Figure 1 illustrates a wide variation of III across 178 NS disciplines and 58 SS disciplines, based on the 2020 version of JCR. The top 33 disciplines with highest III (61.7%–94.1%) are NS disciplines. The highest III owned by SS disciplines achieves 61.5%. Comparing analyses based on data from the 2010 and 2020 versions of JCR, it is found that both the NS and SS disciplines have enhanced their III. However, a larger extent of III growth is observed in NS disciplines. Thus, the gap in III between the NS and SS disciplines has increased. To examine the difference in III for each journal, I take 12,274 NS and SS journals indexed by the 2020 version of JCR as the example to be analyzed, to distinguish journals into six types by their III (Table 1).

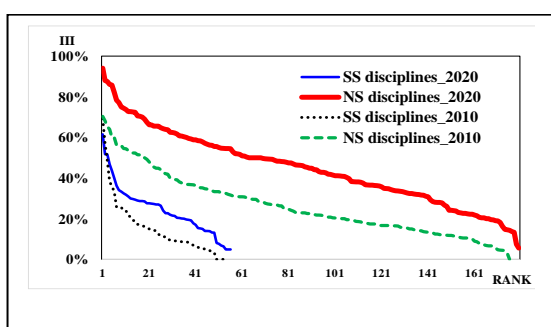


Figure 1. Immediate impact level by discipline

Type	Definition	Journals	%
1	5-year IF > JIF > II	7,699	62.7
2	JIF > 5-year IF	4,206	34.3
3	II > JIF & II > 5-year IF	128	1.0
4	II > JIF > 5-year IF	113	0.9
5	II > JIF	72	0.6
6	JIF > 5-year IF & II > 5-year IF	56	0.5
Total		12,274	100.0

Table 1. Distribution of journals by type

Types of journals by level of immediate impact

Most journal titles (62.7%) relied on the passing of time to enhance their impact. The descending order of values of the three citation-based indicators is consistent with the assumption that 5-year IF > JIF > II. The second-largest group of journals featured a higher value of JIF than that of 5-year IF (34.3%), suggesting that a considerable proportion of journals demonstrated their citation impact in the second and third years after publication. Except for type 6, each journal classified in types 3, 4, and 5 had its highest value in II, rather than in JIF or 5-year IF. This indicates that substantial immediate impact could be observed after papers were published within one year. In particular, characteristics of the impact of journals in type 4 were the opposite of type 1 journals. Type 4 includes SS journals on topics such as history and political science. This indicates that immediate impact journals are not limited to NS journals.

CONCLUSION

This study confirmed that some NS and SS disciplines and journals have unique characteristics in terms of immediate impact level, measured by the combination of JIF, 5-year IF, and II. Journals with higher immediate impact demonstrated their primary impact in less than one year after the relevant papers were published. The results rejected the stereotype that SS journals have a nonimmediate impact. In addition, higher immediate impact journals did not benefit from their values of JIF and 5-year IF, which are used for ranking journals in corresponding disciplines. Therefore, disciplines and journals with an increasing trend in their level of immediate impact are worth monitoring, given that their characteristics have been changing. This tracking can be effectively implemented by using the immediate impact indicator.

ACKNOWLEDGMENTS

This work was financially supported by the Ministry of Science and Technology (MOST), Taiwan, under Grant No. MOST 110-2410-H-002 -233 -

REFERENCES

- Chang, C.-L., Mcaleer, M., & Oxley, L. (2011). How are journal impact, prestige and article influence related? an application to neuroscience. *Journal of Applied Statistics*, 38(11), 2563–2573.
- Liang, G., Hou, H., Lou, X., & Hu, Z. (2019). Qualifying threshold of “take-off” stage for successfully disseminated creative ideas. *Scientometrics*, 120(3), 1193–1208.

- Midorikawa, N., Ogawa, H., Saito, K.-I., Kaneko, M., & Isumura, H. (1984). The relationships among the citation measures and the factors influence on them. *Information Services and Use*, 4(6), 417-424.
- Okagbue, H. I., & Teixeira da Silva, J. A. (2020). Correlation between the CiteScore and Journal Impact Factor of top-ranked library and information science journals. *Scientometrics*, 124(1), 797–801.
- Porter, A. L., & Rafols, I. (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics*, 81(3), 719–745.
- Smart, J. C., & Elton, C. F. (1982). Consumption factor scores of psychology journals: Scientometric properties and qualitative implications. *Scientometrics*, 4(5), 349–360.
- Tahamtan, I., Safipour Afshar, A., & Ahamdzadeh, K. (2016). Factors affecting number of citations: a comprehensive review of the literature. *Scientometrics*, 107(3), 1195–1225.
- Yue, W., Wilson, C. S., & Rousseau, R. (2004). The immediacy index and the journal impact factor: Two highly correlated derived measures. *Canadian Journal of Information and Library Science*, 28(1), 33–48.

A Preliminary Analysis of Geography of Collaboration in Data Papers by S&T Capacity Index

Chen, Pei-Ying

Indiana University Bloomington, USA | peychen@iu.edu

Li, Kai

Renmin University of China, People's Republic of China | kai.li@ruc.edu.cn

Jiao, Chenyue

University of Illinois at Urbana-Champaign, USA | cjiao4@illinois.edu

ABSTRACT

Geography is one of the defining factors in scientific collaboration. Despite the voluminous evidence for how geographical proximity shapes the formation of collaboration in research articles, it has been rarely examined in the emerging genre of data papers, one that describes research data and has enjoyed growing attention in the data-driven paradigm of research. This poster presents preliminary findings from our project that aims to evaluate the geographical dynamics behind the production of data papers. We analyze how researchers from different countries collaborate with one another using 6,821 data papers published in *Scientific Data* and *Data in Brief* between 2014 and 2020. We found that data papers rely heavily upon domestic collaboration and the collaboration pattern largely mirrors that of research articles, although some distinctiveness was also observed. We discuss future work in conclusion, with the ultimate goal of opening a more meaningful conversation about the relationship between the data-driven paradigm and knowledge production.

KEYWORDS

data papers; geography; scientific collaboration

INTRODUCTION

The famous quote from Louis Pasteur that “science knows no country” illustrates the universalism of the scientific system. However, it is also obvious that “science takes place” (Olechnicka, Ploszaj & Celińska-Janowicz, 2018; p. 4). While empirical studies have shown that geographical proximity is positively related to the formation of collaboration networks (Hoekman, Frenken & van Oort, 2009; Pan, Kaski & Fortunato, 2012), cultural/linguistic affinity as well as historical and socio-economic factors are also important determinants in international collaboration (Zitt, Bassecoulard & Okubo, 2000).

As an emerging scientific genre that describes research data objects, data papers introduce a novel mode of knowledge production and presentation under the data-driven paradigm of research (Li & Jiao, 2022). However, we are yet to know whether there are distinct geographical dynamics in the production of data papers (as compared to research articles) in terms of collaboration pattern and correspondence between authors' physical locations and their subjects of study. In this poster, we report preliminary findings on the geography of collaboration among data paper authors, as the first step towards a more thorough understanding of the impact of the data-driven paradigm on knowledge production.

DATA AND METHODS

In this study, we collected 504 data papers published in *Scientific Data* and 6,332 in *Data in Brief* from Scopus on November 15, 2020. The two journals were selected as the two leading exclusively data journals (Kim, 2020; Walters, 2020). We extracted country names from authors' affiliations, using the *countrycode* R package (Arel-Bundock, Enevoldsen & Yetman, 2018) to examine the collaboration pattern at the country level. After excluding papers with no identifiable country-level information on authors' affiliations, the final sample consisted of 6,821 data papers for the present analysis.

We distinguished between single-author, domestic collaboration, as well as leading versus supporting author(s) in international collaboration based on the number of authors, number of countries, and order of countries in authors' affiliation information per paper. To facilitate discussion, we adopted the S&T Capacity Index developed by Wagner, Brahmakulam, Jackson, Wong and Yoda (2001) that classifies 150 countries/territories into four groups: scientifically advanced countries (SAC), scientifically proficient countries (SPC), scientifically developing countries (SDC), and scientifically lagging countries (SLC). Our sample comprises 146 countries/territories, including 22 SAC, 23 SPC, 20 SDC, and 61 SLC.

PRELIMINARY FINDINGS

Over two thirds of the data papers (67.7%) in our sample result from domestic collaboration (i.e., co-authors in the same country), 28.6% international collaboration, and 3.7% single-author papers. The most productive country is the United States (n=1,330), which authored papers more than twice as much as China (n=612), the second most productive country. Figure 1A shows the top 20 most productive countries and their collaboration profiles. While just over half of them are SAC, it is notable the remaining half is composed of not only SPC (China, India, Spain,

Brazil, and South Africa) but also SDC (Iran and Colombia) and SLC (Nigeria and Malaysia). Moreover, there seems to be a trade-off between % of domestic collaboration and % supporting authorship in international collaboration, as the top five countries in the latter are also the bottom five in the former. If we consider only the number of leading and supporting authorship in international collaboration, it appears that SAC and SPC are more likely to be in supporting roles, while countries with more leading than supporting authorship tend to be SDC and SLC (Figure 1B). As both figures show, however, there are considerable within-group variations, indicating there are complex factors that weigh in on collaboration dynamics.

A further examination of the collaboration pattern by S&T capacity in leading and supporting authorship roles reveals that SAC are the most popular collaborating partners for countries across all four groupings, especially among SPCs—nearly 70% of the SPC-led papers have co-authors in SAC. While researchers tend to collaborate with their counterparts in countries with higher S&T capacities, a significant share (26%) of collaboration between SLCs is also present (Figure 1C). In fact, SAC and SLC are both groups with relatively high proportions of within-group collaboration. Although numerous studies have noted the frequent collaboration between SACs (Gazni, Sugimoto & Didegah, 2012; Wagner et al., 2001), the collaboration between SLCs is less well documented, let alone in data papers. Given that the majority of SLCs are in the Global South, which is shown to have comparative advantages in disciplines related to natural resources and infectious diseases (Miao et al., 2022), it would be interesting to investigate how the observed collaboration patterns in data papers vary by discipline.

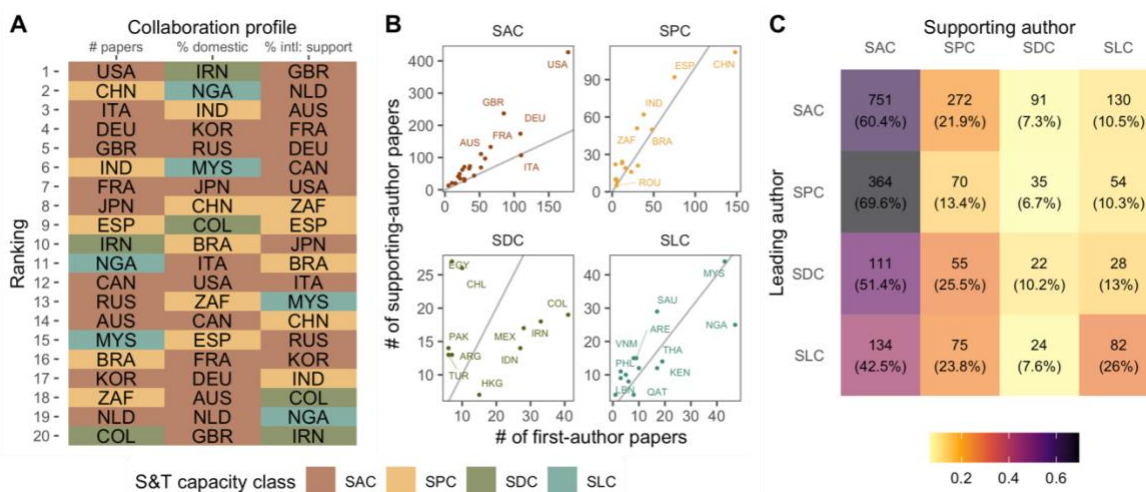


Figure 1. (A) Collaboration profile among top 20 most productive countries. (B) First- vs. supporting-author papers among countries with at least 19 papers (median). (C) Collaboration pattern by S&T capacity. Note: (1) Qatar is manually assigned to SLC. (2) The gray lines represent diagonals with intercept=0 and slope=1.

CONCLUSION

In this poster, we present preliminary findings from our project that aim to investigate the geographical dynamics in the production of data papers. Early evidence suggests that the overall collaboration pattern resembles that of the general scientific research, especially the predominance of domestic collaboration and the high concentration of collaboration not only between SACs but also between other less developed countries and SACs. However, we also found a relatively high proportion of within-group collaboration among SLCs, which warrants further investigation into possible cross-discipline variation.

Based on these results, there are three possible directions to be taken in future research. First, we will examine the geographic proximity between authors' affiliations and their subject of study to better understand whether such proximity plays a role in the division of labor in collaboration. Second, we will take discipline into consideration to illustrate the more granular co-authorship patterns in data papers. Third, we will compare the collaboration patterns of data papers with those of research articles more systematically to get a better idea of the distinct characteristics of data papers as a new academic genre.

ACKNOWLEDGMENTS

We thank three anonymous reviewers who wrote and provided helpful comments on previous versions of this document.

REFERENCES

Arel-Bundock, V., Enevoldsen, N. & Yetman, C. (2018). countrycode: An R package to convert country names and country codes. *Journal of Open Source Software*, 3(28), 848. doi:10.21105/joss.00848

- Gazni, A., Sugimoto, C. R. & Didegah, F. (2012). Mapping world scientific collaboration: Authors, institutions, and countries. *Journal of the American Society for Information Science and Technology*, 63(2), 323-335. doi:10.1002/asi.21688
- Hoekman, J., Frenken, K. & van Oort, F. (2009). The geography of collaborative knowledge production in Europe. *The Annals of Regional Science*, 43(3), 721-738. doi:10.1007/s00168-008-0252-9
- Kim, J. (2020). An analysis of data paper templates and guidelines: Types of contextual information described by data journals. *Science Editing*, 7(1), 16-23. doi:10.6087/kcse.185
- Li, K. & Jiao, C. (2022). The data paper as a sociolinguistic epistemic object: A content analysis on the rhetorical moves used in data paper abstracts. *Journal of the Association for Information Science and Technology*, 73(6), 834-846. doi:10.1002/asi.24585
- Miao, L., Murray, D., Jung, W.-S., Larivière, V., Sugimoto, C. R. & Ahn, Y.-Y. (2022). The latent structure of global scientific development. *Nature Human Behaviour*. doi:10.1038/s41562-022-01367-x
- Olechnicka, A., Ploszaj, A. & Celińska-Janowicz, D. (2018). *The geography of scientific collaboration*. Routledge.
- Pan, R. K., Kaski, K. & Fortunato, S. (2012). World citation and collaboration networks: Uncovering the role of geography in science. *Scientific Reports*, 2(1). doi:10.1038/srep00902
- Wagner, C. S., Brahmakulam, I. T., Jackson, B. A., Wong, A. & Yoda, T. (2001). *Science & technology collaboration: Building capacity in developing countries?*. RAND Corporation. Retrieved from https://www.rand.org/pubs/monograph_reports/MR1357z0.html
- Walters, W. H. (2020). Data journals: Incentivizing data access and documentation within the scholarly communication system. *Insights: the UKSG journal*, 33. doi:10.1629/uksg.510
- Zitt, M., Bassecouard, E. & Okubo, Y. (2000). Shadows of the past in international cooperation: Collaboration profiles of the top five producers of science. *Scientometrics*, 47(3), 627-657. doi:10.1023/a:1005632319799

Engagement for Good or Ill: Comparing Characteristics of Co-Creative and Co-Destructive Online Communities

Chen, Yimin University of Western Ontario, Canada | ychen582@uwo.ca
Cornwell, Sarah, E. University of Western Ontario, Canada | scornwel@uwo.ca
Delellis, Nicole, S. University of Western Ontario, Canada | ndelelli@uwo.ca
Kelly, Dominique University of Western Ontario, Canada | dkelly48@uwo.ca
Liu, Yifan University of Western Ontario, Canada | yliu4369@uwo.ca
Mayhew, Alex University of Western Ontario, Canada | amayhew@uwo.ca
Rubin, Victoria, L. University of Western Ontario, Canada | vrubin@uwo.ca

ABSTRACT

Highly engaged online communities can be characterized in terms of level of user activity and the valence of those activities. Positively-valenced communities engage in co-construction through supportive, uplifting interactions, while negatively-valenced communities engage in co-destruction through shared animosity towards outside individuals and groups. Both types of engagement offer prosocial benefits to community members, but this benefit is achieved through antisocial means in co-destructive communities. This study examines the features of one positively-valenced and one negatively-valenced subreddit community to determine if the same engagement factors drive participation in each, and whether these factors may be influenced to promote positive, rather than negative engagement.

KEYWORDS

Social media; online communities; prosocial and antisocial behavior; reddit

INTRODUCTION

There are a number of ways by which the success of online communities has been defined. Among the most frequently cited of these measures is the level of participation by individuals within the group (Arguello et al., 2006; Cunha et al., 2019; Malinen, 2015). Forms of participation may include asking questions and sharing information, chatting with or messaging other community members, or posting content (Oh et al., 2014). For individuals, participation influences their perceptions of reciprocity and belonging, which helps newcomers integrate into and become engaged within the community (Casaló et al., 2013). In turn, a greater sense of community may encourage people to more actively engage in supportive interactions in those communities, which is associated with greater life satisfaction and subjective well-being (Oh et al., 2014). We can broadly describe such supportive communities as prosocial, in that individuals act in ways that benefit other members and the community as a whole (Dovidio et al., 2006). Many such communities are formed as spaces where people can share and discuss common interests, such as hobbies, ethnicity or culture, or political ideology. However, while highly engaged communities may be successful in terms of creating and maintaining a cohesive sense of shared identity and reciprocation, they do not always act in ways that are prosocial towards people outside of those communities. One example is the notoriously antisocial 4chan, which has historically been a prolific and influential online community with a strong group identity, but where much of the community participation and content is intentionally offensive or hateful (Bernstein et al., 2011).

The prosocial and antisocial styles of online community participation can be conceptualized through the typology of social media engagement behaviors proposed by Dolan et al. (2015), which describes engagement along two axes: activity and valence. Under this model, high-engagement behaviors can either be considered co-creation, where users collaborate in positive content creation and sharing, or co-destruction, where content creation and sharing still occur, but take antagonistic forms (Dolan et al., 2015). Viewed through this lens, even antisocial communities can be considered “successful” insofar as they exhibit high levels of engagement, shared sense of identity, and reciprocation between members. For example, a study of individuals involved in misogynistic and anti-feminist involuntarily celibate (Incel) groups online found that many “entered the Incel community seeking support, belonging, and above all, friendships and relationships” (Regehr, 2022, p. 148). Although clearly negatively-valenced, these Incel communities nonetheless offered “a supportive environment for lonely individuals” (Regehr, 2022, p. 152). Both co-creative and co-destructive communities seem to provide supports and benefits to their members, but co-destructive communities do so at the expense of others outside the community.

RESEARCH OBJECTIVE

Thus far, studies of highly-engaged online communities have tended to focus on either positively-valenced or negatively-valenced groups, exclusively. There has yet to be substantial work investigating how these positive and

negative groups compare to each other. Do the same engagement factors drive participation in co-creative and co-destructive communities? Do these factors influence whether a community tends toward co-creation or co-destruction? Could these factors be used to encourage more positively-valenced engagement, rather than negative? Are there other latent factors that may impact user participation?

This exploratory work-in-progress will provide a descriptive analysis of two active, oppositely-valenced reddit communities: r/MadeMeSmile and r/KotakuInAction. The r/MadeMeSmile subreddit self-describes as “a place to share things that made you smile or brightened up your day” (n.d.) and explicitly calls upon users to post “uplifting” content. By contrast, r/KotakuInAction is the self-described “main hub for GamerGate on Reddit” (n.d.) and has been characterized as a haven for misogynistic internet trolls (Jhaver et al., 2018). Using r/MadeMeSmile as an example of a co-creative community and r/KotakuInAction as co-destructive, this study will examine 1) participation patterns, 2) types and frequencies of content posts, user comments, and other interactions between members, and 3) demographics. Preliminary analysis of one week of subreddit comments collected from June 5-11, 2022, via Commalytic is shown below (Table 1).

	Negative Sentiment*	Neutral Sentiment*	Positive Sentiment*	Toxicity†
r/MadeMeSmile				
r/KotakuInAction	41.5%	16.2%	42.3%	0.25

Table 1. Sentiment and toxicity analysis of r/MadeMeSmile and r/KotakuInAction

*Based on VADER sentiment analysis: <https://pypi.org/project/vaderSentiment>

†Perspective API toxicity analysis: <https://www.perspectiveapi.com>

CONCLUSION

Positive interpersonal interactions on social networking sites have been associated with greater happiness and life satisfaction (Zell & Moeller, 2018), but in co-destructive communities, this prosocial benefit is sometimes achieved through antisocial means. This study examines the features of one positively-valenced and one negatively-valenced subreddit community in order to determine the engagement factors that drive participation in each and whether these factors influence the community’s engagement valence. Recognizing that the affordances of social networking sites play a role in incentivizing and disincentivizing different types of user behaviours (Kelly et al., 2022), we hope that insights gained from this study could be used to design online spaces which promote more positive user engagement.

ACKNOWLEDGEMENTS

This work has been funded by an institutional Strategic Support for the Canadian Sciences and Humanities Research Council (SSHRC) Success grant, entitled “Rewiring for Happiness: Computing for Kinder, Healthier, and Compassionate Life Experiences with Information Commutation Technologies (ICTs).”

REFERENCES

- Arguello, J., Butler, B. S., Joyce, E., Kraut, R., Ling, K. S., Rosé, C., & Wang, X. (2006, April). Talk to me: foundations for successful individual-group interactions in online communities. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 959-968). <https://doi.org/10.1145/1124772.1124916>
- Bernstein, M., Monroy-Hernández, A., Harry, D., André, P., Panovich, K., & Vargas, G. (2011). 4chan and/b: An Analysis of Anonymity and Ephemerality in a Large Online Community. In *Proceedings of the international AAAI conference on web and social media* (Vol. 5, No. 1, pp. 50-57). <https://ojs.aaai.org/index.php/ICWSM/article/view/14134>
- Casaló, L.V., Flavián, C., & Guinalú, M. (2013). New members’ integration: Key factor of success in online travel communities. *Journal of Business Research*, 66(6), 706-710. <https://doi.org/10.1016/j.jbusres.2011.09.007>
- Cunha, T., Jurgens, D., Tan, C., & Romero, D. (2019, May). Are all successful communities alike? Characterizing and predicting the success of online communities. In *The World Wide Web Conference* (pp. 318-328). <https://doi.org/10.1145/3308558.3313689>
- Dolan, R., Conduit, J., Fahy, J., & Goodman, S. (2016). Social media engagement behaviour: a uses and gratifications perspective. *Journal of strategic marketing*, 24(3-4), 261-277. <https://doi.org/10.1080/0965254X.2015.1095222>
- Dovidio, J.F., Piliavin, J.A., Schroeder, D.A., & Penner, L.A. (2006). *The Social Psychology of Prosocial Behavior*. Lawrence Erlbaum: Hillsdale, NY.

- Jhaver, S., Chan, L., & Bruckman, A. (2018). The view from the other side: The border between controversial speech and harassment on Kotaku in Action. *First Monday*, 23(2). <https://doi.org/10.5210/fm.v23i2.8232>
- Kelly, D., Liu, Y., Mayhew, A., Chen, Y., Cornwell, S. E., Delellis, N. S., & Rubin, V. L. (2022). Supporting Prosocial Behaviour in Online Communities through Social Media Affordances. *AM22: 85th Annual Meeting of the Association for Information Science and Technology*.
- Malinen, S. (2015). Understanding user participation in online communities: A systematic literature review of empirical studies. *Computers in human behavior*, 46, 228-238. <https://doi.org/10.1016/j.chb.2015.01.004>
- Oh, H.J., Ozkaya, E., & LaRose, R. (2014). How does online social networking enhance life satisfaction? The relationships among online supportive interaction, affect, perceived social support, sense of community, and life satisfaction. *Computers in Human Behavior*, 30, 69-78. <https://doi.org/10.1016/j.chb.2013.07.053>
- Regehr, K. (2022). In (ce)l doctination: How technologically facilitated misogyny moves violence off screens and on to streets. *New Media & Society*, 24(1), 138-155. <https://doi.org/10.1177/1461444820959019>
- r/KotakuInAction. (n.d.). *Reddit*. <https://www.reddit.com/r/KotakuInAction>
- r/MadeMeSmile. (n.d.). *Reddit*. <https://www.reddit.com/r/MadeMeSmile>
- Zell, A. L., & Moeller, L. (2018). Are you happy for me... on Facebook? The potential importance of “likes” and comments. *Computers in Human Behavior*, 78, 26-33. <https://doi.org/10.1016/j.chb.2017.08.050>

Census Racial Metadata: Categories Used for Different Racial Groups Across Countries and Time

Cheng, Siyao

University of Illinois at Urbana-Champaign, USA | siyaoc2@illinois.edu

Twidale, Michael

University of Illinois at Urbana-Champaign, USA | twidale@illinois.edu

ABSTRACT

A national census enables a government to comprehend its population. Questions concerning race and ethnicity are often included in census forms. We examined the racial category options provided in 35 countries' historical census forms over multiple decadal censuses. Analysis of these metadata options can help in understanding the way that official measures of race are socially constructed and reconstructed over place and time according to changing circumstances.

KEYWORDS

Race, ethnicity, racism, census metadata, self-identification

BACKGROUND

Census forms typically serve as a demographic gold standard, allowing government agencies to understand and plan for the evolving needs of their population. In accordance with the United Nations census creation suggestions, different countries' census forms typically include questions about age, gender, occupation, education, etc. (United Nations, 2017). Racial and ethnicity-related questions are also included. The way the census is filled out has evolved, often associated with rising literacy levels in a country. Increasingly respondents fill out the form themselves, instead of the form being filled out for them. A consequence is that with respect to questions about race & ethnicity, respondents are now enabled (or required) to define themselves in census forms instead of census takers identifying respondents based on their physical appearance (Saperstein & Penner, 2014).

Such an opportunity for self-identification can be empowering, but issues of metadata and interface design can have an impact. What are the options that you are given to choose from for the purpose of self-identification? How do the design of the form and the language used to describe the category options align with how you see yourself, how other people see you, how the government sees you, and how you see others seeing you? Inspired by Bowker & Star's (2000) work on the issues around definitions of racial classification and the associated metadata, we are investigating this issue in the census metadata of multiple countries.

METHOD

As an exploratory research approach, we collected demographic census metadata from the following countries: Australia, Austria, Bangladesh, Belgium, Brazil, Canada, China, Cyprus, Egypt, Fiji, France, Ghana, Guyana, India, Indonesia, Ireland, Israel, Italy, Jamaica, Kenya, Malaysia, Malta, Mexico, Netherlands, New Zealand, Nigeria, Pakistan, Philippines, Singapore, South Africa, Sri Lanka, Sweden, Switzerland, Trinidad & Tobago, and the USA. Countries were selected as a convenience sample, with a bias towards those where information was available in English. The metadata was obtained from: the countries' government websites. We retrieved the race and ethnicity-related questions and their options, tracing them back to the earliest decadal census forms that were available on the website. This allowed us to make comparisons of changes in racial categories both within a given country over time, and the similarities and differences between selected countries. We paid particular attention to the degree of granularity and when and whether people are permitted to use multi-racial combinations in the census forms.

PRELIMINARY FINDINGS

Racial categories are continually changing

Unsurprisingly, each country has its own interpretations of the meaning of 'race' and 'ethnicity', and different approaches when designing the related questions. Those racial categories and questions vary in response to each country's political currents, public attitudes, and social constructions, especially variations in patterns of immigration. 27 out of 35 selected countries do not use the term 'race'. This may be for a range of reasons including (particularly for European countries) the term's negative associations with Nazi racial laws. Instead, questions of ethnic group, nationality, ancestry, country of origins, and even the language frequently spoken at home are used. This can result in permitted values from different conceptual categories being jumbled together into a single fusion category whose precise definition may be unsaid or unsayable. Other countries (e.g. the USA) have race and ethnicity as two distinct categories and respondents have to provide a value for each. Even providing definitions of certain values can be politically and socially fraught. Moreover, we found that within countries, classificatory metadata change was the norm; the category options evolving from one decade's census to the next. The options and terminology available in the metadata (Kilty & Videl de Haymes, 2004) and the wording of questions (Kirman et al., 2001) are debated and change over time (Thompson, 2020). As we know from social science methodology, the

design of a form and the exact phrasing of questions can affect the results that are obtained; an issue that applies not just to research but to the design of government forms. Providing many options embraces a richer diversity of identities but can also make the form more complex, harder to read, and more daunting to answer. The provision of a value of “Other” can be perceived as a kind acknowledgement of diversity, or rather offensive—a literal ‘othering’. Allowing people to write in their own self-identification increases freedom but complexifies quantification. Considered as an interface design issue, census forms necessarily have to deal with multiple design trade-offs: legal, political, and social.

Identity options are evolving: both fragmenting and merging

In the interests of subsequent statistical analysis, permitted responses are often constrained to one of a few options. Such analytic tidiness may be convenient for bureaucratic purposes but may not fit well with human diversity. Especially with rising levels of immigration and evolving self-identity, there may be a need to provide more options. Requests from certain minorities may lead to recognition of distinct identities separate from broader categorizations. Furthermore, there may be increasing pressure to permit people to express more than one identity—even if that causes problems with counting and classifying people into tidy boxes. For example, it is only since 2000 that people in the USA have been permitted to say that they have more than one race.

CONCLUSION

Organizations of all sizes, including countries, may collect data on race and ethnicity for a range of purposes, both benign and malign. It is now well established by contemporary social scientists that race is a social construct, but one that can have a significant impact on how people are treated by others, are treated by society and how they see themselves. It can be easy to say and to agree that “race is a social construct”, but actually seeing race being socially constructed differently in different places (Fiola, 2004; Reddy, 2016) and at different times makes the idea of social construction much more salient. Accurate data can be helpful in recognizing and then addressing concerns of racist social inequality (Kelly, 2020) and health (Scott & Unger, 2018), but that requires the data to reflect the reality of how people see themselves—And how others see them. When data about race is collected by governments on official forms, it can be seen as being more ‘official’, or even definitive. So when a census does not allow people to select more than one race, it can be interpreted as meaning that people of more than one race do not exist, or do not matter, or that they must be seen as only having one race. And when that rule changes in the next census, what has changed? Our data or our reality? Realizing that people can be categorized differently by different countries can help in providing a different perspective of how metadata embodied in bureaucratic forms about race and ethnicity can be at any one place and time overly rigid, while also revealing perhaps surprising fluidity (Saperstein & Penner, 2014) between places and times. The implications of this study establish a foundation for future research and investigation into the variability and potential unreliability that might occur in demographic reporting and the choosing of categories.

REFERENCES

- Bowker, G. C., & Star, S. L. (2000). The Case of Race Classification and Reclassification under Apartheid. *Sorting things out: Classification and its consequences*. MIT press.
- Fiola, J. A. (2004). Blackness Without Ethnicity: Constructing Race in Brazil (Book). *Choice: Current Reviews for Academic Libraries*, 41(7), 1352–1353.
- Kelly, M. D. (2020). Examining Race in Jamaica: How Racial Category and Skin Color Structure Social Inequality. *Race and Social Problems*, 12(4), 300–312.
- Kilty, K. M., & Videl de Haymes, M. (2004). What’s in a name? Racial and ethnic classifications and the meaning of Hispanic/Latino in the United States. *Ethnic Studies Review*, 27(1), 32–56.
- Kirnan, J., Bragge, J. D., Brecher, E., & Johnson, E. (2001). What race am I? The need for standardization in race question wording. *Public Personnel Management*, 30(2), 211–220.
- Reddy, G. (2016). Race rules in Singapore. *Singapore: Negotiating State and Society, 1965–2015*, 54–75.
- Saperstein, A., & Penner, A. (2014). Beyond the looking glass: Exploring fluidity in racial self-identification and interviewer classification. *Sociological Perspectives*, 57, 186–207.
- Scott, P., & Unger, H. V. (2018). Questioning Categorisation Practices. *Ethical, Legal and Social Aspects of Healthcare for Migrants: Perspectives from the UK and Germany*, 78.
- Thompson, D. (2020). Race, the Canadian Census, and Interactive Political Development. *Studies in American Political Development*, 34(1), 44–70.
- United Nations (2017). Principles and recommendations for population and housing censuses (Rev. 3). (2017). United Nations.

Reading Between the Lines in Discussing Multicultural Books for Children: Preliminary Results

Choi, Yunseon

Valdosta State University, USA | yunchoi@valdosta.edu

Drouillard, Colette

Valdosta State University, USA | cldrouillard@valdosta.edu

ABSTRACT

Selecting multicultural books for children is very important because multicultural literature helps children understand the issues of cultural diversity and gender biases. This study is a part of a larger research project that aims to build a culture-rich environment in organizing and selecting books for children. In this study, we investigate whether the sentiments in online reviews of children's multicultural literature would represent readers' opinions on the books and whether readers' positive, negative, or neutral attitudes would be directly associated with the overall ratings of books. The preliminary results show that the review rating scores did not accurately indicate readers' minds in discussing books. This study has an implication of increasing awareness of the ethnic or racial minority themes in children's literature by uncovering readers' minds in discussing children's multicultural books.

KEYWORDS

Online reviews; social media; social networking; text analysis; children's books

INTRODUCTION

The American Library Association's statement on Equity, Diversity, Inclusion states the embodiment of equitable and inclusive access to all the materials regardless of origin, background, or views (American Library Association, 2017). It is also addressed that all members should be able to see themselves reflected in the collections and resources provided by libraries. Multicultural literature helps children see their own culture and better understand issues on diversity and inclusiveness by making themselves open to other cultures (Adam, 2021; Colby & Lyon, 2004). Therefore, all children must see themselves reflected in multicultural collections developed by librarians. Librarians have used authoritative sources to guide book-selection decisions. Examples of the authoritative sources include book reviews that librarians have been long used for building library collections (Jenkins, 1996; Jenkins, 1999). Although the number of printed book reviews has been reduced, online book reviews have been increased (Hartley, 2018). The usefulness of online book reviews was discussed regarding their roles in forming a new type of book culture (Driscoll & Sedo, 2019). This study is a part of a larger research project that aims to build a culture-rich environment in organizing and selecting books for children. Selecting multicultural books for children is very important because multicultural literature helps children understand the issues of cultural diversity and gender biases. In this study, we investigate whether the sentiments in online reviews of children's multicultural literature would represent readers' opinions on the books and whether the overall rating scores of books would serve as an indicator of readers' minds in discussing multicultural books for children. This study, therefore, addresses the following research question: Q. Do the sentiment values of readers' reviews on children's multicultural books correspond to readers' overall rating scores of the books?

METHODOLOGY

This study investigates the polarity of online reviews with neutral, positive, or negative sentiment books selected for ALA Youth Media Honors and Awards for multicultural literature (Pura Belpré, Coretta Scott King, Schneider, and Stonewall). The study selected 27 books published between 2015 and 2019 and collected 2,640 reader reviews of the books on Goodreads. This study conducted sentiment analysis to examine the relationship between the overall rating of a book and the sentiment value of the book. Sentiment analysis, known as opinion mining and subjectivity analysis, analyzes people's opinions and attitudes from written reviews to rate products or services (Liu, 2010). Sentiment analysis identifies the emotional content or opinions in texts and determines their polarities (Huang, Yen, & Zhang, 2008). This project used the Natural Language Toolkit (NLTK) (www.nltk.org) for analyzing sentiments to analyze subjectivity and polarity (i.e., neutral, positive, or negative sentiment) of book reviews.

RESULTS AND DISCUSSION

The preliminary results of this study showed that the sentiment values of reviews on books did not necessarily result in corresponding reader rating scores for books (Figure 1). Figure 1 depicts a stacked bar chart showing sentiment values by book with color segments. The red part indicates negative sentiments, while the green and blue represent neutral and positive sentiments. Figure 1 also shows the rating scores for each book by marking them as black dots, which demonstrates that high rating scores of each book by marking them as black dots, which demonstrates that high rating scores of books do not always match the positive sentiments of the books. Despite high average ratings of some books, a majority of reader reviews were not found to have positive sentiment polarity. Coretta Scott King Award Winner, "March: Book Three" was found to have only 42% positive sentiment polarity in reviews (Figure 2)

85th Annual Meeting of the Association for Information Science & Technology | Oct. 29 – Nov. 1, 2022 | Pittsburgh, PA. Author(s) retain copyright, but ASIS&T receives an exclusive publication license.

even though the average rating of the book is high (4.14 out of 5) (Figure 1). The book titled “Step up to the plate, Maria Singh,” the winner of an Asian/Pacific American Libraries Association Award, realistically depicts a family that merges two cultures, Mexican and Indian, as the family confronts prejudice and discrimination in the 1940s. 79% of Goodreads reviews were found to have neutral sentiment polarity (Figure 2), with an average reader rating of 3.82/5.0. The Stonewall Book Award winner, George, is a story about a young transgender girl, and the overall ratings of reviews regarding this book had an average reader rating of 4.05/5.0. However, 16% of Goodreads reviews were found to have negative sentiment polarity (which is higher than found in the other two books). In comparison, 64% of reviews were found to have positive sentiment polarity (also higher than two other books) (Figure 2).

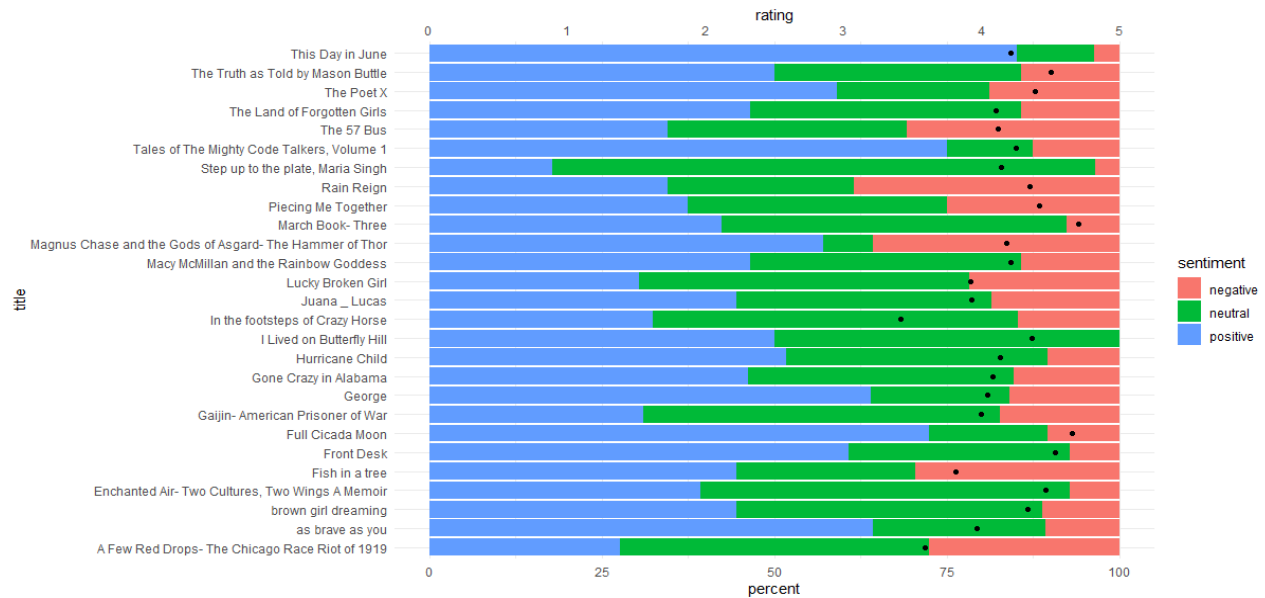


Figure 1. Ratings vs. Sentiment values (Neutral, Positive, and Negative sentiment)

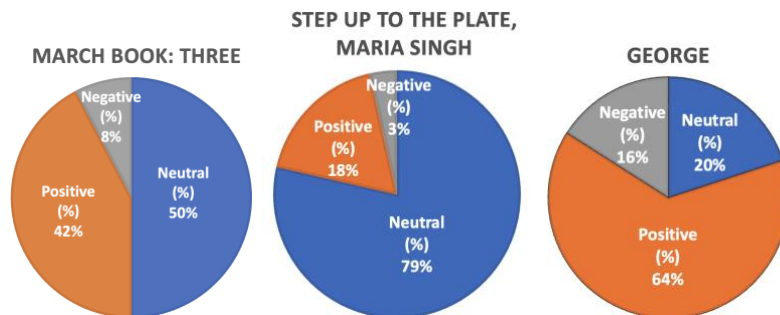


Figure 2. Neutral, Positive, and Negative Sentiment Values about Three Book Titles

CONCLUSION

This study analyzed the sentiment values of reviews on children’s multicultural books to investigate whether the review rating scores would indicate readers’ minds in discussing multicultural books for children. This study discovered that the review rating scores did not represent the accurate indicators of readers’ minds in discussing books. Despite high average ratings of some books, a majority of reader reviews were not found to have positive sentiment polarity. For the determination of whether there is a statistical association between sentiment values and the Goodreads reader book rating scores, this study will perform a correlation analysis. Furthermore, this study will conduct content analysis to investigate whether sentiments include significant factors which would affect the selection of multicultural books for children. This study has an implication of increasing awareness of the ethnic or racial minority themes in children’s literature by uncovering readers’ minds in discussing children’s multicultural books.

REFERENCES

Adam, H. (2021). When Authenticity Goes Missing: How Monocultural Children’s Literature Is Silencing the Voices and Contributing to Invisibility of Children from Minority Backgrounds. *Education Sciences*, 11(1), 32.

- American Library Association Council (2017) Equity, diversity, inclusion: An interpretation of the Library Bill of Rights. Retrieved from (<http://www.ala.org/advocacy/intfreedom/librarybill/interpretations/EDI>).
- Colby, S.A., & Lyon, A.F. (2004). Heightening awareness about the importance of using multicultural literature. *Multicultural Education*, 11(3), 24–28.
- Driscoll, B. & Sedo, D. R. (2019). Faraway, So Close: Seeing the Intimacy in Goodreads Reviews. *Qualitative Inquiry*, 25(3), 248-259.
- Hartley, J. (2018). Some observations on the current state of book reviewing in the social sciences. *Learned Publishing*, 31(2), 169-171.
- Huang, A. H., Yen, D. C., & Zhang, X. (2008). Exploring the potential effects of emoticons. *Information & Management*, 45(7), 466–473.
- Jenkins, P. (1996). Faculty priorities: where does material selection stand? *Collection Building*, 15(1), 19-20.
- Jenkins, P. (1999). Book reviews and faculty book selection. *Collection Building*, 18(1), 4-5.
- Liu, B. (2010). Sentiment analysis and subjectivity. *Handbook of Natural Language Processing* 2:627-666.

A Plug and Play Approach to Vocabulary Mapping and Document Annotation

Clunis, Julaine

University of Wisconsin-Milwaukee, United States | jclunis@uwm.edu

ABSTRACT

This study investigates the utility of open-source data analytics, reporting, and integration tools for mapping knowledge organization systems (KOS) dedicated to COVID-19. Using various natural language processing and machine learning methods, a workflow was created with the KNIME Analytics Platform for term mapping and document annotation tasks using string-based, sense-based, and rule-based algorithmic methods. Results suggest strong support for mapping vocabularies within similar domains and appropriate tagging of unstructured data with concepts from an integrated dictionary created from mapped terms. This study demonstrates the use of a shareable, easily adaptable workflow for KOS mapping and annotating documents focused on COVID-19 research. These mapped terms were then used for annotation of clinical trials. The results suggest that researchers could select and align vocabularies of interest and then use the results to annotate documents of interest.

KEYWORDS

Ontologies, mapping, document annotation, COVID-19, natural language processing, machine learning

INTRODUCTION

The number of terms needed to describe health-related concepts is estimated to be about 45 million (ISO, 2018). These terms are represented in clinical coding schemes, a crucial element of the infrastructure needed for enabling the proper functioning of healthcare systems, facilitating data-driven research discoveries (Schriml et al., 2020), achieving meaningful and accurate exchange and use of information, enriching knowledge, and facilitating improved information analysis (Arvanitis, 2014; Zeng et al., 2020). Sharing and assessing accurate and detailed clinical data is a critical challenge in dealing with the pandemic. In response, new terms have been added to coding schemes, and new specialized vocabularies and ontologies have been created. A critical principle of ontologies is the idea of reuse, which can mean either adapting concepts from another ontology or merging different ontologies into a unified whole. Vocabularies such as the Coronavirus Infectious Disease Ontology (CIDO), COVOC Coronavirus Vocabulary (COVOC), and COVID-19 Ontology (COVID-19) have either not yet been mapped to each other or have limited mappings to other coding schemes. Mapping is critical since these vocabularies contain concepts crucial for understanding COVID-19 and can provide clinical insights through annotating the thousands of clinical trial documentation containing unstructured elements. Semantic annotation of unstructured data with terminologies and ontologies requires a "common, uniform and comprehensive approach" to clinical knowledge representation, thus enhancing discovery, interpretation, and reuse (Tchechmedjiev et al., 2018; de Quiros et al., 2018). However, those currently used for this task are not optimized to tag COVID-19-specific terms. Aggregating specialized terms for annotation of clinical trials will support continued research discoveries. Currently, mapping and annotation require specialized knowledge, manual effort, or complex programming skill, leaving room to explore whether alternate methods might be useful. Therefore, the research addressed the following questions, 1) How can an Extract Transform Load (ETL) workflow tool support the clinical coding scheme mapping task? and 2) How does the mapping output of the novel workflow support and affect annotation of clinical trials in COVID-19 research?

METHODS

The KNIME Analytics Platform was used to create a workflow that leveraged NLP and machine learning techniques to map terms and annotate clinical trial data. Three ontologies with a combined total of 10691 terms were selected and mapped to each other using string-based matching algorithms that have demonstrated efficiency for locating occurrences of terms and specific patterns of text and matching and clustering entity names (Aho & Corasick, 1975; Cohen et al., 2003; Monge & Elkan, 1997); sense-based algorithms that function by making matches based on relations (Giunchiglia et al., 2004); and rule-based methods that determine the semantic type of matches. These were implemented in KNIME, as shown in Figure 1. Mappings between RDFS, OWL, and SKOS ontological classes, properties, concepts from concept schemes, or transitive super properties were identified. The integrated set of concepts was used to create a dictionary for the NLP pipeline to annotate 575 clinical trials on COVID-19 from www.clinicaltrials.gov. Precision, recall, and F-measure was calculated against a gold standard set of terms. The gold standard comprises existing mapped terms from BioPortal based on matching URIs and UMLS unique concept identifiers or generated with the NCBO lexical mapping (LOOM) algorithm. These existing mappings between the three tested vocabularies were extracted and used to compare and evaluate the mapping accuracy of the workflow output. Additionally, the Logical Observation Identifiers Names and Codes (LOINC) terminology was used as a control to test the mapping results.

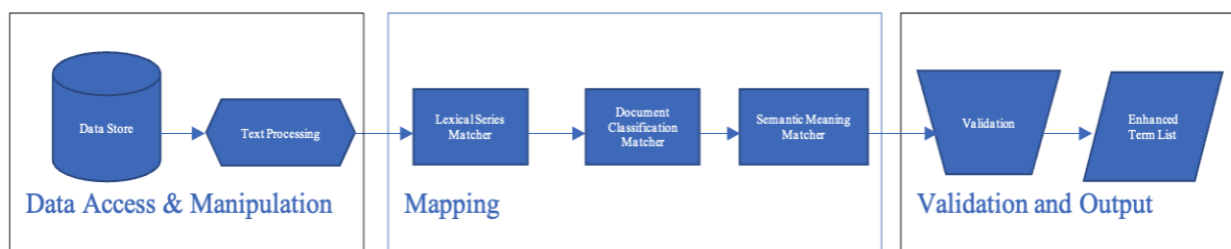


Figure 1. A flowchart illustrating implemented processes in workflow segments

RESULTS

The lexical matchers identified 602 of the 666 mappings shown in Figure 2A for CIDO/COVID19 ontology and identified an additional 294 matches that were not present in the Gold Standard set of terms. Between CIDO and LOINC, the matcher identified 747 mappings of those 426 were correctly identified when compared with the gold standard mapping. There were an additional 252 matches found that should have been present in the Gold Standard but were not. Finally, between the COVID-19 ontology and LOINC, the matcher identified 398 mappings. There were 248 correctly identified terms from that set; however, there were 145 mappings found that are not in the Gold Standard. Performance scores are reported in Figure 2B. Precision is high in most cases, even though recall is low in the tests with LOINC. This can occur when there is a class imbalance between the datasets.

Additionally, this may be because additional data transformation is needed. Accuracy results for the semantic similarity matcher, which uses cosine similarity (see Figure 2C), are moderate and calculated as correct mappings divided by found mappings. Only accuracy is reported since results suggest more heuristic rule-based methods would need to be included in the workflow to identify mappings with lower scores that are related but not the same. Within KNIME, the StanfordNLP NE scorer node was leveraged to validate the clinical trial annotation workflow, achieving 99% precision and 99% recall, as shown in Figure 2D, indicating that the annotation workflow identifies more relevant results than irrelevant results and that most of the dictionary terms are identified.

A

Mappings	SAME URI	LOOM	TOTAL
CIDO / COVID-19	587	79	666
CIDO/LOINC	0	871	871
COVID-19/LOINC	0	489	489

B

	Precision	Recall	TOTAL	F-measure
CIDO ↔ COVID-19	0.979	0.926		0.961
CIDO ↔ LOINC	0.861	0.477		0.681
COVID-19 ↔ LOINC	0.996	0.511		0.767

C

	Found Mappings	Correct Mappings	Accuracy
CIDO ↔ COVID-19	314	206	0.656
COVOC ↔ CIDO	369	291	0.789
COVOC ↔ COVID-19	19	8	0.421

D

Precision	Recall	F1	TP	FP	FN
0.994	0.994	0.994	711	4	4

Figure 2. A) Mappings used for validation. B) Performance Measures from Lexical Matchers. C) Semantic Similarity Mappings Accuracy Scores. D) NLP Model Scorer Results.

CONCLUSION

This study demonstrates the utility of performing mapping and annotation tasks with Open-Source Data Analytics ETL tools requiring less cognitive load and reduced complexity and facilitating easy loading and analysis of datasets, data cleaning and transformation, reductions in operating, product, personnel, and project-related costs, and insights into community-based development, that anyone, expert and non-expert alike, can use. Additionally, it supports assessing and improving data quality through specific Functional, Impactful, and Transformable Metrics for vocabulary transformation through mapping (Zeng & Clunis, 2020). Moreover, the mapping output directly innovatively supports the semantic enrichment of words by providing a list of tailored terms that can be used to train the annotation model to identify the desired entity types within the unstructured text. The workflow can easily refine results, connect annotation to mapping tasks, and can easily be extended to domains beyond those tested in this study. The annotation workflow output demonstrated that the use of vocabulary terms enriched within the workflow with mappings from COVID-19-specific vocabularies offers the ability to provide rich indexing of clinical data for researchers to use or for downstream use in applications. Future development should include methods that use external knowledge bases or unsupervised representation learning. Another direction to explore is building rules to automatically determine narrower and broader matches and refine the annotation workflow.

REFERENCES

- Aho, A. V., & Corasick, M. J. (1975). Efficient string matching: An aid to bibliographic search. *Communications of the ACM*, 18(6), 333–340. <https://doi.org/10/crw9b4>
- Arvanitis, T. (2014). Semantic interoperability in healthcare. In J. Mantas, M. Househ, & A. Hasman (Eds.), *Integrating Information Technology and Management for Quality of Care* (p. 5). IOS Press.
- Cohen, W. W., Ravikumar, P., & Fienberg, S. E. (2003). A Comparison of string distance metrics for name-matching tasks. In *IIWeb*, 2003, 73–78.
- De Quiros, F. G. B., Otero, C., & Luna, D. (2018). Terminology services: Standard terminologies to control health vocabulary: Experience at the Hospital Italiano de Buenos Aires. *Yearbook of Medical Informatics*, 27(01), 227–233. <https://doi.org/10.1055/s-0038-1641200>
- Giunchiglia, F., Shvaiko, P., & Yatskevich, M. (2004). S-Match: An algorithm and an implementation of semantic matching. In C. J. Bussler, J. Davies, D. Fensel, & R. Studer (Eds.), *The Semantic Web: Research and Applications* (pp. 61–75). Springer. <https://doi.org/10/b4t35k>
- ISO. (2018). ISO 17117-1:2018(en), Health informatics—terminological resources—part 1: characteristics (Standard ISO 17117-1:2018(en)). International Standards Organization. <https://www.iso.org/obp/ui/#iso:std:iso:17117:-1:ed-1:v1:en>
- Monge, A., & Elkan, C. (1997). An efficient domain-independent algorithm for detecting approximately duplicate database records.
- Schriml, L. M., Chuvochina, M., Davies, N., Eloë-Fadrosh, E. A., Finn, R. D., Hugenholtz, P., Hunter, C. I., Hurwitz, B. L., Kyrpides, N. C., Meyer, F., Mizrahi, I. K., Sansone, S.-A., Sutton, G., Tighe, S., & Walls, R. (2020). The COVID-19 pandemic reveals the peril of ignoring metadata standards. *Scientific Data*, 7(1), 188. <https://doi.org/10/gg3knq>
- Tchechmedjiev, A., Abdaoui, A., Emonet, V., Melzi, S., Jonnagaddala, J., & Jonquet, C. (2018). Enhanced functionalities for annotating and indexing clinical text with the NCBO Annotator+. *Bioinformatics*, 34(11), 1962–1965. <https://doi.org/10/gdk4vz>
- Zeng, M. L., & Clunis, J. (2020). FAIR + FIT: Guiding principles and functional metrics for linked open data (LOD) KOS products. *Journal of Data and Information Science*, 5(1). <https://doi.org/10.2478/jdis-2020-0008>
- Zeng, M. L., Hong, Y., Clunis, J., He, S., & Coladangelo, L. P. (2020). Implications of knowledge organization systems for health information exchange and communication during the COVID-19 pandemic. *Data and Information Management*, 4(3), 148–170. <https://doi.org/10.2478/dim-2020-0009>

Screen Media Modeling & Mentoring: Parent's Views

Cox, Evelyn D. University of Oklahoma, United States | ecox@ou.edu
Abbas, June M. University of Oklahoma, United States | jmabbas@ou.edu
Munyao, Moses M. University of Oklahoma, United States | Moses.Muendo.Munyao-1@ou.edu
Agosto, Denise E. Drexel University, United States | dea22@drexel.edu

ABSTRACT

Screen media modeling and mentoring is an important factor of family media practices and for tech readiness of digital youth. The authors conducted a qualitative study to examine the media practices within families with children ages 5-11. Semi structured interviews with 51 parents revealed that parents and other significant people, such as grandparents, siblings, friends, and community members, mentor or model media practices with children but practices vary between families and within contexts. Further, there are direct and sustained impacts on children's screen media behaviors when parents actively participate in mentoring or modeling of screen media with their children. Findings also shed light on other influencing factors of screen media engagement, such as modeling styles, parent's prior experience and comfort with technology, personal backgrounds, and concerns about their own use of screen media. Findings are useful as information professionals engage with children when navigating the post-Covid digital landscape.

KEYWORDS

screen media, media advocacy, children's media use, tech-readiness, post Covid-19, navigating digital landscape, medial modeling, media mentoring

INTRODUCTION

The continued technological innovations have resulted in the proliferation of devices that have been integrated into our everyday lives whether at school, work, or home environments. Despite the benefits in the integration of screen media in different aspects of our lives, screen media poses dilemmas that do not have simple solutions. Not only is there a question on the time spent using screen media (AAP, 2016), there are concerns about the social impact among children through the exposure to inappropriate content (including violence and language), overstimulating a child, a child losing their creativity (Wartella and Jennings, 2000), cyberbullying, online privacy, and issues around tech readiness. Additionally, the onset of the Covid-19 pandemic has put an increased burden on parents as gatekeepers of what media their children were consuming, thus having a greater influence on their children's screen media use. Studies have demonstrated that parents play a vital role in modelling healthy behaviors as well as influencing child engagement in active play and other screen time activities (Schoeppe, et al., 2016; Trost & Loprinzi 2011). These findings indicate that active parent participation in the lives of children can yield positive results. The research reported in this paper sought to provide a more nuanced understanding of family media practices, decision-making around digital media use in individual families, and other important factors around children's screen media activities.

LITERATURE REVIEW

According to Haines, Campbell, & Donohue (2016), media mentorship involves "supporting the literacy, information and media needs of children, teens and their families." (p. 4). Studies conducted found that many parents report using media technology with their children but this "joint media engagement" decreases in children ages six or older (Connell, Lauricella & Wartella 2015; Takeuchi and Stevens, 2011). There is also the question of how media modeling and mentoring affects "tech readiness" of our digital youth. This "technology-readiness construct refers to people's propensity to embrace and use new technologies for accomplishing goals in home life and at work." (Parasuraman, 2000, p. 308). Heitner (2016) addresses the need for parents to act more as active mentors who instead of monitoring a child's every move, act as guides helping children make good decisions and foster their creativity. Heitner argues that with the aid of parents who help and mentor children to navigate the challenges, technology has amazing potential to empower both parents and children to traverse this new realm (2016). Another influencing factor to modeling and mentoring are parenting styles. Factors studied by Clark (2013) when discussing parenting styles are socioeconomics, interpersonal family relationships, ethical considerations, individual expectations, and cultural implications which all factor into these "styles" and how they are mediated with technology and modeling (p. 133-185). Also, while parents, teachers and librarians have been widely recognized as media mentors, there are other people beyond the home and school who have the necessary knowledge and skills to guide children's screen media use. Family members such as older siblings, cousins, and grandparents and children's friends provide mentoring or modeling of screen media use (Amaro, et al., 2016; Ballagas, et al., 2013). Online communities in which children participate may also be an influencing factor on children and may model norms or behaviors within the online community (Spencer, 2014).

85th Annual Meeting of the Association for Information Science & Technology | Oct. 29 – Nov. 1, 2022 | Pittsburgh, PA. Author(s) retain copyright, but ASIS&T receives an exclusive publication license.

RESEARCH DESIGN

This research project (Navigating Screens) sought to provide a fuller understanding of family media practices and decision-making around digital media use in individual families, targeting parents of children ages 5-11. This age range was chosen because it is when children are expected to be using technologies for learning as well as for entertainment and when they are developing more independence in online environments. This paper focuses on the preliminary research findings from the parent interviews in relation to parents' perceptions of their modeling and mentorship practices and those of other individuals in the family's community circle. Four research questions guided this portion of the research:

1. What are common screen media practices in households with children ages 5-11?
2. When or how are children's screen media practices guided by parental modeling and mentoring?
3. What are other resources and influencing factors guiding modeling and mentoring of children's screen media use?
4. What are justifications and concerns around modeling and mentoring regarding children's screen media use (for tech readiness)?

Sampling and Data Analysis

Semi structured interviews with 51 parents, representative of populations from the Midwest, Eastern, and Central states in the United States were conducted. Parents were recruited in public libraries, other community events, and using snowball sampling. The sample included 39 females and 12 males from 45 urban and 6 rural families from a range of ethnicities and socio-economic status. Inductive thematic analysis was used for data analysis as "a method for identifying, analyzing, and reporting patterns (themes) within data," (Braun and Clarke, 2006). The code development process involved multiple iterative readings of the data and revisions to the codes. The interview transcripts were uploaded into Dedoose and then coded, reviewed, and recoded by the researchers. This paper focuses on Code # 7, *Modeling and Mentorship* and subcodes: a) Tension around parents' own mentoring and modeling practices; b) Modeling context (e.g., who models, devices modeled, activity modeled, and c) Reasons for / for not modeling). The authors conducted further analysis by relating the themes of each excerpt to the models of mentoring and modeling as identified in the research literature. Several subthemes were identified: 1) Active participation in modeling/mentoring; 2) Influencing factors such as modeling styles, roles, or responsibilities (e.g., to teach life lessons or modeling to check content), and technology use background or comfort of parent(s); 3) Mentoring/modeling by people other than parents; 4) Parental concerns when modeling/mentoring; and 5) Control measures that aid when modeling/mentoring.

Findings and Discussion

Findings related to questions one and two, indicated that active participation in modeling and mentoring was a recurring central subtheme. Parents play an active role in screen media experiences of their child(ren), having direct and sustained impact on children's behavior, such as modelling healthy behaviors and influencing child engagement in active play as well as screen time activities, as also found in Schoeppe, et al., (2016); Trost & Loprinzi, (2011). Additionally, analysis of this subtheme demonstrated purposeful and positive engagement that benefits both the child and the parent involved. This finding supports the research by Reiser, et al., (1984); Valkenburg et al., (1999); and AAP, (2016) that children learn more when parents are involved with the co-viewing process. Regarding question three and the subtheme related to influencing factors to modeling and mentoring, respondents indicated that differing modeling styles was a factor for some parents or caregivers who were driven by personality or upbringing influences when modeling and mentoring. This is in line with the literature around parenting styles in the work of Baumrind (1966; 1971) and later expanded by Maccoby and Martin (1983) and Trost and Loprinzi (2011). The findings also indicated that professionals and other family members sometimes act as "loco parentis," taking on some parental responsibilities when interacting with children in formal and informal settings. Regarding question four, the justification or concern around modeling and mentoring, the findings are in line with Nikken & de Hann (2015) who also outlined concerns when modelling/mentoring such as time spent on devices, exposure to inappropriate content, lack of parental screen media awareness, and peer pressure to access certain devices. Also related to question four and confirming the research by Ying et al., (2015), a reoccurring subtheme around control measures that aid in modelling/mentoring, suggests that measures are not standard, and vary from parent to parent.

CONCLUSION

This study provided an understanding of the different approaches and influencing factors utilized in modelling and mentoring children's screen media behaviors. While the findings indicate that active participation by parents results in positive screen media habits among children, media mentorship should be viewed as a collaborative effort that not only involves parents but other individuals (e.g., relatives, librarians, and teachers). Media is prevalent in the lives of children and therefore parents will continue to have concerns about the short and long-term impact of their children's screen media practices like exposure to inappropriate content; however, the findings demonstrate that active participation during mentoring has positive implications on the tech readiness of our digital youth.

ACKNOWLEDGMENTS

This project was made possible in part by the Institute of Museum and Library Services, IMLS grant #LG-96-17-0220-17.

REFERENCES

- Amaro, A. C., Oliveira, L., & Veloso, A. I. (2016). Let's build our family tree: Grandparents and grandchildren using tablets together. *Procedia Computer Science*, 100, 619-625.
- Ballagas, R.T., Dugan, T.E., Revelle, G., Mori, K., Sandberg, M., Go, J., Reardon, E., & Spasojevic, M. (2013). Electric agents: fostering sibling joint media engagement through interactive television and augmented reality. *Proceedings of the 2013 conference on Computer supported cooperative work*.
- Baumrind, D. (1966). Effects of authoritative parental control on child behavior. *Child Development*, 37(4), 887-907.
- Baumrind, D. (1971). Current patterns of parental authority. *Developmental Psychology*, 4(1), 1-103.
- Braun, V. and Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.
- Clark, L. (2013). *The parent app: Understanding families in the digital age*. Oxford University Press.
- Connell, S., Lauricella A. R., & Wartella E. (2015). Parent co-use of media technology with their young children in the USA. *Journal of Children and Media*, 9(1), 1-17.
- Haines, C., Campbell, C., & ALSC, (2016). *Becoming a media mentor: A guide for working with children and families*. American Library Association.
- Heitner, D. (2016). *Screenwise: Helping kids thrive (and survive) in their digital world*. HPB Diamond.
- Maccoby, E., & Martin, B. (1983). Socialization in the context of the family: Parent-child interaction. In E.M. Hetherington (Eds.), *Handbook of child psychology*, (Vol. 4, pp. 1-101). Wiley.
- Nikken, P., & de Haan, J. (2015). Guiding young children's internet use at home: Problems that parents experience in their parental mediation and the need for parenting support. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 9(1), 44-57.
- Parasuraman, A. (2000). Technology teadiness index (Tri): A multiple-item scale to measure readiness to embrace new technologies. *Journal of Service Research*, 2(4), 307-320.
- Reiser, R. A., Tessmer M. A., & Phelps P. C. (1984). Adult-child interaction in children's learning from Sesame Street. *Education and Communication Technology Journal*, 32, 217-223.
- Rideout, V. (2017). *The Common Sense Census: Media use by kids age zero to eight*. San Francisco, CA: Common Sense Media.
- Schoeppe, S., Rebar, A., Short, C., Alley, S., Lippevelde, W., & Vandelanotte, C. (2016). How is adults' screen time behaviour influencing their views on screen time restrictions for children? A cross-sectional study. *BMC Public Health*, 16(209), 1-5.
- Spencer, D. C. (2014). From many masters to many students: YouTube, Brazilian Jiu Jitsu, and communities of practice. *Journalism, Media and Cultural Studies*, 5, 1-12.
- Takeuchi, L., & Stevens, R. (2011). The new co-viewing: Designing for learning through joint media engagement. New York: The Joan Ganz Cooney Center at Sesame Workshop. Retrieved from www.joanganzcooneycenter.org.
- Trost, S., & Loprinzi, P. (2011). Parental influences on physical activity behavior in children and adolescents: A brief review. *American Journal of Lifestyle Medicine*, 5(2), 171-181.
- Valkenburg, Patti M., Krcmar, Marina, Peeters, Allerd L., & Marseille, Nies M. (1999). Developing a scale to assess three styles of television mediation: "Instructive mediation," "restrictive mediation," and "social coviewing". *Journal of Broadcasting & Electronic Media*, 43(1), 52-66.
- Wartella, E. A., & Jennings, N. (2000). Children and computers: New technology - Old concerns. *The Future of Children*, 10(2), 31-43.
- Ying, L., Ma, F., Huang, H., Guo, X., Chen, C., & Xu, F. (2015). Parental monitoring, parent-adolescent communication, and adolescents' trust in their parents in China. *PLoS ONE*, 10(8), E0134730.

The Incessancy of #gobacktothekitchen and Responses to Normalized Online Misogyny

Daly, Diana
Smith, Maddie Raeann
Bao, Duo

University of Arizona School of Information, USA | didaly@arizona.edu
Undergraduate student, University of Arizona School of Information, USA
Graduate student, University of Arizona School of Information, USA

ABSTRACT

This research arose from the iVoices project collection of student technology experiences guiding research. In response to students being told to "go back to the kitchen" while gaming and reading as "female", our team analyzed TikTok for videos hashtagged #gobacktothekitchen and #backtothekitchen across a one-year period. We also performed deeper analysis on comeback appeals eliciting or offering suggestions of responses to "Go back to the kitchen" and related misogyny and their responses. We found videos were typically created by "girl gamers" toward whom "back to the kitchen" misogyny had been directed, and who tagged them to assign networked meanings to their experiences, encapsulate their struggles for broad publics, and find validation with users sharing similar experiences. A salient theme in comeback appeal posts was performing positions of power to gain leverage over aggressors, while comments frequently offered support from other "girl gamers" and reinforcement of misogynistic stereotypes by male-identified aggressors.

KEYWORDS

misogyny; media production; online games; video games; social media

INTRODUCTION

This pilot inquiry arose from a three-year project called iVoices Media Lab, in which faculty lead a team of students in collecting and producing stories based on students' social media experiences, to forge student-directed paths toward teaching, research, and advocacy around new media. The path of this research began with a 2021 iVoices story-turned-podcast episode by a student named Kierstin, a Rainbow Six Siege gamer whose encounters with misogyny from fellow players—exemplified by the phrase "Go back to the kitchen" incessantly directed at her and at other gamers who read as "female"—lowered her expectations around interpersonal in-game communication, enabling an unhealthy personal relationship with a troubled gamer. (Social Media & Ourselves, 2021.) Kiersten's story echoes those of other young women. "That's just how online is," according to one of numerous young women threatened and harassed by the Uvalde mass shooter online before he killed 21 people. Rather than serving as a crucial red flag toward violent intention, mistreatment of girls and women is normalized in popular online spaces.

SECTIONS

Background

Online misogyny is persistent, now new. Unchecked vitriol against those identified as girls and women online has been the focus of academic research and popular media, including focus on the "e-bile" often accompanying "Go back to the kitchen" in online harassment documented by Jane (2016), and dangers such communication poses for national security (Hunter & Jouenne, 2021). Misogynistic discourse on social networking sites has been found to present significant threats to democracy including the exclusion of women from public participation (Barker & Jurasz, 2019). Hatred for women and feminism is a core value of networked publics spreading disinformation and propaganda (Marwick & Lewis, 2017). Yet despite its many troubling associations, misogyny in online spaces has historically been treated as a personal matter rather than a concern of government or threat to public security (Ging & Siapera, 2018). When we learned Kiersten's story, we recognized that "Go back to the kitchen" was emblematic of systems of abuse that young women face alone in male dominated spaces. This raised a new question: What strategies do young women deploy to remain in these spaces, even when challenged with repeated directives of exclusion?

Methods

Our team of researchers including Dr. Diana Daly, undergraduate student Maddie Rae Smith, and graduate student Duo Bao searched TikTok for videos hashtagged #gobacktothekitchen and #backtothekitchen across a one-year period, from July 25th 2021 to July 24th, 2022. We worked iteratively, first analyzing how many videos studied referenced misogyny, and then analyzing the most prominent type of content in these videos as a pilot subcategory. We found 83 Tiktok videos tagged #gobacktokitchen or #backtothekitchen, and designated the 65 of these (78%) with misogyny-related content as our initial corpus. We used data visualization in the form of word clouds to analyze co-occurring hashtags in the corpus of 65 misogyny-related videos for patterns.

We also selected one salient trend of the 65 misogyny-related video posts to examine more closely through ten examples. This trend was what we called comeback appeals: users appearing or identifying as young women

85th Annual Meeting of the Association for Information Science & Technology | Oct. 29 – Nov. 1, 2022 | Pittsburgh, PA. Author(s) retain copyright, but ASIS&T receives an exclusive publication license.

gamers, eliciting from or offering to their viewing publics suggestions of responses for when they encounter "Go back to the kitchen" and related misogyny. Following Shifman (2013), we coded this pilot corpus through the mimetic dimensions of Content (Asking for comeback suggestions, or Demonstrating comeback); Form (Gameplay demo; Lip-synching; Talking head; Enhanced onscreen messaging [such as talking through a sim avatar]; or Other); and Stance (Sincere/bothered; Humorous; Angry; Triumphant/not bothered; and Other). To get a sense of conversations branching from these posts, we also coded the two most liked comments per video and the one most liked reply to each of these comments as Offering comeback; Insulting or countercomeback; Supportive; or Other / Unclear.

Discussion

Videos in our corpus were typically created by gamers who read as young women and toward whom "back to the kitchen" had been directed. Hashtags they used included appeals for visibility in general (#viral), on TikTok (#fyp), and on gaming platforms (such as #callofduty, #valorant, #r6siege); identity pronouncements to signify belonging (#girlgamer and #femalestreamer); calling out systematic oppression (#feminism, #toxic) and defying it (#imjustgood, #staymad); and echoes of misogyny (#omgitsagirl and sequential tags like #thisgirl #isntgood and #not #her.) Use of these tags highlights creators' intentions to broadly share and assign networked meanings to their experiences, encapsulate their struggles for broad publics, and validate them through aggregation with users sharing similar experiences.

The comeback appeals content we analyzed had 5,222,900, 965,527 likes, and 17,685 comments. Our analysis of comeback appeals discourse found patterns and distinctive themes in the creative strategies those identifying as girls and women deploy to counter misogyny in their digital lives. Patterns included humor as the most prominent video stance, presented to elicit witty comebacks and to counterstrike their male-identified attackers. One common theme in these conversations was performing positions of power to gain leverage over an aggressor, through sexuality or through therapy positioning by diagnosing aggressors as having "daddy issues;" for example, in one video the creator uses audio from a different woman player recorded during gameplay in which she comes back with, "I'm going to fuck your dad and give him a child he actually loves;" this elicits supportive comments like "smooth and sharp as a brand new Knife Sis," empathy ("that's why I turn off the chat"), and critiques for not giving enough credit to the author of the sound. Power positioning also included interpretation of male aggression as a signifier of homosexuality, prompting both comments of support and condemnation of homophobia. We also found themes in counterattacks from what appeared to be male-identified users, including reinforcement of misogynistic stereotypes such as that women are not funny, or "tech bro mansplaining" that creators are using wrong equipment or gaming incorrectly or poorly.

CONCLUSION

This project will continue with full analysis of a broader corpus, production of iVoices media around outcomes, and direction of advocacy toward more inclusivity in online gaming. The persistence of a phrase as anachronistic as "Go back to the kitchen" today shows one the result when decades of misogyny are left unchallenged by those with the power to curtail it, including governments and the leadership and development teams of social networking sites and gaming platforms. We found that those facing harassment online appropriated the affordances of the TikTok platform for quickly posting, upvoting, and responding, to gather resources to weather anticipated future misogyny, and to build community around their continued struggles to participate safely and equally online. In the face of networked harassment, these gamers perform positions of power to gather networked resistance, to stay in the game and its associated social arenas.

ACKNOWLEDGMENTS

This research was supported by the University of Arizona Center for University Education and Scholarship and Vertically Integrated Projects programs.

REFERENCES

- Barr, M., & Copeland-Stewart, A. (2022). Playing video games during the COVID-19 pandemic and effects on players' well-being. *Games and Culture*, 17(1), 122-139.
- Foster-Frau, S. et al (May 28, 2022). "Before massacre, Uvalde gunman frequently threatened teen girls online." *The Washington Post*.
- Ging, D., & Siapera, E. (2018). Special issue on online misogyny. *Feminist media studies*, 18(4), 515-524.
- Hunter, K., & Jouenne, E. (2021). All Women Belong in the Kitchen, and Other Dangerous Tropes: Online Misogyny as a National Security Threat. *Journal of Advanced Military Studies* 12(1), 57-85.
- Jane, E. A. (2014). 'Back to the kitchen, cunt': Speaking the unspeakable about online misogyny. *Continuum*, 28(4), 558-570.
- Marwick, A. E., & Caplan, R. (2018). Drinking male tears: Language, the manosphere, and networked harassment. *Feminist Media Studies*, 18(4), 543-559.
- Marwick, A., & Lewis, R. (2017) Media Manipulation and Disinformation Online. *Data & Society*.

- Perrin, Andrew. (2017). Five facts about Americans and video games. Pew Research Center,
- Shifman, L. (2013). Memes in a digital world: Reconciling with a conceptual troublemaker. *Journal of computer-mediated communication*, 18(3), 362-377.
- Social Media & Ourselves* (2021). Girl Meets Chud. December 1st, 2021. iVoices Media Lab.
<https://ivoices.ischool.arizona.edu/social-media-ourselves>.
- Vizcaíno-Verdú, A., & Aguaded, I. (2022). # ThisIsMeChallenge and Music for Empowerment of Marginalized Groups on TikTok. *Media and Communication*, 10(1), 157-172.

Applying Positive Psychology's Subjective Well-Being to Online Interactions

Delellis, Nicole, S. University of Western Ontario, Canada | ndelelli@uwo.ca
Kelly, Dominique University of Western Ontario, Canada | dkelly48@uwo.ca
Liu, Yifan University of Western Ontario, Canada | yliu4369@uwo.ca
Mayhew, Alex University of Western Ontario, Canada | amayhew@uwo.ca
Chen, Yimin University of Western Ontario, Canada | ychen582@uwo.ca
Cornwell, Sarah, E. University of Western Ontario, Canada | scornwel@uwo.ca
Rubin, Victoria, L. University of Western Ontario, Canada | vrubin@uwo.ca

ABSTRACT

This paper outlines the complexity of the psychological construct of individuals' subjective well-being (SWB) and argues for the importance of examining behaviours and linguistic expression of individuals online social interactions in relation to self-reported SWB. This paper calls for a systematic review of the psychology research which examines SWB and its association with various character strengths, personality traits, and behaviours. While the Big Five personality traits (OCEAN) have an underlying neuropsychological basis and are considered as universal dimensions of personality along which humans differ one from another, minimal research has attempted to evaluate the relationship between personality traits, SWB, and online interactions.

KEYWORDS

Subjective Well-Being (SWB); Positive Psychology; Personality Traits; Online Interaction Context; Natural Language Processing (NLP)

INTRODUCTION

With the emergence of the COVID-19 pandemic and associated infection rates and decreased human mobility, researchers have noted a global "increase in the prevalence of major depressive disorders and anxiety disorders during 2020" (Santomauro et al., 2021, p. 1706). Individuals with limited human mobility may turn to information and communication technologies (ICTs) to connect with others. However, use of social networking sites (SNS) may exacerbate feelings of depression. Cunningham et al. (2021) found that "depression symptoms were significantly, but weakly, associated with time spent using SNS and intensity of SNS use" (p. 241). Prior work has examined the motivations behind voluntary non-use of social media platforms (Baumer et al., 2013; Grandhi et al., 2019). Overall, the perception that social media erodes personal well-being is a key theme in users' decisions to leave (Grandhi et al., 2019). This work calls for a shift to evaluate how social media users can increase their happiness.

SUBJECTIVE WELL-BEING: WHAT IS IT?

In positive psychology discourse, subjective well-being (SWB) "is the preferred scientific conceptualization of the more colloquial term 'happiness'" (Heintzelman & Tay, 2017, p. 10). SWB is conceptualised as a multifaceted umbrella construct. As derived from Diener's (1984) tripartite model, SWB is "[...] frequently defined as having three inter-related qualities: regular positive affect, elevated life satisfaction, and sporadic negative affect" (Dunn et al., 2009, p. 652). SWB is rooted in individuals' subjective experience. Self-reported measures are traditionally used to measure SWB because of "the internal and experiential nature of affective experiences and satisfaction judgements" (Heintzelman & Tay, 2017, p. 11) used to gauge life satisfaction, positive affect, and negative affect.

Three factors (genetics, circumstantial, and intentional actions) have been proposed as having a long-term influence on individuals' SWB. The genetic factor or 'set point of happiness' has previously been viewed as stable and fixed; however recent research has indicated that the 'set point of happiness' is not as fixed as previously conceived (Lucas et al., 2021, p. 7). Personality psychologists "[...] have reached a consensus that people vary from one another along five basic dimensions: the Big Five traits" (Little, 2017, p. 12). These five dimensions are often spelled out as an acronym, OCEAN: Open (to experiences vs. closed), Conscientious (vs. casual), Extraverted (vs. introverted), Agreeable (vs. disagreeable), and Neurotic (vs. stable). As reviewed in Ozer and Benet-Martínez (2007), "the Big Five have major consequences for how our lives play out" (Little, 2017, p. 12) in education, marriage, health, and work outcomes. Research has revealed that different personality traits across humans correlate to different components and conceptualizations of SWB, and that personality traits may "[...] account for between 32 and 56% of the variance in subjective well-being scores" (Hayes & Joseph, 2003, p. 725).

While genes are a significant factor in SWB, "[...] both life circumstances and intentional activities also play important, and perhaps larger roles" (Heintzelman & Tay, 2017, p. 16). Circumstantial factors include "[...] personal qualities (e.g., age, gender, life history), events (e.g., marriage, unemployment, and retirement), and situations (e.g., nation, location, culture)" (Dunn et al., 2009, pp. 652 - 653). Intentional actions refer to "what people do and how

85th Annual Meeting of the Association for Information Science & Technology | Oct. 29 – Nov. 1, 2022 | Pittsburgh, PA. Author(s) retain copyright, but ASIS&T receives an exclusive publication license.

they think in everyday life” (Dunn et al., 2009, pp. 652 - 653). Interventions pertaining to increasing SWB when engaging in online SNS interactions should aim to change maladaptive intentional actions and ICT design environmental features that decrease SWB. Randomised controlled experiments have shown that in offline environments SWB can be increased “[...] by practising simple positive activities with effort and commitment. Such activities including performing kind acts, expressing gratitude or optimism, and re-experiencing joyful events represent the most promising route to enhanced happiness” (Lyubomirsky et al., 2021, p. 945).

APPLICATION OF POSITIVE PSYCHOLOGY TO ONLINE INTERACTIONS AND ICT DESIGN

Given the complexity of SWB, determining what makes a ‘happier’ online community/interaction could vary across individuals. In online communities, prosocial interactions such as sharing, helping, and co-operating may occur in contexts that, at first glance, may seem likely to lower SWB. Co-destructive communities may provide prosocial behaviours through positive interpersonal interactions (Chen et al., 2022). Self-deprecating statements, for example, have been found to be detrimental to SWB (Steinmayr et al., 2016), but within some internet meme groups, self-deprecating memes are forms of emotional or affective labour where the liking and sharing of memes are ways to transform lived pain into relatable and humorous (and therefore palatable) struggles (Ask & Abidin, 2018, p. 845). By laughing at their own failures, anxieties, and inadequacies, these communities can build surprisingly supportive and cohesive group identities. Research has indicated that different usage of SNS may both increase and decrease SWB. Gerson et al. (2016) found that while using Facebook as a means of social comparison was negatively related to SWB, Facebook intensity was positively associated with SWB. More research is required to uncover if patterns exist between an individual's personality traits, online community engagement, online behaviours, and SWB.

While SWB has been studied in the field of Natural Language Processing (NLP) for over a decade, research using NLP to examine linguistic expression in SNS has drawn minimally from positive psychological discourse. Research of SWB in NLP grew out of sentiment analysis, which aims to accurately identify human emotion in text. As a result, the field’s attempts at SWB have largely skewed towards the facet of emotional well-being, the detection of either positive emotion, negative emotions, or neutrality. Work detecting extreme negative emotions such as suicidality and major depression has been a major subfield with three systematic reviews published since 2020 (Bernert, et al., 2020; Heckler et al., 2022; Ji et al., 2021). Research on positive emotion detection is less common and has applied a more limited number of NLP tools; Luhmann’s review (2017) showed that most studies focussed on emotional well-being assessed at the word level with tools like Linguistic Inquiry and Word Count (LIWC). This finding was re-confirmed in Sametoğlu et al.’s systematic review (2022) which reported most papers used closed word lists or word-level analyses, though more papers are now comparing these data to broader ‘satisfaction with life’ self-report measures. Emotional well-being, though important, can be transient and self-reports are subject to mood fluctuations at the time of the self-report (Pavot & Diener, as cited in Kjell, 2022).

NEXT STEPS

Recent models of SWB underscore the complexity of the construct across a multitude of factors. Seligman’s (2011) PERMA model encapsulates five elements of well-being that enable flourishing – Positive Emotion, Engagement, Relationships, Meaning and Accomplishment. The PERMA model demonstrates that SWB is influenced by a range of factors including feelings of positive emotions, active involvement, close interpersonal relationships, life purpose/sense of belonging, and experiencing mastery. While past research has examined individuals’ behaviour such as gratitude (Deng et al., 2019) or prosocial behaviour (Gherghel et al., 2021) in relation to SWB, a systematic literature review compiling a range of such character strengths/behaviours/personality traits associated with SWB is required to provide a solid conceptual basis for our future research and inform its empirical methodology.

CONCLUSION

The implications of this research for the field of library and information science and technology (LIS&T) lie in its planned deep dive into the psychology of ICT users and their interactions in online communities. While much has been said about information needs of information seekers or sharers (e.g., Bukhari et al., 2020; Savolainen, 2019), few models look at the tri-part modelling of the psychological factors underlying group interactions via ICTs: the impact of users’ personality traits, life circumstances, and intentional activities. What would make online social media users happier, and how do we enact change towards more positive interactions and healthier communities? It is time for researchers to assess if positive psychological findings pertaining to SWB can be translated into individuals’ online behaviours and ICT design, and if these alterations can make users happier.

ACKNOWLEDGMENTS

This work has been funded by the University of Western Ontario Strategic Support for the Canadian Sciences and Humanities Research Council (SSHRC) Success grant, awarded to Dr. Rubin in fall 2021. The grant was entitled "Rewiring for Happiness: Computing for Kinder, Healthier, and Compassionate Life Experiences with Information Commutation Technologies (ICTs)."

REFERENCES

- Ask, K., & Abidin, C. (2018). My life is a mess: Self-deprecating relatability and collective identities in the memification of student issues. *Information, Communication & Society*, 21(6), 834-850. <https://doi.org/10.1080/1369118X.2018.1437204>
- Baumer, E. P. S., Adams, P., Khovanskaya, V. D., Liao, T. C., Smith, M. E., Sosik, V. S., & Williams, K. (2013). Limiting, leaving, and (re)lapsing: An exploration of Facebook non-use practices and experiences. *CHI '13: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3257-3266. <https://dl.acm.org/doi/pdf/10.1145/2470654.2466446>
- Bernert, R. A., Hilberg, A. M., Melia, R., Kim, J. P., Shah, N. H., & Abnoui, F. (2020). Artificial Intelligence and Suicide Prevention: A Systematic Review of Machine Learning Investigations. *International Journal of Environmental Research and Public Health*, 17(16), 5929. <http://dx.doi.org/10.3390/ijerph17165929>
- Bukhari, S., Hamid, S., Ravana, S. D., Kurnia, S., Chang, S., Norman, A. A., & Abdul Ghani, N. (2020). The Use of Facebook by International Students for Information-seeking in Malaysia: A Social Network Analysis. *Libri: International Journal of Libraries & Information Services*, 70(3), 251-268. <https://doi.org/10.1515/libri-2019-0033>
- Chen, Y., Cornwell, S., Delellis, N. S., Kelly, D., Liu, Y., Mayhew, A., & Rubin, V. L. (2022). Engagement for Good or Ill: Comparing Characteristics of Co-Creative and Co-Destructive Online Communities. *The 85th Annual Meeting of the Association for Information Science and Technology (ASIS&T2022): "Information for a Sustainable World: Addressing Society's Grand Challenges"*, October 29 - November 1, 2022, Pittsburgh, PA.
- Cunningham, S., Hudson, C. C., & Harkness, K. (2021). Social Media and Depression Symptoms: a Meta-Analysis. *Journal of Abnormal Child Psychology*, 49(2), 241-253. <https://doi.org/10.1007/s10802-020-00715-7>
- Deng, Y., Xiang, R., Zhu, Y., Li, Y., Yu, S., & Liu, X. (2019). Counting blessings and sharing gratitude in a Chinese prisoner sample: Effects of gratitude-based interventions on subjective well-being and aggression. *The Journal of Positive Psychology*, 14(3), 303-311. <http://dx.doi.org/10.1080/17439760.2018.1460687>
- Diener, E. (1984). Subjective Well-Being. *Psychological Bulletin*, 95(3), 11-58. https://doi.org/10.1007/978-90-481-2350-6_2
- Dunn, D. S., Uswatte, G., & Elliott, T. R. (2009). Happiness, Resilience, and Positive Growth Following Physical Disability: Issues for Understanding, Research, and Therapeutic Intervention. In S. Lopez & C. R. Snyder (Eds.), *The Oxford Handbook of Positive Psychology* (2nd ed). Oxford University Press.
- Gerson, J., Plagnol, A. C., & Corr, P. J. (2016). Subjective well-being and social media use: Do personality traits moderate the impact of social comparison on Facebook? *Computers in Human Behavior*, 63, 813-822. <https://doi.org/10.1016/j.chb.2016.06.023>
- Gherghel, C., Nastas, D., Hashimoto, T., & Takai, J. (2021). The relationship between frequency of performing acts of kindness and subjective well-being: A mediation model in three cultures. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues*, 40(9), 4446-4459. <http://dx.doi.org/10.1007/s12144-019-00391-x>
- Grandhi, S. A., Plotnick, L., & Hiltz, S. R. (2019). Do I stay or do I go? Motivations and decision making in social media non-use and reversion. *Proceedings of the ACM on Human-Computer Interaction*, 3, 1-7. <https://doi.org/10.1145/3361116>
- Hayes, N., & Joseph, S. (2003). Big 5 correlates of three measures of subjective well-being. *Personality and Individual Differences*, 34(4). [https://doi.org/10.1016/S0191-8869\(02\)00057-0](https://doi.org/10.1016/S0191-8869(02)00057-0)
- Heckler, W. F., de Carvalho, J.V., & Barbosa, J. L.V. (2022). Machine learning for suicidal ideation identification: A systematic literature review. *Computers in Human Behavior*, 128 (2022), 107095. <https://doi.org/10.1016/j.chb.2021.107095>.
- Heintzelman, S., & Tay, L. (2017). Subjective well-being: Payoffs of being happy and ways to promote happiness. In *Positive Psychology: Established and Emerging Issues* (pp. 9-28). <https://doi.org/10.4324/9781315106304>
- Ji, S., Pan, X., Li, E., Cambria, G., Long, & Huang, Z. (2021). Suicidal Ideation Detection: A Review of Machine Learning Methods and Applications. *IEEE Transactions on Computational Social Systems*, 8 (1), p. 214-226, Feb. 2021, doi: 10.1109/TCSS.2020.3021467.
- Kjell, O. N. E., Sikström, S., Kjell, K., & Schwartz, H. A. (2022). Natural language analyzed with AI-based transformers predict traditional subjective well-being measures approaching the theoretical upper limits in accuracy. *Scientific Reports*, 12(1), 3918. <https://doi.org/10.1038/s41598-022-07520-w>
- Little, B. R. (2014). Well-doing: Personal projects and the quality of lives. *Theory and Research in Education*. 12(3), 329-346. doi:10.1177/1477878514545847
- Lucas, E., Oishi, S., & Diener, E. (2021). Subjective Well-Being: The Science of Happiness and Life Satisfaction. In *The Oxford Handbook of Positive Psychology*. Oxford University Press.
- Luhmann, M. (2017). Using Big Data to study subjective well-being. *Current Opinion in Behavioral Sciences*, 18, 28-33. <https://doi.org/10.1016/j.cobeha.2017.07.006>
- Lyubomirsky, Ruberton, P. M., & Boehm, J. K. (2021). The Promise of Fostering Greater Happiness. In *The Oxford Handbook of Positive Psychology* (pp. 945-958). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199396511.013.55>
- Ozer, D. J., & Benet-Martínez V. (2006). Personality and the prediction of consequential outcomes. *Annu Rev Psychology*, 57, 401-421. doi: 10.1146/annurev.psych.57.102904.190127.
- Sametoğlu, S., Pelt, D., Eichstaedt, Johannes C., Ungar, L. H., & Bartels, M. (2022). *The Value of Social Media Language for the Assessment of Wellbeing: A Systematic Review and Meta-Analysis* [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/qnx2v>

- Santomauro, D., Mantilla Herrera, A. M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott, D. M., Abbafati, C., Adolph, C., Amlag, J. O., Aravkin, A. Y., Bang-Jensen, B. L., Bertolacci, G. J., Bloom, S. S., Castellano, R., Castro, E., Chakrabarti, S., Chattopadhyay, J., Cogen, R. M., Collins, J. K., ... Deen, A. (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *The Lancet (British Edition)*, 398(10312), 1700–1712. [https://doi.org/10.1016/S0140-6736\(21\)02143-7](https://doi.org/10.1016/S0140-6736(21)02143-7)
- Savolainen, R. (2019). Modeling the interplay of information seeking and information sharing: A conceptual analysis. *Aslib Journal of Information Management*, 71(4), 518–534. <https://doi.org/10.1108/AJIM-10-2018-0266>
- Seligman, M. (2011). *Flourish : a visionary new understanding of happiness and well-being*. Free Press.
- Steinmayr, R., Crede, J., McElvany, N., & Wirthwein, L. (2016). Subjective well-being, test anxiety, academic achievement: Testing for reciprocal effects. *Frontiers in psychology*, 6, 1994. <https://doi.org/10.3389/fpsyg.2015.01994>

Towards Data-driven GIM tools: Two Prototypes

Dinneen, Jesse David

Humboldt-Universität zu Berlin, Germany | jesse.dinneen@hu-berlin.de

Donner, Sascha

evana.ai, Germany | saschadonner76@gmail.com

Bubinger, Helen

Humboldt-Universität zu Berlin, Germany | helen.bubinger@student.hu-berlin.de

Low, Jwen Fai

McGill University, Canada | jwen.low@mail.mcgill.ca

Krtalić, Maja

Victoria University of Wellington, New Zealand | maja.Krtalic@vuw.ac.nz

ABSTRACT

Here we describe two approaches to improve group information management (GIM) and draw on the results of prior works to implement them in software prototypes. The first aids browsing and retrieving from large and unfamiliar collections like shared drives by dynamically reducing and re-organising them. The second supports the transfer and re-use of collections (e.g. to/by successors, descendants, or curators) by integrating novel sorting and annotation features. The prototypes' source code is shared online and screenshots are presented in the accompanying poster.

KEYWORDS

Personal information management, group information management, personal digital archiving, information retrieval

INTRODUCTION

People in the information society are tasked everyday with storing, organising, and refinding an increasing number and variety of digital items in a challenging task known as *personal information management* (PIM; Jones et al., 2017). Such activities are made further challenging when other people are involved in some way – in *group information management* (GIM) – for example when finding items in collections organised by coworkers or when organising a collection that will be later re-used by family (Dinneen & Julien, 2020). To help address the challenges of GIM we offer two prototype interfaces, the ideation and design of which are informed by data collected in prior studies. Both prototypes were developed using the Python programming language and Qt graphics framework. *Their interfaces can be seen on the accompanying poster and their source code is shared online* (links below).

GIMZOOMER

Personal and group-managed collections of digital files are often very large, consisting of tens of thousands of files and folders organised into tree shapes that are deep and wide (Dinneen et al., 2019). Large information structures of all kinds are usually difficult to browse (Julien et al., 2013) and therefore also to organise and retrieve from; for example they require users to review and decide among many labels (e.g. names of folders, genres, nodes) and maintain a sense of their location as they navigate down and across the tree and identify one or more relevant items (Savolainen, 2018). Group-managed structures introduce further challenges, as users often lack the benefit of familiarity that comes from having previously seen, organised, or labelled the folders they are browsing or retrieving from (Bergman et al., 2014). As a result, retrievals require more effort, take longer, and are more prone to failure (Bergman et al., 2019, Julien et al., 2013). In some cases, users may customise the structure to suit their individual retrieval strategy and aid later relevance judgments (de Fremery & Buckland, 2022), but doing so may violate group conventions or decrease the intelligibility of the structure for others users (Berlin et al., 1993; Dourish et al., 1999; Erickson, 2006). Such problems manifest today on shared drives during collaboration, turnover, and onboarding.

One way to address large and unfamiliar collections is to provide flexible views wherein users can see and modify the collection in a way that benefits them and is determined less by other users (Dourish et al., 1999; Ellis & Dix, 2007). A promising start to this approach is to reduce the structure so that the displayed portion and its overall browsing complexity are reduced, as has been done with two information structures that organise published scientific texts: LCSH and MeSH (Julien et al., 2013). There, the extremely uneven distribution of most books into a relatively small number of genres enabled two algorithmic reductions: (a) compressing the structure by moving book-heavy genres (and their siblings) up in the tree, and (b) pruning (i.e. removing) book-light genres. With this approach the browsing complexity of MeSH was reduced by ~80%, and of LCSH by ~50% without precluding access to the majority of the books. Later, a between-subjects experiment confirmed the usefulness of the approach in practice: when using a strongly pruned-and-compressed LCSH tree, participants completed retrieval tasks significantly more accurately and faster, and with no fewer (potentially valuable) interactions with the structure than participants using unmodified or only compressed trees (Dinneen et al., 2018).

Although developed for LCSH and MeSH, the reduction approach may also work well for GIM structures: recent works have established that digital files are often distributed unevenly into folders similarly to the books and genres of LCSH and MeSH (Dinneen et al., 2019), so if a user is retrieving a file, it is likely (*ceteris paribus*) to be located in one of only a few large folders, and rapidly reducing the display to only those folders should enable faster access.

Driven by these observations, we designed GIMZoomer, an interface implementing the compression and pruning algorithms provided by Julien et al. (2013). Specifically, the interface features a slider control between tree views showing (1) an unmodified folder hierarchy and (2) its compressed and pruned version. The slider allows the user to control the strength of the algorithms and thus flexibly and iteratively reduce/increase how many folders are displayed in the second tree view and the maximum downward navigation steps; the *ends* of the slider thus represent extremes of the algorithm strengths, producing no modification (bottom) and the strongest modification (top), wherein only a single folder is shown (the fullest). The user can thus reduce the structure, navigate within it to their desired file or folder, or in more complex cases of information seeking (e.g. berrypicking), continue modifying and exploring the structure. The source code is available at <https://github.com/jddinneen/GIMZoomer>.

COLLECTION ANNOTATION ASSISTANT

We note above that individuals and institutions accumulate digital file collections of extensive scale and variety. For individuals, collections contain a variety of everyday items such as private photos, family plans and finances, professional activities, and information reflecting many other of life's activities and tasks (Dinneen & Julien, 2019; Jones et al., 2017; Krtalić & Ihejirika, 2022). Accumulated over years or a lifetime, these digital assets can have long-term value for family members, and the collections of culturally significant people can also have value for society in the form of cultural heritage (Krtalić et al., 2021). But such collections also pose management challenges that entail various risks of loss (Digital Preservation Coalition, 2021; Krtalić & Dinneen, 2022). When collections are transferred to heirs or cultural heritage repositories (Day & Krtalić, 2021), whether for immediate use or preservation, they are often shared without detailed metadata about the contents or context, and it can thus be very difficult for the new owner or custodian to navigate and grasp the often vast and complex collections (Dinneen & Krtalić, 2020; Society of American Archivists, 2013). There is thus a frequent risk that collections will not be available for current or future generations because parts of the collection are inadvertently deleted, the meaning of the content is not sufficiently clear, too large a portion of the collection is irrelevant (i.e. precludes finding the relevant items), or some portion may be too private to be shared (e.g. insufficient time and labour to filter such items before the collection is shared).

Software developers have acknowledged that such collections or *digital legacies* need to be considered before problems arise. For example, Apple recently released a Legacy Contacts feature where up to five people can be named as managers of one's digital assets after death (Digital Legacy Association, 2016). This has the potential to help collections avoid total loss after death and thus avoid losing their value (in the sense indicated above). There remains, however, no tool available with which to indicate which *portions* of a collection might be relevant to families, curators, etc., nor which parts should not be transferred or seen, and so the practical challenges of identifying, re-using, or excluding relevant portions of a collection remain. While prior works have emphasised the importance of supporting digital legacies (e.g. Dinneen et al., 2016; Jones et al., 2016), no tool has yet been developed to help people make sense of large inherited collections, determine which portions are relevant for specific stakeholders, and ensure that collections can continue to be used in the spirit of the collection creator.

A simple but promising approach to avoiding the problems above is to facilitate annotation of the majority of a collection. We thus designed a Collection Annotation Assistant, a backup-like software where parts of a collection can be marked as relevant or valuable to some context (e.g. to a theme, career, family, institution, or society). In the interface a tree-view presents the folder collection together with tick-boxes to indicate relevancy or, alternatively, exclusion (i.e. unable to be labelled relevant). As collections are often so large that reviewing the relevant portions could be time intensive, we also leverage (a) the uneven collection distribution described above and (b) trends in access times (most folders have not been accessed in the last six months; Dinneen, 2018) to add two related sorting options: by accessible files (i.e. those contained in the folder and sub-folders) and by last date modified. In this way, users should be able to quickly sort the collection and annotate the majority of it in relatively few clicks. A further screen in the interface allows users to note any software required for accessing content stored within files, for example in the case of atypical or obsolete formats. The annotations and notes can then be saved in the interoperable JSON format and loaded later for review or updating. The same interface can thus be used later by the inheritors to get a quick impression of what is in the collection, where the relevant parts are, and what should not be accessed or shared. The source code is available at <https://github.com/jddinneen/collection-annotation-assistant>.

CONCLUSION

Although the prototypes described here should be evaluated (e.g. for their efficacy and usability), demonstrating new approaches to GIM is important step forward because inadequate tools can prevent individuals from acclimating to new work places, finding information needed in daily life (Jones et al., 2017), maintaining and sharing their collections, and preserving them for future generations of family or society (Krtalić et al., 2021).

ACKNOWLEDGMENTS

The authors thank Charles-Antoine Julien for helping to create GIMZoomer and creating the related algorithms.

REFERENCES

- Bergman, O., Israeli, T., & Whittaker, S. (2019). Factors hindering shared files retrieval. *Aslib Journal of Information Management*, 72(1), 130–147.
- Bergman, O., Whittaker, S., & Falk, N. (2014). Shared files: The retrieval perspective. *Journal of the Association for Information Science and Technology*, 65(10), 1949–1963.
- Berlin, L. M., Jeffries, R., O'Day, V. L., Paepcke, A., & Wharton, C. (1993) Where Did You Put It? Issues in the Design and Use of a Group Memory. *Proceedings of ACM INTERCHI'93 Conference on Human Factors in Computing Systems*, 23–30.
- Day, A. S., & Krtalić, M. (2022). “Evidence of Me” Becoming “Evidence of Us”: A Case Study of the Policy, Processes, Donor Relations and Responses of Selected New Zealand GLAM Institutions to Personal Donations of Collections and Artifacts. *Collection Management*, 47(1), 49–73.
- de Fremery, W. & Buckland, M. K. (2022). Context, relevance, and labor. *Journal of the Association for Information Science and Technology*, Early View: <https://doi.org/10.1002/asi.24631>.
- Digital Preservation Coalition. (2021). The 'Bit List' 2021: the global list of digitally endangered species. Retrieved April 12, 2022, from <https://www.dpconline.org/docs/miscellaneous/advocacy/wdpd/2521-bitlist2021/file>
- Digital Legacy Association. (2016). The Digital Legacy Association’s response to Colin and Sue Hehir speaking on BBC Breakfast. Retrieved April 8, 2022, from <https://digitallegacyassociation.org/the-digital-legacy-associations-response-to-colin-and-sue-hehir-speaking-on-bbc-breakfast>
- Dinneen, J. D. (2018). Analysing file management behaviour. [Doctoral dissertation, McGill University]. eScholarship@McGill. <https://escholarship.mcgill.ca/concern/theses/4q77fv027>
- Dinneen, J. D., Asadi, B., Frissen, I., Shu, F., & Julien, C.-A. (2018, March). Improving exploration of topic hierarchies: Comparative testing of simplified library of congress subject heading structures. In *CHIIR '18: Proceedings of the 2018 ACM Conference on Human Information Interaction & Retrieval* (pp. 102–109).
- Dinneen, J. D., & Julien, C.-A. (2019). What's in people's digital file collections? *Proceedings of the Association for Information Science and Technology*, 56(1), 68–77.
- Dinneen, J. D., & Julien, C.-A. (2020). The ubiquitous digital file: A review of file management research. *Journal of the Association for Information Science and Technology*, 71(1), E1–E32.
- Dinneen, J. D., Julien, C.-A., & Frissen, I. (2019, May). The scale and structure of personal file collections. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–12).
- Dinneen, J. D., & Krtalić, M. (2020). E-mail as legacy: managing and preserving e-mail as a collection. *portal: Libraries and the Academy*, 20(3), 413–424.
- Dinneen, J. D., Odoni, F., & Julien, C.-A. (2016). Towards a desirable data collection tool for studying long-term PIM. In *Personal Information Management Workshop at CHI '16: ACM Conference on Human Factors in Computing Systems*. New York: ACM.
- Dourish, P., Lamping, J., & Rodden, T. (1999). Building bridges: customisation and mutual intelligibility in shared category management. In *Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work* (pp. 11–20). ACM.
- Ellis, G., & Dix, A. (2007). A taxonomy of clutter reduction for information visualisation. *IEEE Transactions on Visualization and Computer Graphics*, 13(6), 1216–1223.
- Erickson, T. (2006). From PIM to GIM: personal information management in group contexts. *Communications of the ACM*, 49(1), 74–75.
- Julien, C. A., Tirilly, P., Dinneen, J. D., & Guastavino, C. (2013). Reducing subject tree browsing complexity. *Journal of the American Society for Information Science and Technology*, 64(11), 2201–2223.
- Jones, W., Bellotti, V., Capra, R., Dinneen, J. D., Mark, G., Marshall, C., ... & Van Kleek, M. (2016, May). For richer, for poorer, in sickness or in health... the long-term management of personal information. In *Proceedings of the 2016 ACM CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 3508–3515).
- Jones, W., Dinneen, J. D., Capra, R., Diekema, A. R. & Pérez-Quiñones, M. A. (2017). Personal information management. Chapter in Levine-Clark, M., & McDonald, J. (Eds.), *Encyclopaedia of Library and Information Science, Fourth Edition*, 2017, pp. 3584–3605. Taylor & Francis. <https://doi.org/10.1081/E-ELIS4-120053695>.
- Krtalić, M., & Dinneen, J. D. (2022). Information in the personal collections of writers and artists: Practices, challenges and preservation. *Journal of Information Science*. <https://doi.org/10.1177/01655515221084613>.
- Krtalić, M., Dinneen, J.D., Liew, C.L., & Goulding, A. (2021). Personal collections and personal information management in the family context. *Library Trends*, 70(2), 149–179.
- Krtalić, M., & Ihejirika, K. T. (2022). The things we carry: migrants' personal collection management and use. *Journal of Documentation*. <https://doi.org/10.1108/JD-12-2021-0236>.
- Savolainen, R. (2018). Berrypicking and information foraging: Comparison of two theoretical frameworks for studying exploratory search. *Journal of Information Science*, 44(5), 580–593.
- Society of American Archivists. (2013). Donating your personal or family records to a repository. Retrieved March 19, 2022, from <https://www2.archivists.org/publications/brochures/donating-familyrecs>.

A Survey of Undergraduate Engineering Students' Personal Information Management Practices

Dodson, Samuel

University at Buffalo, USA | smdodson@buffalo.edu

ABSTRACT

Personal information management (PIM) includes the activities and preferences that support future information use. PIM is especially important in the context of engineering education, where learners are expected to engage with a substantial amount of information. This paper builds on the findings of a research project on the information practices of undergraduate engineering students, examined through multiple qualitative and quantitative methods (Dodson, 2021). This paper reports on one aspect of the project, an online questionnaire ($n = 103$). Findings suggest learners regularly engage in PIM activities, and experience different degrees of challenge when managing multiple types of information.

KEYWORDS

Engineering Education; Information Literacy; Information Seeking and Use; Personal Information Management

INTRODUCTION

Personal information management (PIM) is a crucial skill for many individuals. In particular, PIM is essential for university students to develop and use in order to complete their coursework and transition into professionals. PIM encompasses the activities and preferences for gathering information, managing it, and (re)retrieving it for future tasks (Bergman & Whittaker, 2016; Jones, 2007). In this study, we surveyed the PIM practices of undergraduate engineering students. Other library and information science researchers have recognized the role of engineers in building, maintaining, and operating critical infrastructure, and have studied the information needs and seeking behaviors of professionals (Fidel & Green, 2004; Freund, 2015; Hertzum & Pejtersen, 2000) and students (Andrews & Patil, 2007; Fosmire, 2012; Mercer et al., 2019). However, there remains a limited understanding of the PIM practices of learners, including the prevalence of activities for managing course- and program-based information. This led us to ask: *How are undergraduate engineering students managing information?*

Dodson et al. (2019) found that undergraduate engineering students interact with diverse information ecologies. For example, learners are expected to create, manage, and use a range of document genres (e.g., journal articles, patents, and technical reports). Artemeva and Fox (2010) argued that it is imperative for students to develop an awareness of these genres, which can be a lengthy and difficult learning process (Artemeva, 2008; 2009). Course-based information is accessed through a number of information systems (e.g., digital libraries, learning management systems, and Web-based homework systems). Information systems commonly used in North American engineering education programs often lack “*interoperability*”, or the ability to easily access, manage, and use information across different services (Dodson et al., 2019; Harandi et al., 2018). This can result in “*information islands*” (Jones, 2007), where students maintain multiple collections of information, fragmented by specialty tools. This led us to ask: *In what ways do undergraduate engineering students find it challenging to manage multiple types of information?*

Previous research has also found that undergraduate engineering students supplement the information they receive from their instructors by actively seeking additional resources (Atman et al., 2005; Fosmire, 2012; Wertz et al., 2011). This led us to ask: *How do undergraduate engineering students supplement the course materials provided by their instructors by seeking information from other sources?*

METHODS

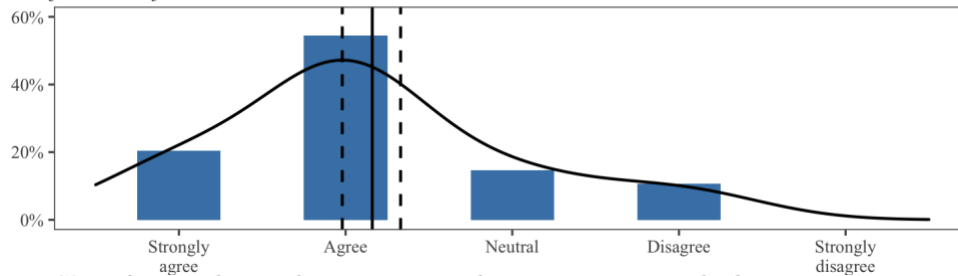
To examine undergraduate engineering students' perspectives on these questions, we administered an online questionnaire. A call for participation was e-mailed to all undergraduates enrolled in computer (CE), electrical (EE), and mechanical engineering (ME) bachelor's programs at the University of British Columbia through each department's listserv. In total, 103 undergraduates completed the questionnaire, including 16 CE, 22 EE, and 65 ME respondents. We asked respondents three questions about their PIM practices, as described in the following section (see also Figure 1). For each question, respondents selected one of five Likert-type scale options to indicate how much they agreed with a statement — “Strongly agree” (1), “Agree” (2), “Neutral” (3), “Disagree” (4), and “Strongly disagree” (5). We analyzed the responses as numeric data, in accordance with Harpe (2015), calculating descriptive statistics and visualizing the underlying probability distribution of the data (Figure 1).

RESULTS

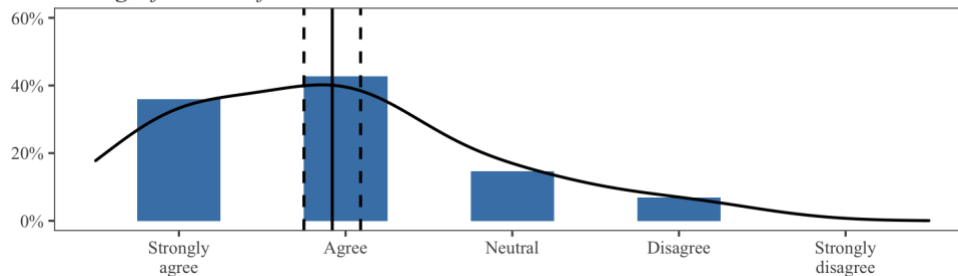
The majority of respondents ($n = 78\%$) agreed or strongly agreed with the first statement, *I carefully manage my course-related information, so I can easily find the information I need* ($n = 103$, $M = 2.16$, $SD = 0.87$). In response to the second statement, *I often supplement the course materials my instructor provides by seeking information from*

other sources, the majority of respondents ($n = 81$, 79%) agreed or strongly agreed ($n = 103$, $M = 1.92$, $SD = 0.88$). There was less consensus in response to the third statement, *I find it challenging to manage multiple types of information when doing some or all of my academic activities* ($n = 103$, $M = 2.88$, $SD = 1.08$). For this question, 39 respondents (38%) agreed or strongly agreed, 31 (30%) were neutral, and 33 (32%) disagreed or strongly disagreed.

S1: I carefully manage my course-related information, so I can easily find the information I need.



S2: I often supplement the course materials my instructor provides by seeking information from other sources.



S3: I find it challenging to manage multiple types of information when doing some or all of my academic activities.

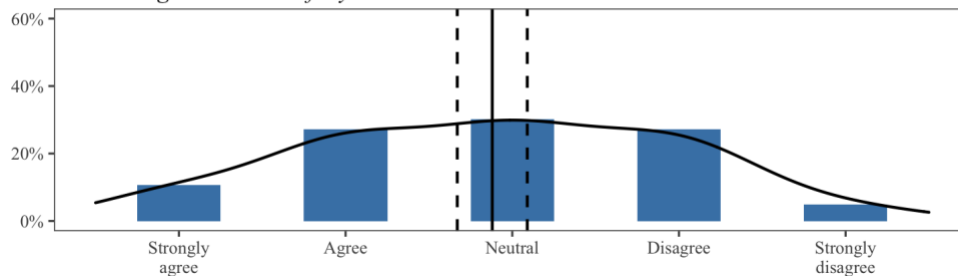


Figure 1. The bar and density plots for the statements. The solid vertical line in each plot represents the mean, as reported in this section. The dashed vertical lines to the left and right of the mean lines represent the lower and upper bounds of the 95% confidence interval, respectively. The continuous curve in each plot visualizes the probability distribution of the data. The curves were calculated using kernel density estimation, with a Gaussian kernel and a bandwidth of 0.5.

CONCLUSION & FUTURE WORK

Questionnaire responses indicate that undergraduate engineering students regularly engage in PIM and information seeking activities. Self-reported PIM practices appear to align with findings identified with other methods (e.g., Dodson et al., 2019). Findings have implications for strengthening understandings of the ways in which learners interact with a range of information types and systems, as well as the factors that influence how information is managed. However, the research questions cannot be fully answered by the questionnaire alone. In-depth qualitative or mixed methods studies should be carried out to better understand how undergraduate engineering students manage information. In a subsequent study, we also plan to investigate whether PIM practices change in response to the genres students are using or the tasks they are engaged in. Responses also suggest that students experience considerably different degrees of challenge when faced with managing multiple types of information. Consequently, PIM-specific information literacy instruction may be useful to a large proportion of undergraduate engineering students. Future work may explore whether the shift to online learning in response to the COVID-19 pandemic has amplified students' difficulties managing information, given the lack of interoperability between information systems.

REFERENCES

- Andrews, T., & Patil, R. (2007). Information literacy for first-year students: An embedded curriculum approach. *European Journal of Engineering Education*, 32(3), 253–259.
- Artemeva, N. (2008). Toward a unified social theory of genre learning. *Journal of Business and Technical Communication*, 22(2), 160–185.
- Artemeva, N. (2009). Stories of becoming: A study of novice engineers learning genres of their profession. In C. Bazerman, A. Bonini, & D. Figueiredo (Eds.), *Genre in a changing world* (pp. 158–178). The WAC Clearinghouse and Parlor Press.
- Artemeva, N., & Fox, J. (2010). Awareness versus production: Probing students' antecedent genre knowledge. *Journal of Business and Technical Communication*, 24(4), 476–515.
- Atman, C. J., Cardella, M. E., Turns, J., & Adams, R. (2005). Comparing freshman and senior engineering design processes: An in-depth follow-up study. *Design studies*, 26(4), 325–357.
- Bergman, O., & Whittaker, S. (2016). *The science of managing our digital stuff*. MIT Press.
- Dodson, S. (2021). *Becoming engineers: How students leverage relationships between documents and learning activities* [Doctoral dissertation, University of British Columbia].
- Dodson, S., Roll, I., Harandi, N. M., Fels, S., & Yoon, D. (2019). Weaving together media, technologies, and people: Students' information practices in flipped classrooms. *Information and Learning Sciences*, 120(7/8), 519–540.
- Fidel, R., & Green, M. (2004). The many faces of accessibility: Engineers' perception of information sources. *Information Processing & Management*, 40(3), 563–581.
- Fosmire, M. (2012). Information literacy and engineering design: Developing an integrated conceptual model. *International Federation of Library Associations and Institutions Journal*, 38(1), 47–52.
- Freund, L. (2015). Contextualizing the information-seeking behavior of software engineers. *Journal of the Association for Information Science and Technology*, 66(8), 1594–1605.
- Harandi, N. M., Agharebparast, F., Linares, L., Dodson, S., Roll, I., Fong, M., Yoon, D., & Fels, S. (2018). Student video-usage in introductory engineering courses. *Proceedings of the Canadian Engineering Education Association*, 1–8.
- Harpe, S. E. (2015). How to analyze Likert and other rating scale data. *Currents in Pharmacy Teaching and Learning*, 7(6), 836–850.
- Hertzum, M., & Pejtersen, A. M. (2000). The information-seeking practices of engineers: Searching for documents as well as for people. *Information Processing & Management*, 36(5), 761–778.
- Jones, W. (2007). Personal information management. *Annual Review of Information Science and Technology*, 41(1), 453–504.
- Mercer, K., Weaver, K., & Stables-Kennedy, A. (2019). Understanding undergraduate engineering student information access and needs: Results from a scoping review. *Proceedings of the American Society of Engineering Education*.
- Wertz, R. E., Ross, M., Fosmire, M., Cardella, M., & Purzer, S. (2011). Do students gather information to inform design decisions? Assessment with an authentic design task in first-year engineering. *Proceedings of the American Society for Engineering Education Annual Conference*.

Exploring News Tagging on Spanish Digital Press

Domínguez-Delgado, Rubén

University of Seville, Spain | rdd@us.es

López-Hernández, María-Ángeles

University of Seville, Spain | alhernan@us.es

Muñoz-García, Francisco-Javier

University of Huelva, Spain | fjpgarcia@uhu.es

ABSTRACT

A proper news tagging is essential for a suitable search and retrieval of information by news libraries users, as well as for a suitable web SEO positioning of these media outlets, what is an aspect to take increasingly into account in digital journalism. Through a case study, we carried out a comparative exploratory analysis on how the most read Spanish online newspapers assign the tags which identify the informational content of their news, given the scarcity of this kind of empirical studies and the enormous importance of this task in the field of news librarianship. The resulting data demonstrate the need for specific training in news librarianship for journalists and the importance of educating media outlets in using controlled vocabularies to ensure an efficient retrieval of news in databases a posteriori.

KEYWORDS

News libraries; news archives; news tagging; tags; metadata; journalism

INTRODUCTION

In current digital journalism, not exclusively the news librarians—as it happened some years ago in the news libraries of printed newspapers—are in charge of determining and assigning, in the form of content tags or keywords, the relevant content that best identifies each news article and that will allow users to access them. As a result of the deep crisis that media outlets—especially press—and their archives have been suffering in this century (Paul, 2009), they are their own journalists who carry out this task (Rubio Lacoba, 2012:71; García-Jiménez et al., 2019:62) and other archiving duties (Micó-Sanz et al., 2009; Guallar, 2011:53). Thus, at the same time of publishing online news items journalists must tag them, these tags being reviewed by news librarians only in those newsrooms which still count on these professionals (Rubio Lacoba, 2012:75; Marcos Recio & Edo, 2015:398-399). Hence the special need, in the current context, that university journalism education provides an appropriate training in news libraries to future journalists (Domínguez-Delgado et al., 2019, 2021).

News tagging is today considered an important aspect for a good SEO—Search Engine Optimization—positioning of online newspapers on the web (Zamith, 2008:173; Iglesias-García & Codina, 2016; Lopezosa et al, 2021:40). For this reason, this task should be seriously taken into account in digital journalism (Richmond, 2008; Usher, 2010). Authors such as Zamith (2008) or Guallar (2012) consider the news tagging system as a quality indicator for digital newspapers. Moreover, the most common metadata standards, such as Dublin Core, make reference to these content keywords as essential information related to any digital document (Baños-Moreno et al., 2015). In addition, this tagging makes possible semantic navigation (Castellanos Díaz, 2011:11), thanks to which users can access, by clicking on tags, *dossiers* on specific contents, these tools being traditionally considered very useful in media outlets libraries.

RESEARCH OBJECTIVE AND METHODOLOGY

Given the scarcity of empirical works which explore whether or not this important archiving task of news tagging is being carried out in an appropriate way by online newspapers today, we aimed to carry out a case study to comparatively analyze the way how the most popular Spanish online newspapers tag their news.

We firstly selected all those newspapers in which the tags assigned to their news articles were visible to readers, from the ranking of the 10 Spanish digital newspapers with the highest number of unique visitors in January 2022, according to *GfK DAM (Dircomfidencial, 2022)*. These media outlets were, in a descending order of visitors: *ElPaís.com*, *ElMundo.es*, *20Minutos.es*, *ABC.es*, *ElEspañol.com*, *ElConfidencial.com*, *ElPeriódico.com*, *ElDiario.es* and *OKDiario.com*. *LaVanguardia.com*, ranked in fifth place, was left out, since no tags were visible on its news articles.

Secondly, we used the constructed week sampling technique (Stempel, 1952). Thus, we successively took one different day of the week per week, for 7 weeks in total, in order for the selected news to be different. And we took 15 news items per day and newspaper, those being posted on the uppermost location of their websites—which were supposed to be the most relevant news. Considering this, the tags on 945 digital news articles were analyzed, in the period from Monday, March 14 to Sunday, May 1, 2022.

For this analysis, a table was used in which we recorded: number of item, newspaper, date, headline, topic, total number of tags assigned to each news item and number of tags corresponding to each of the main 4 tags categories

85th Annual Meeting of the Association for Information Science & Technology | Oct. 29 – Nov. 1, 2022 | Pittsburgh, PA. Author(s) retain copyright, but ASIS&T receives an exclusive publication license.

resulting from the essential considerations on news content analysis by experts (e.g., ISO, 1985; Cunha, 1989; Pinto-Molina et al., 2002; Fox, 2005; Cleveland, 2013; García Gutiérrez, 2014; Domínguez-Delgado & López-Hernández, 2016, 2017). These categories are: thematic, onomastic, geographic and temporal tags. We also recorded—by category—the literal transcription of each tag, in order to conduct a comparative qualitative analysis, observing: pertinency, relevance, rigor, terminology/vocabulary and coherence of the news tagging inside each newspaper and among them.

RESULTS AND DISCUSSION

As can be seen in Table 1, which shows the quantitative results, the total average number of tags assigned to each news article was 5.14. While *ElPaís.com* positively stood out as the newspaper that annotated the highest number of tags on its news items (11.98), *ElPeriódico.com* was the newspaper with the lowest number of tags per article (3.69). Regarding the categories of tags, it should be noted that only on *ElPaís.com* some exceptional temporal tags were assigned—all of them relating to periods of history such as “*neolithic*” or “*paleolithic*”. This fact implies that times which are mentioned by journalists inside news articles—dates, years, decades, centuries—are not usually taken into account in their indexing, this being an obstacle to their retrieval by users.

Online newspaper	Average number of tags per news item	Average number of thematic tags per news item	Average number of onomastic tags per news item	Average number of geographic tags per news item	Average number of temporal tags per news item	% of news with 1 only category of tags	% of news with 2 categories of tags	% of news with 3 categories of tags	% of news with 4 categories of tags
<i>ElPaís.com</i>	11.98	6.96	2.58	2.41	0.03	0%	31%	69%	0%
<i>ElMundo.es</i>	4.12	1.07	2.09	0.96	0.00	46%	41%	13%	0%
<i>20minutos.es</i>	3.73	1.40	1.73	0.60	0.00	20%	34%	46%	0%
<i>ABC.es</i>	4.31	1.44	1.82	1.05	0.00	26%	54%	20%	0%
<i>ElEspañol.com</i>	4.73	2.12	1.91	0.70	0.00	5%	69%	26%	0%
<i>ElConfidencial.com</i>	4.21	1.39	2.52	0.30	0.00	41%	47%	12%	0%
<i>ElPeriódico.com</i>	3.69	1.16	1.48	1.05	0.00	38%	42%	20%	0%
<i>ElDiario.es</i>	4.68	2.27	1.62	0.79	0.00	19%	42%	39%	0%
<i>OKDiario.com</i>	4.74	1.14	3.13	0.47	0.00	31%	24%	45%	0%
Total average	5.14	2.11	2.10	0.93	0.00	25%	43%	32%	0%

Table 1. Quantitative results of the analysis of news tagging

In relation to the other categories, the tags that were assigned to a greater extent to news, taking into account the total averages, were thematic (2.11 per news item), closely followed by onomastics (2.10) and further by geographic ones (0.93). Regarding the number of categories, it should be pointed out as the most frequent tendency in tagging to use tags linked to 2 different categories (in 43% of news). Only on *ElPaís.com*, *20Minutos.es* and *OKDiario.com* it was the most common practice that tags related to 3 different categories were assigned to news articles (in 69%, 46% and 45% of their items, respectively). On the contrary, *ElMundo.es* preferably opted for 1 only category of tags (46%).

On the other hand, from the qualitative study of tags, it is mainly inferred a general poor selection of indexing terms—many relevant terms or concepts are usually missing in the news tags—and a lack of control of vocabulary, detecting linguistic problems, such as synonymy and polysemy, in *ElPeriódico.com*, in geographic tags, as well as in *ElMundo.es*, *20Minutos.es* and in *ElConfidencial.com*, in onomastic tags. We specifically observed an inaccuracy and low specificity of thematic tags, except in the case of *ElPaís.com*, and geographic tags, except in *ABC.es*, which added references to “*city*” or “*province*”, in brackets, when both tags shared a same name. In addition, we detected a lack of homogeneity when using the same type of term—for example, in *ElConfidencial.com*, political parties were sometimes named by their acronyms (“*PSC*”), others by their full names (“*Partido Popular*”) and others by their full names followed, in brackets, by their acronyms (“*Esquerra Republicana de Catalunya (ERC)*”). These problems could cause the lack of results or noise after a search on newspapers databases (Muñoz-García et al., 2021:796).

CONCLUSIONS

In the most read Spanish online newspapers, the task of indexing their news articles, which is essential for an appropriate access to them by users, generally have considerable shortcomings, except on *ElPaís.com*, these being mainly related to the low number of assigned tags and the lack of coherence and rigor in news tagging both intra and enter newspapers. Therefore, it is urgent, on the one hand, to make online newspapers aware of the importance of having controlled vocabularies—thesauri—to overcome these deficiencies; and, on the other hand, to provide at universities a suitable academic training in news libraries for journalism students, those who will be working in the future in newsrooms and tagging their everyday news articles.

REFERENCES

Baños-Moreno, M.J.; Felipe, E.R.; Pastor-Sánchez, J.A.; Martínez-Béjar, R. & Lima, G. (2015). Metadatos en noticias: un análisis internacional para la representación de contenidos en periódicos, In *XII Congreso ISKO España y II Congreso ISKO*

España-Portugal, 19-20 de noviembre. *Organización del conocimiento para sistemas de información abiertos*, University of Murcia.

- Castellanos Díaz, J. (2011). De lo impreso a lo digital: la migración de los periódicos impresos de América Latina a los entornos digitales. *Razón y Palabra*, 77. <https://www.redalyc.org/pdf/1995/199520010063.pdf>
- Cleveland, D. & A. (2013). *Introduction to Indexing and Abstracting*. Littleton: Libraries Unlimited. Santa Barbara, CA: Libraries Unlimited.
- Cunha, I. (1989). *Análise documental: considerações teóricas e experimentações*. Sao Paulo. Febab.
- Domínguez-Delgado, R.; Muñoz-García, F.J. & López-Hernández, M.A. (2021). A historical introduction to Library and Information Science education in Journalism studies. *Ibersid*, 15(2), 57-65.
- Domínguez-Delgado R.; Hansen K.A. & Paul N. (2019). Educating journalism students about news archives: a global comparison with special focus on Spain and the United States. *Journalism & Mass Communication Educator*, 74(1), 3-16.
- Domínguez-Delgado, R., & López-Hernández, M.-A. (2017) A methodological proposal of content analysis for non-fiction movies at film archives. *Revista General De Información Y Documentación*, 27(2), 527-550.
- Domínguez-Delgado, R., & López-Hernández, M.-A. (2016). The retrieval of moving images at Spanish film archives: the oversight of content analysis. *Proceedings of the Association for Information Science And Technology*, 59(1), 1-4.
- Dircomfidencial.com (2022). El País lidera el primer mes completo auditado por GfK DAM. *Dircomfidencial.com*, February 22. <https://dircomfidencial.com/medios/el-pais-lidera-el-primer-mes-completo-auditado-por-gfk-dam-20220216-0403>
- Fox, V. (2005). *Análisis documental de contenido. Principios y prácticas*. Buenos Aires: Alfagrama Ediciones
- García Gutierrez, A. (2014). Document analysis of news items for factual information systems. *Revista Española de Documentación Científica*, 37(2): e046.
- Guallar, J. (2011). La documentación en la prensa digital. Nuevas tendencias y perspectivas. *III Congreso Internacional de Ciberperiodismo y Web 2.0. La transformación del espacio mediático*. <https://core.ac.uk/download/pdf/11888476.pdf>
- Guallar, J. (2012). *Las hemerotecas de la prensa digital. Análisis de diarios españoles Javier Guallar* [Ph diss]. Universitat de Barcelona. <http://hdl.handle.net/10803/96817>
- Iglesias-García, M. & Codina, L. (2016). Online media and the strategic importance of Search Engine Optimization (SEO). *Opción*, 32(9), 929-944.
- ISO (International Standardization Organization) (1985). *ISO 5963:1985. Documentation - Methods for examining documents, determining their subjects, and selecting indexing terms*. <https://www.iso.org/standard/12158.html>
- Lopezosa, C.; Trillo-Domínguez, M.; Codina, L. & Cabrera Méndez, M. (2021). SEO in the journalistic company: perceptions and key elements for its adoption in writing. *Revista Latina de Comunicación Social*, 79, 27-45.
- Marcos Recio, J.C. & Edo, C. (2015). Analysis of the new perspective of journalistic documentation in the Spanish media. *Revista General de Información y Documentación*, 25(2), 389-423. 873.
- Micó-Sanz, J.L.; Masip-Masip, P. & García-Avilés, J.A. (2009). Journalists as news librarians (and viceversa?). New relationships between newsroom and archive after media digitization. *Profesional de la información*, 18(3), 284-290.
- Muñoz-García, F.J; López-Hernández, M.A. & Domínguez-Delgado, R. (2021). Written news search engines and retrieval systems of the databases of spanish digital newspapers. *Proceedings of the Association for Information Science and Technology*, 58(1), 795-797.
- Paul, N. (2009). Elegy for the news library. *Profesional de la información*, 18(3), 249-254.
- Pinto-Molina, M.; García-Marco, F.J. & Agustín-Lacruz, M.C. (2002). *Indización y resumen de documentos digitales y multimedia. Técnicas y procedimientos*. Gijón: Ediciones Trea
- Richmond, S. (2008). How SEO is changing journalism. *British Journalism Review*, 19(4), 51-55.
- Rubio Lacoba, M. (2012). New documentary skills for journalists: the collaborative vocabulary at El País newspaper. *Trípodos*, 31, 65-78.
- Stempel, G.H. (1952) Sample size for classifying subject matter in dailies: research in brief. *Journalism Quarterly*, 29, 333-334.
- Usher, N. (2010). What impact is SEO having on journalists? Reports from the field. *Nieman Journalism Lab*, September 23. <http://www.Niemanlab.org/2010/09/what-impact-is-seo-having-on-journalists-reports-from-the-field>.
- Zamith, F. (2008). A methodological proposal to analyze the news websites' use of the potentialities of the Internet. *Observatorio (OBS*) Journal*, 5, 165-191.

Transitioning Towards Sustainability: The Information Practices of Sustainable Food Consumers

Du, Xiaotong

Rutgers University, New Brunswick, USA | xd103@rutgers.edu

ABSTRACT

Sustainability is an emergent topic due to the increased climate and social issues caused by human activities. While sustainability is not a new concept in library and information science, its scope is often limited to library and information systems. This study aims to expand sustainability to human information behavior and practice. This poster reports initial findings from a pilot study in Jan – April 2022. Interviews and photo diaries with 6 sustainable food consumers about their information behaviors and practices were conducted and analyzed using the constructivist grounded theory methodology. We found that participants use sensory and embodied information in assessing the real impact of their sustainable food behaviors, and that this process is essential in the transition to sustainable food behaviors. Participants assess the impact of their transition to sustainable food by comparing information at different times, places, and between actors (individuals and food providers). This research sheds light on the role of time in everyday information use, and demonstrates that sustainability is a fruitful research area in human information behavior and practice.

KEYWORDS

Sustainability; Human Information Behavior; Food Consumption; Transition; Information Use

INTRODUCTION

Sustainability and sustainable development have drawn increased attention in various disciplines over the past few decades (Schoolman et al., 2012). Information will play an essential role in the transition toward a sustainable future (Kamińska et al., 2022). Sustainability research in library and information science (LIS) mainly focuses on reducing the energy consumption or carbon emissions in library and information systems, such as sustainable library buildings, collections, digital services, green Information and Communication Technologies (ICTs) (Kamińska et al., 2022). However, sustainability in everyday life from the human-centered information behavior or information practice (IB/IP) perspective has been understudied. Therefore, this study expands sustainability research in LIS with a focus on everyday IB/IP.

In everyday life, household food consumption has a significant impact on health and the environment. Food is a nuanced field and has been an information-rich research area in LIS and IB/IP (e.g., Hartel, 2010; Ocepek, 2016, 2018). However, sustainable food consumers have not drawn much attention in the IB/IP field, perhaps in part because environmental and ethical concerns are not necessarily primary drivers in food choices. However, Polkinghorne (2021) recently described how moral concerns (e.g., animal welfare, environment) influence food-related information practices. Therefore, we aim to extend previous research on food-related everyday information practices by focusing on sustainable food consumers. The question we aim to address here is: *How do sustainable food consumers engage with information in everyday food-related consumption?*

METHOD

Constructivist grounded theory (Charmaz, 2014) was employed to explore the actions, processes, and meanings of sustainable food consumers' food choices and theorize their information practices. Constructivist grounded theory recognizes the active role of participants and researchers to co-construct knowledge in the process. Sustainable food consumers, in this study, refer to people who, in the last month, have made food choices based on environmental, political, social, or ethical concerns (Peattie, 2010; Schoolman, 2020).

Data were collected from pre-screening phone calls, photo diaries, and semi-structured interviews via email and video-conferencing tools (i.e., Zoom). Six participants, all sustainable food consumers who live in a large Northeastern state in the U.S., were recruited through special interest groups of sustainability from social media (e.g., Facebook) and snowball sampling between January and March of 2022. Participants are all women, white, and Democrats, which aligns with previous research on a typical demographic of sustainable consumers (Schoolman, 2020). Pre-screening phone calls ensured participants' eligibility and collected demographics and preferred pseudonyms. We also asked them to take photos of their sustainable food consumption behaviors (e.g., grocery shopping, vegetable gardens). These photo diaries were discussed during the interview. Semi-structured interviews focus on participants' (1) perception of sustainability, (2) food consumption and information practices based on photo-elicitation techniques (Clark-Ibáñez, 2004; Harper, 2002), (3) information behaviors/practices based on critical incident techniques (Flanagan, 1954; Sonnenwald et al., 2001). Notes were taken during the pre-screening phone calls. Interviews were audio-recorded and lasted from 45 to 90 minutes. In total, 46 photos and 381 minutes of recording were collected. Participants were compensated with a \$15 gift card after each interview. The recorded

interviews were transcribed verbatim, and photos were imported for data analysis. We used theoretical sampling to collect data and constant comparing, line-by-line coding, asking questions, and saturation to analyze data (Charmaz, 2014; Corbin & Strauss, 2014). Extensive jotting, analytical memoing, and weekly peer debriefings were also employed to increase the credibility of the findings (Charmaz, 2014; Corbin & Strauss, 2014; Miles et al., 2018).

FINDING: INFORMATION USE OF SUSTAINABLE FOOD CONSUMERS

The primary findings are centered around the use of information to guide sustainable food consumption. This section only reports on a relatively well-saturated concept in the analysis thus far: *assessing the real impact of individuals' sustainability transitions*.

Assessing the Real Impact of Individuals' Sustainability Transitions

Sustainable food consumption is not one single behavior in a linear process but involves various types of sustainable food shopping or acquisitions at different transitions simultaneously. Sustainability transitions include maintaining current sustainable food behaviors, adopting new ones, replacing one with another, or giving one up. For instance, Ann was not only buying food from local farms and growing vegetable gardens (maintaining a sustainable food behavior) but also paying more attention to labor rights (adopting a new sustainable food behavior).

During sustainability transitions, participants used information to assess the real impacts of their efforts. The process involves comparing sensory and embodied information about the impact of sustainable food practices at different times, spaces, or between individuals and food providers. First, participants compared visual information in different time. Valva had noticed more options of sustainably sourced foods in the restaurants since she started her sustainable food journey 17 years ago: *"there's been more things on menus, there's been more options in restaurants."* Similarly, after planting more native plants, Aama saw an increased number of pollinators and insects in her garden: *"I can see when more if I plant milkweed that the monarch butterflies are going to come, so if you know that I'm doing something that is impact that I can see more caterpillars."*

Second, participants compared embodied information at different times and spaces. Aama noted her bodily feeling of the hot temperature lasted longer in summer but praised the lower temperature under the tree in her backyard: *"I know in my backyard sitting under the trees how climatically different it, it feels."* Accordingly, the observed positive changes over time made participants optimistic about their real impact, which, in turn, encourages them to maintain sustainable food behaviors. As Valva noted: *"I feel like the more I move forward on that, the more companies will rise to that occasion."*

However, one participant also gave up current sustainable behaviors when comparing their own sustainable practices to companies' behaviors in a short period of time. During the COVID-19 pandemic, Louise recalled her negative feeling by observing the wasteful food packaging used by a restaurant while picking up food with her own reusable containers: *"It was sad because I would be there with my bag and my reusable container... I stand there and wait for them to fill the containers with food and see the amount [of plastic packaging] that's being sent out in just that short time that I'm standing there."* Therefore, participant perceived limited impact due to the observed unsustainable behaviors of food providers and discontinued their sustainability transition.

DISCUSSION

While there is no consensus on the definition of information use (Savolainen, 2008; Spink & Cole, 2006; Todd, 1999), researchers advocate the need to study how people use information to transform their everyday behaviors (Fidel, 2012; Spink & Cole, 2006). Our research supports this with empirical data. In addition, prior studies have shown how sensory and bodily experience informed daily food activities, such as grocery shopping (Ocepek, 2018), cooking, and eating (Polkinghorne, 2021). Our study further demonstrates how participants use embodied information to reinforce or discontinue sustainable food transitions. Our finding also illustrates that *time* impacts information use, because participants were informed by comparing changes of sensory (e.g., numbers of pollinators) and embodied information (e.g., temperature) before and after their sustainable behaviors. It echoes Dervin's Sense-Making (Dervin, 1999) that conceptualizes a human living in a time-space. Lastly, food providers could present more sensory and embodied information (e.g., visualizing sustainable changes over time) to encourage sustainability transitions.

LIMITATIONS AND FUTURE RESEARCH

First, participants lack gender, racial, geographical, and political diversity. Second, we have a small sample size (n=6) in this pilot study. Research is ongoing and the theoretical sampling process will continue until we reach saturation (Charmaz, 2014; Corbin & Strauss, 2014; Low, 2019). Third, participants in our study went beyond sustainable food consumption by engaging in other aspects of their life (e.g., fashion, waste disposal). Therefore, our study shows a fruitful result of studying the nuanced and multifaceted concept of sustainability in the IB/IP field

ACKNOWLEDGMENTS

We are deeply grateful for Dr. Ross Todd's and Dr. Ethan Schoolman's guidance and peer-debriefing. We thank Dr. Kaitlin Costello's for their help in reviewing and revising the manuscript and all participants for their time and inputs.

REFERENCES

- Charmaz, K. (2014). *Constructing Grounded Theory*. SAGE.
- Clark-Ibañez, M. (2004). Framing the Social World With Photo-Elicitation Interviews. *American Behavioral Scientist*, 47(12), 1507–1527. <https://doi.org/10.1177/0002764204266236>
- Corbin, J., & Strauss, A. (2014). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. SAGE Publications.
- Dervin, B. (1999). On studying information seeking methodologically: The implications of connecting metatheory to method. *Information Processing & Management*, 35(6), 727–750. [https://doi.org/10.1016/S0306-4573\(99\)00023-0](https://doi.org/10.1016/S0306-4573(99)00023-0)
- Fidel, R. (2012). *Human Information Interaction: An Ecological Approach to Information Behavior*. MIT Press.
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51(4), 327–358.
- Harper, D. (2002). Talking about pictures: A case for photo elicitation. *Visual Studies*, 17(1), 13–26. <https://doi.org/10.1080/14725860220137345>
- Kamińska, A. M., Opaliński, Ł., & Wyciślik, Ł. (2022). The Landscapes of Sustainability in the Library and Information Science: Systematic Literature Review. *Sustainability*, 14(1), 441. <https://doi.org/10.3390/su14010441>
- Low, J. (2019). A Pragmatic Definition of the Concept of Theoretical Saturation. *Sociological Focus*, 52(2), 131–139. <https://doi.org/10.1080/00380237.2018.1544514>
- Miles, M. B., Huberman, A. M., & Saldana, J. (2018). *Qualitative Data Analysis: A Methods Sourcebook*. SAGE Publications.
- Ocepek, M. G. (2018). Sensible Shopping: A Sensory Exploration of the Information Environment of the Grocery Store. *Library Trends*, 66(3), 371–394. <https://doi.org/10.1353/lib.2018.0008>
- Peattie, K. (2010). Green Consumption: Behavior and Norms. *Annual Review of Environment and Resources*, 35(1), 195–228. <https://doi.org/10.1146/annurev-enviro-032609-094328>
- Polkinghorne, S. C. (2021). *Exploring everyday information practices: Embodied mutual constitution of people's complex relationships with food*. Swinburne University of Technology.
- Savolainen, R. (2008). *Everyday Information Practices: A Social Phenomenological Perspective*. Scarecrow Press.
- Schoolman, E. D. (2020). Building community, benefiting neighbors: “Buying local” by people who do not fit the mold for “ethical consumers.” *Journal of Consumer Culture*, 20(3), 285–304. <https://doi.org/10.1177/1469540517717776>
- Schoolman, E. D., Guest, J. S., Bush, K. F., & Bell, A. R. (2012). How interdisciplinary is sustainability research? Analyzing the structure of an emerging scientific field. *Sustainability Science*, 7(1), 67–80. <https://doi.org/10.1007/s11625-011-0139-z>
- Sonnenwald, D. H., Wildemuth, B. M., & Harmon, G. L. (2001). *A Research Method to Investigate Information Seeking using the Concept of Information Horizons: An Example from a Study of Lower Socio-economic Students' Information Seeking Behavior*. 22.
- Spink, A., & Cole, C. (2006). Human information behavior: Integrating diverse approaches and information use. *Journal of the American Society for Information Science and Technology*, 57(1), 25–35. <https://doi.org/10.1002/asi.20249>
- Todd, R. J. (1999). Utilization of heroin information by adolescent girls in Australia: A cognitive analysis. *Journal of the American Society for Information Science*, 50(1), 10–23. [https://doi.org/10.1002/\(SICI\)1097-4571\(1999\)50:1<10::AID-ASI4>3.0.CO;2-B](https://doi.org/10.1002/(SICI)1097-4571(1999)50:1<10::AID-ASI4>3.0.CO;2-B)

A Preliminary Analysis of Outreach Librarian's Knowledge and Skills

Du, Yunfei

University of North Texas, USA | Yunfei.du@unt.edu

Byrne, Sephra

University of North Texas, USA | Sephra.byrne@unt.edu

ABSTRACT

Library services go beyond knowledge gatekeepers in the current information-rich environment and provide outreach services to their communities. This paper analyzed 85 job postings from ALA Job lists that included the term “outreach” and analyzed job titles, duties, and qualifications in a 3-dimensional plot. Implications for information science education are also discussed. Findings from this study may help library and information science programs strive to adapt their curriculum to meet market needs.

KEYWORDS

Community outreach, libraries, data visualization, knowledge and skills

INTRODUCTION

In the current information-rich environment, there is increasing demand to connect people to information and to succeed in a resilient information environment. Many libraries have adapted from knowledge gatekeepers and begun providing innovative outreach services to their communities. This preliminary study aims to explore expected knowledge and skills for outreach librarians and offer suggestions for library education based on current market needs. The research questions for this study are 1) What are the expected knowledge, skills, and qualifications for outreach librarians? 2) How are libraries adopting outreach services?

LITERATURE REVIEW

Library outreach is reflected in multiple ways, including collaborative programs and joint events with community members and, more formally, partnerships with community members that will share the cost of revenues of the collective effort. Library community outreach can be reflected in expected qualifications, knowledge, and skills from some public job postings. Library services have evolved from a traditional desk model and their assigned areas. Studies found reference, subject, and liaison librarians tend to be more visible to users and communities and an integral part of online and in-person instruction (Johnson, 2018; Tumbleson, Burke, & Long, 2019; Zanin-Yost, 2018). In addition, academic libraries can market their services at campus events and share resources for sponsoring events, such as open houses, orientations, and book sales (LeMire & Ballestro, 2019).

METHOD

The authors sampled 85 job postings from the American Library Association job posting site (joblist.ala.org) that included the keyword “outreach” and were open from April 22 to May 26, 2021. The authors conducted content analysis on job postings. For each job posting, the title, job duties, and qualifications were coded by researchers on a scale of 1-5, where 1 represents the first stage of the adoption of library outreach, and 5 represents the last stage. Each job posting and its title, duties, and qualifications were covered in a 3-dimensional vector space (title score, duties score, and qualification score). Histograms of each score and the vector space were illustrated in a 3D coordinate to demonstrate the distribution of each job posting (see Figures 1-4 below).

DATA AND RESULTS

Out of these 85 job postings, all but 3 were full-time positions. Among these 82 full-time positions, only one stated full-time entry-level, seven did not answer entry level or not, and all others were not entry-level positions, requiring 1-2 years, 3-5, or more than five years of experience. The titles of these jobs reflected different adoption stages of library outreach: from “reference/instruction librarian” (knowledge or persuasion stage), to “instruction and outreach librarian” (implementation stage), or “outreach librarian” (confirmation stage). Figure 1 shows the histogram of job titles related to outreach.

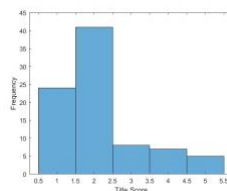


Figure 1. Histogram of Job Titles Related to Outreach

Regarding job duties, the percentage of work related to outreach activities was calculated for each posting. Figure 2 illustrates the histogram of job duties related to outreach.

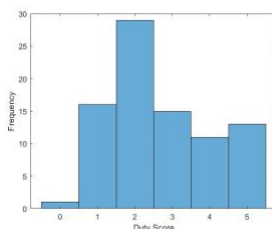


Figure 2. Histogram of Job Duties on Outreach

Job qualification was coded based on description compared to different stages of innovation. For example, the description of knowledge, skills, and abilities related to outreach – such as teaching, research assessment, public communications, promotion, and marketing – were coded and analyzed using a histogram. Figure 3 illustrates the histogram of job qualifications on outreach.

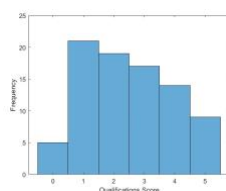


Figure 3. Histogram of Job Qualifications on Outreach

The authors also visualize job postings related to outreach in a 3D coordinate to compare the distance between job postings on scores on stages of innovation. The clustered dots mean these job postings are more related and deserve a detailed look at trends and commonalities. Figure 4 illustrates the frequency of different scores along the title, duties, and outreach axes, with the larger spheres indicating more job postings with that combined score.

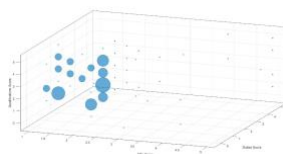


Figure 4. Cluster Job Postings in the 3D Environment

Here are examples of the knowledge and qualifications of outreach librarians:

“Collaborates with library team and university stakeholders to develop, foster, and market library events and exhibits that are educational, informative, innovative, and entertaining. Works in collaboration with librarians to identify and develop strategies to enhance perceptions of the library and improve user engagement with library programming.”

“Essential Job Functions: ... implementing outreach services in accordance with organizational strategic plan and priorities; trains new and current library staff at library locations in effective community outreach; oversees the implementation of outreach-related policies and procedures for staff and customers; provides inclusive services to patrons from a large array of national, lingual, and cultural backgrounds.”

IMPLICATIONS AND CONCLUSION

Job descriptions revealed traditional information science training still covers needed skill sets, such as communications, critical thinking, teaching, assessment, and research. In addition, outreach librarian positions emphasize marketing and promotion skills, which do not seem to be represented in information schools. An unfamiliar environment brings new challenges for outreach librarians. Currently, users may feel overwhelmed with information or misinformation in social interactions, and there is a need for librarians to be able to teach, promote, communicate, and perform research and assessment to conduct better information services. A curriculum on library outreach may meet such market needs.

REFERENCES

- Johnson, A. M. (2018). Connections, conversations, and visibility: How the work of academic reference and liaison librarians is evolving. *Reference & User Services Quarterly*, 58(2), 91-102.
- LeMire, S., & Ballestro, J. M. (2019). Insourcing library outreach: Engaging technical services in outreach to student organizations. *College & Undergraduate Libraries*, 26(2), 149-162. <https://doi.org/10.1080/10691316.2019.1636442>
- Tumbleson, B., & Burke, J, Long, J. (2018). Assessment, analytics, and analysis: Demonstrating the impact of LMS embedded librarians on student learning. *Journal of Library & Information Services in Distance Learning*, 13(1/2), 196-214.
- Zanin-Yost, A. (2018). Academic collaborations: Linking the role of the liaison/embedded librarian to teaching and learning. *College & Undergraduate Libraries*, 25(2), 150-163.

Who Talked about What Regarding Derek Chauvin's Trial: A Work-In-Progress Analysis

Duan, Yiran Syracuse University, USA | yduan12@syr.edu
Hemsley, Jeff Syracuse University, USA | jjhemsle@syr.edu
Smith, Alex Syracuse University, USA | aosmith@syr.edu
Gray, LaVerne Syracuse University, USA | lgray01@syr.edu
Gandhi, Dhvani Syracuse University, USA | dgandhi@syr.edu

ABSTRACT

This work-in-progress study investigates the information sharing behavior of different levels of Twitter influencers within the context of the #BlackLivesMatter social movement and its related discussions #AllLivesMatter and #BlueLivesMatter during the 2-week period around Derek Chauvin's trial. Using qualitative content analysis and quantitative machine learning methods, we analyzed over one million retweets to test if different levels of influencers tend to spread different kinds of information in the discussions around #All/Black/BlueLivesMatter on Twitter. We found out that different levels of influencers tend to spread different information within and between the #All/Black/BlueLivesMatter, and we offer some explanations through the lens of curation logics. We suggest that different levels of influencers may be exposed to different incentives, and be facing different social norms, which leads to different information behaviors. This work-in-progress study contributes to updating the theory of curated logics, virality, and influencers, as well as provides empirical data for the discussions of the #BlackLivesMatter social movement and its related discussions of #AllLivesMatter and #BlueLivesMatter.

KEYWORDS

BlackLivesMatter, Curated flows, Virality, Influencers

INTRODUCTION

This study leverages the concept of social media influencers (Freberg et al., 2011), people who have some influence over other social media users, along with the theory of *curation logics*, which posits that actors' decision to share, or curate content is driven by the incentives they face and the social norms within which they are embedded. Our influencers are the actors who spread Twitter messages related to Derek Chauvin's trial between April 13, 2021, to April 27, 2021. The trial found Chauvin guilty on all counts of killing George Floyd and was closely followed in the Twitter discussion space of #BlackLivesMatter.

Using both qualitative and quantitative methods, this work-in-progress study shows that different levels of influencers tend to spread different kinds of messages. From this we argue that influencers face different curation logics, which impacts what they share. We also argue that Nahon and Hemsley's (2013) theory of how content goes viral could have more explanatory power with these updated concepts. This research contributes to the study of viral events on social media, the theory of curation logics and the literature around influencers

BACKGROUND

#BlackLivesMatter (BLM) began in 2013 as a collective of Black radical organizers and self-identified Black/Queer feminists, Patrisse Cullors, Alicia Garza, and Opal Tometi (Freelon et al., 2016, 2018; Gallagher et al., 2018; Stewart et al., 2017). They organized in response to the death of Trayvon Martin and George Zimmerman's acquittal (Tillery, 2019). While initial uptake of the hashtag #BlackLivesMatter was small (Freelon et al., 2016), it has grown into a large movement in the last eight years, far beyond the original organization. Not long after BLM formed, #BlueLivesMatter and #AllLivesMatter as hashtag movements quickly followed (Carney, 2016).

To understand the logics around sharing viral content in such a social movement we turn to (Thorson & Wells, 2015, 2016)'s theory of *curated flows*, which describes how people select information to share into their own networks. Curation decisions are based on the incentives and social norms that actors face. As such, we might expect that actors in #BlackLivesMatter and #BlueLivesMatter networks might face different incentives and norms. Similar to some other studies (Hemsley, 2019; Jackson, 2020), we break down the levels of influencers in our study based on their follower numbers. Specifically, our research questions are:

RQ1: Do different levels of influencers curate different types of information about #BlackLivesMatter on Twitter?

RQ2: Are there differences in message type and influencer curation for #BlackLivesMatter, #BlueLivesMatter, and #AllLivesMatter?

METHOD

We used Twitter's streaming application programming interface (API) to collect Tweets during the week before and after Derek Chauvin's trial: April 13, 2021, to April 27, 2021. We then used supervised machine learning to analyze the data. During this time Twitter had 2,634,951 tweets matching the terms BlackLivesMatter, BlueLivesMatter, and AllLivesMatter. Of these, 1,586,376 are retweets (60.21%). Using qualitative coding techniques (Elliott, 2018), three researchers worked together to first code a random sample of 200 tweets (retweets were not coded). This led to an initial emergent set of categories that the researchers felt might shed light on differences in influencer's curation behavior. We then developed 6 categories: Support, Opposition, Disruption, Informative, Reframing, and Not Relevant. Next, we coded another 1100 tweets separately and the intercoder reliability was higher than 75%. We used the 1300 tweets as a "gold label" data set to train machine learning models (Géron, 2019). An initial run using the Gradient Boosting Classifier with default parameters achieved the highest accuracy at 63%, but we plan to further improve the models using hyperparameter tuning and adding significantly more data to the gold labeled dataset. We then ran the models on all the retweets.

While we are still coding and improving the models, this work-in-progress paper presents the results we have at the moment. Having the models trained on all the kinds of messages that people sent and run on the kinds of messages that influencers retweeted helps us understand the kinds of messages that influencers are curating. Similar to what others have done (Hemsley, 2019; Jackson, 2020), we break down the levels of influencers in our study based on their follower numbers. Table 1 summarizes the different levels of influencers in this study.

Influencer levels	Follower number
Mega	$X > 1$ million
Big	$100,000 \leq X < 1$ million
Middle	$10,000 \leq X < 100,000$
Small	$1000 \leq X < 10,000$
Tiny	$X \leq 1000$

Table 1. The Influencer Levels defined in This Study

RESULTS

We used Fisher's exact test to answer RQ1, and we find that different levels of influencers do tend to curate, or retweet, different message categories (p -value < 0.0005). As can be seen in figure 1 (we excluded the Not Relevant category in figure 1 because it does not require a specific #All/Black/BlueLivesMatter group), within #BlackLivesMatter data, Mega-level influencers tended to post relatively more Informative and fewer Disruption tweets, whereas those in the Middle-level tended to post relatively more disruption tweets. Given that influencers at the mega level are often news channels, who have an incentive to appeal to the widest audience possible, and have a journalistic norm for neutrality, we expect them to avoid being disruptive, and instead to be informative, which is what we have found. Indeed, in the #BlackLivesMatter group, Mega-level influencers only curated support and informative messages, while small or tiny influencers curated content in all 5 categories. In general, except for Mega-level influencers, reframing messages were curated more frequently than informative. Tiny and Small-level influencers tended to be more similar and tended to current the highest rates of Support messages. These small and tiny level influencers are the largest groups and so probably face the widest range of norms and incentives, therefore it makes sense that they cover a wide range of content.

With respect to RQ2, we do see differences in message types and influencer curations. For example, support was the most frequently curated within #BlackLivesMatter and #AllLivesMatter, while for #BlueLivesMatter it is Opposition. #BlueLivesMatter has a much even distribution of message types that influencers curated, while #AllLivesMatter is more similar, but not as uneven, as #BlackLivesMatter. Based on the information behavior differences between different levels of influencers, we suggest that there is nuance at different levels and influencers face different curation logics (Thorson & Wells, 2015, 2016) at these different levels, which impacts what information they choose to spread.

CONCLUSION

In this work-in-progress study we aim to update Nahon and Hemsley's (2013) conception of virality by applying the framework of curated flows (Thorson & Wells, 2015, 2016) to discuss the information behaviors related to the #BlackLivesMatter social movement. Using both qualitative and quantitative methods, we find differences in different levels of influencers curation behavior within and between #All/Black/BlueLivesMatter. We offer some observations about the differences in logics (incentives and norms) that they might face. In future work we intend to expand our qualitatively coded gold label dataset, which we believe will improve the performance of our models. And while space requirements limited our ability to delve more deeply into how the concept of influencers and curation logics can update the concept of virality, we intend to follow up this theoretical work. Overall, this work-in-progress study contributes to literature about curated flows as well as the #BlackLivesMatter social movement.

REFERENCES

- Bakshy, E., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). Everyone's an influencer: Quantifying influence on twitter. *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining - WSDM '11*, 65. <https://doi.org/10/fmzn5j>
- Bennett, W. L., & Segerberg, A. (2013). *The Logic of Connective Action: Digital Media and the Personalization of Contentious Politics*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139198752>
- Carney, N. (2016). All lives matter, but so does race: Black lives matter and the evolving role of social media. *Humanity & Society*, 40(2), 180–199. <https://doi.org/10/gf3gsp>
- Elliott, V. (2018). Thinking about the Coding Process in Qualitative Data Analysis. *The Qualitative Report*, 23(11), 2850–2861.
- Freberg, K., Graham, K., McGaughey, K., & Freberg, L. A. (2011). Who are the social media influencers? A study of public perceptions of personality. *Public Relations Review*, 37(1), 90–92. <https://doi.org/10.1016/j.pubrev.2010.11.001>
- Freelon, D., McIlwain, C., & Clark, M. (2018). Quantifying the power and consequences of social media protest. *New Media & Society*, 20(3), 990–1011. <https://doi.org/10/gdfpb5>
- Freelon, D., McIlwain, C. D., & Clark, M. (2016). Beyond the hashtags:# Ferguson,# Blacklivesmatter, and the online struggle for offline justice. *Center for Media & Social Impact, American University, Forthcoming*.
- Gallagher, R. J., Reagan, A. J., Danforth, C. M., & Dodds, P. S. (2018). Divergent discourse between protests and counter-protests:# BlackLivesMatter and# AllLivesMatter. *PloS One*, 13(4), e0195644. <https://doi.org/10/gdchmm>
- Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, Inc.
- Hemsley, J. (2019). Followers Retweet! The Influence of Middle-Level Gatekeepers on the Spread of Political Information on Twitter. *Policy & Internet*, 11(3), 280–304. <https://doi.org/10/gfwnpm>
- Jackson, S. (2020). *Differentiated Information Flows: Social Media Curation Practices in the 2016 U.S. Presidential Elections*. 22.
- Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? *Proceedings of the 19th International Conference on World Wide Web*, 591–600. <https://doi.org/10.1145/1772690.1772751>
- Mundt, M., Ross, K., & Burnett, C. M. (2018). Scaling social movements through social media: The case of Black Lives Matter. *Social Media+ Society*, 4(4), 2056305118807911.
- Stewart, L. G., Arif, A., Nied, A. C., Spiro, E. S., & Starbird, K. (2017). Drawing the Lines of Contention: Networked Frame Contests Within #BlackLivesMatter Discourse. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW), 1–23. <https://doi.org/10/gf3gsk>
- Thorson, K., & Wells, C. (2015). *How Gatekeeping Still Matters: Understanding Media Effects in an Era of Curated Flows* (pp. 25–44).
- Thorson, K., & Wells, C. (2016). Curated Flows: A Framework for Mapping Media Exposure in the Digital Age: Curated Flows. *Communication Theory*, 26(3), 309–328. <https://doi.org/10/f878cg>
- Tillery, A. B. (2019). What kind of movement is Black Lives Matter? The view from Twitter. *Journal of Race, Ethnicity and Politics*, 4(2), 297–323. <https://doi.org/10/gh7smx>

Implementing Virtual Reality Training for Librarians: Supporting Patrons in Crisis

Dumas, Catherine Simmons University, USA | catherine.dumas2@simmons.edu
Williams, Rachel D. Simmons University, USA | rachel.williams@simmons.edu
Flanagan, Joanna Simmons University, USA | Joanna.flanagan@simmons.edu
Porwol, Lukasz National University of Ireland, Galway, Ireland | lukasz.porwol@insight-centre.org

ABSTRACT

This research proposes a study that is part of a larger, interdisciplinary, and multi-institutional project that examines the usability of VR training for library and information science (LIS) graduate students and professionals in gaining skills for interacting effectively with patrons in crisis. This project is entirely novel and innovative in its approach. To date, no virtual reality training for crisis management in libraries exists. Additionally, we hypothesize that our project will ultimately enable us to: reach a wider audience with training; consider the usability of VR training in LIS education; introduce training in low-stakes environments that does not pose harm to patrons in crisis; engage students in an immersive learning experience. This project can create possibilities for how we approach LIS education within our department and support training students in vital, but difficult to teach, soft skills such as empathy and communication (Saunders & Bajjaly, 2021).

KEYWORDS

virtual reality; social work education; library and information science education; crisis intervention; Mozilla Hubs

INTRODUCTION

Library staff tirelessly continue to support individuals in their communities, and those efforts have been particularly challenging in recent years, and since the onslaught of the COVID-19 pandemic. As a result, library staff face particular difficulties when it comes to balancing the needs of library patrons with taking care of themselves. A crucial component in finding that balance involves bolstering training for librarians so that they feel empowered to support patrons in crisis effectively and confidently (Ogden and Williams, 2022). Trauma-informed social work approaches that focus on well-being theory can frame and enhance librarian training and overall professional practice (Wahler et al., 2020). Recent trends in higher education emphasize the use of virtual reality for training pre-service professionals (Kourgiantakis et al., 2020; McGarr, 2020). Studies in education and social work highlight the benefits of virtual simulations, including repeated practice without risk of harm to students, increase in student motivation, ability for instructors to provide feedback, and virtual environments as a "safe space" to make mistakes (McGarr, 2020). Although the use of virtual reality (VR) in training social workers and educators has become more popular, similar approaches have not been adopted in LIS education. This study addresses this need in the context of training LIS students to interact more effectively with patrons in crisis. Our project examines the following:

- 1) How virtual reality training approaches using head-mounted displays (HMDs) and the social VR platform, Mozilla Hubs, can be implemented to support training librarians and LIS students to build communication and de-escalation skills and confidence in interacting with patrons in crisis.
- 2) The usability of virtual reality training approaches using head-mounted displays (HMDs) and the social VR platform, Mozilla Hubs, for training librarians and LIS students.

METHODOLOGY

Ten participants (six female, two male, one non-binary, one unsure (between male and non-binary), all between 20-39 years old) were recruited from the LIS school graduate program at Simmons University to take part in the study. The study consisted of a training session with each of the participants using an Oculus Quest 2 VR headset and a public library environment created by the researchers in the social VR platform, Mozilla Hubs. Each session included pre-and post-session questionnaires. The pre-questionnaire is from Herrera et al. (2018), Building long-term empathy: A large-scale comparison of traditional and virtual reality perspective-taking. The post-questionnaire is created by the researchers where some questions are based on the mental illness questionnaire developed by Cohen and Struening (1962) and other questions that address the effectiveness of training, the usability of VR headset and platform, and the potential of VR for education in library and information science. The questionnaires also address the following: demographic data; experience in crisis management and communication/de-escalation skills; and experience with different technologies. The exercises implemented in VR have been used by one of the authors during in-person and zoom-based training workshops with staff at libraries throughout the state of Massachusetts.

The outline below describes the activities for a typical session:

Participants sent a profile picture to the researchers one week before their session. The researchers created avatars for the participants using their pictures and ReadyPlayerMe, an online tool for creating avatars. The sessions were held in the researcher's lab. Prior to the session, participants completed the informed consent and pre-session questionnaire. Next, participants were given instructions on how to use the headset and enter the session room in Mozilla Hubs which was created by the researchers. The following session structure is based on learning de-escalation techniques. In the session, the participant watched a video with the instructor presenting a training session on implementing TACOS (Threaten/Argue/Challenge/Order/Shame), a set of de-escalation skills, inside the virtual environment. After instruction, the participant watched a video created by the researchers role-playing inside the virtual environment in Mozilla Hubs which researchers rendered to look like a public library (see Figures 1 and 2).



Figures 1 and 2: Screenshots of the simulated public library and the scenario video created by the researchers in Mozilla Hubs watched by the participant

Scenario: Your regular patron, Keith, is using the computer again today (he spends about an hour or two most days the library is open on the computer). You have noticed recently that Keith has started mumbling to himself while using the computer. He has also mentioned that other patrons using the computer nearby are watching him. Today, while using the computer, all of a sudden, he jumps up, pushing his chair over, and walks toward the front desk. He tells you that Laura was sent here from a different planet to kill him and that she was watching his every move, and he's sick of it.

After the close of the virtual training session and role-play, students completed the post-session questionnaires.

Additionally, since Mozilla Hubs is open source, we have created an instance that allows us to collect user behavior data from the participants (gazing, position, whether people are talking and how loud, and hand gestures). We will use artificial intelligence (AI) to train the models for data analysis. Communication and de-escalation skills involve many physical behavioral cues, which may correlate to specific user behavior data that will be collected during sessions.

RESULTS

We are currently in the process of analyzing the data. One finding is that all of the participants said that they felt more confident in their ability to work with library patrons with mental illness as a result of the training in the VR session. Additionally, all of the participants felt that teaching LIS students' skills like crisis communication in a public library in VR is effective. We will have more results to report on in time for the Annual Meeting in October 2022. Additionally, we will be bringing Oculus Quest 2 headsets to the Annual Meeting so attendees can interact with our training session and our virtual library environment in Mozilla Hubs.

CONCLUSION

Our analysis of recent trends in the literature emphasizes the potential of VR for LIS education. Our systematic literature review provides the basis for a study that examines how virtual reality training approaches may be implemented to support LIS students to build communication and de-escalation skills and confidence in interacting with patrons in crisis. Therefore, our project examined the viability and usability of VR training for this skill development using the social VR platform, Mozilla Hubs.

ACKNOWLEDGMENTS

We thank Simmons University for funding from the COCIS Innovation Grant Award that is supporting this research project. We also want to note that this study was approved by the IRB at Simmons University.

REFERENCES

Cohen, J., & Struening, E. L. (1962). Opinions about mental illness in the personnel of two large mental hospitals. *Journal of Abnormal and Social Psychology*, 64, 349–360.

- Herrera, F., Bailenson, J., Weisz, E., Ogle, E., & Zaki, J. (2018). Building long-term empathy: A large-scale comparison of traditional and virtual reality perspective-taking. *PLoS ONE*, 13(10), 1–37. <https://doi.org/10.1371/journal.pone.0204494>
- Huttar, C. M., & BrintzenhofeSzoc, K. (2020). Virtual reality and computer simulation in social work education: A systematic review. *Journal of Social Work Education*, 56(1), 131–141. <https://doi.org/10.1080/10437797.2019.1648221>.
- Khan, K., Kunz, R., Kleijnen, J., & Antes, G. (2011). *Systematic reviews to support evidence-based medicine* (2nd ed.). London, UK: Hodder & Stoughton.
- Kourgiantakis, T., Sewell, K. M., Hu, R., Logan, J., & Bogo, M. (2020). Simulation in social work education: A scoping review. *Research on Social Work Practice*, 30(4), 433–450.
- McGarr, O. (2020). The use of virtual simulations in teacher education to develop pre-service teachers' behaviour and classroom management skills: implications for reflective practice. *Journal of Education for Teaching*, 47(2), 274–286.
- Ogden, L. P., & Williams, R. D. (2022). Supporting Patrons in Crisis through a Social Work-Public Library Collaboration. *Journal of Library Administration*, 1–17. <https://doi.org/10.1080/01930826.2022.2083442>.
- Oliveira, C., Simões de Almeida, R., & Marques, A. (2021). Virtual reality in social skills training programs for people with schizophrenia: A systematic review and focus group. *British Journal of Occupational Therapy*, 84(9), 571–581.
- Porwol, L., & Dumas, C., (accepted with revisions). VR-Dialogue: An AI-supported Framework for VR Meetings Social Interaction Investigation. *Transforming Government: People, Process and Policy*.
- Saunders, L., & Bajjaly, S. (2021). The importance of soft skills to LIS education. *Journal of Education for Library and Information Science*, e20200053.
- Ticknor, B. (2017). Pilot 1.0: Creating a virtual environment for the treatment of offenders. *Corrections Today*, 79(3), 46–50.
- Ticknor, B. (2018). Using virtual reality to treat offenders: An examination. *International Journal of Criminal Justice Sciences*, 13(2), 316–325.
- Wahler, E. A., Provence, M. A., Helling, J., & Williams, M. A. (2020). The changing role of libraries: How social workers can help. *Families in Society*, 101(1), 34–43.
- Washburn, M., & Zhou, S. (2018). Teaching note—Technology-enhanced clinical simulations: Tools for practicing clinical skills in online social work programs. *Journal of Social Work Education*, 54(3), 554–560. <https://doi.org/10.1080/10437797.2017.1404519>.

Collection Development Policies for Special Collections at Selected Academic Libraries

Faulkner, James University of North Texas, USA | James.Faulkner@unt.edu

Kim, Jeonghyun University of North Texas, USA | Jeonghyun.Kim@unt.edu

ABSTRACT

Collection development policies for university special collections and archives guide archival appraisal, which is the process by which archivists determine the value of an item. Much has been written about appraisal theory, but little is known about archivists' actual appraisal practice and how collection policy guides such practice. The object of this paper is to explore the appraisal practices of archivists by examining collection development policies. As a preliminary examination, this study analyzed the collection development policies of seven Texas tier-one institutions. The most common component of the policies was a statement on the types of items collected or excluded. The study also found that the language used for these formats varied considerably with no standardization. The policies themselves also varied widely in detail.

KEYWORDS

University special collections, Appraisal, Selection, Collection development policy

INTRODUCTION

Special collections are departments of libraries that house "material in any format, such as rare books, manuscripts, photographs, institutional archives, that are generally characterized by their artifactual or monetary value, physical format, uniqueness or rarity, and/or an institutional commitment to long-term preservation and access" (Dooley & Luce, 2010, p. 16). The items in such collections often have restricted access and are not available in open stacks because of their uniqueness and rarity. They are regarded as special importance to the community that the library serves. For instance, special collections in academic libraries support the academic curriculum and research needs unique to the institution (American Library Association, 2017). It should be noted that university archives are another type of archive that deals with collecting items related to a university's history. University archives and university special collection and are often seen together within academic libraries.

Appraisal is an intellectual process carried out by archivists who determine what items have value and will be preserved in the archive. It is the process that a collection development policy is meant to assist. There has been much discussion about the archival theory of appraisal, but little examination of the archivist's actual appraisal practice and process has taken place (Marshall, 2002). Further, how collection development policies guide process and decision-making related to appraisal is not well documented.

In the case of academic libraries' special collections, a survey conducted by the Society of American Archivists (2017) found university archives and university special collections were more likely than other types of archives to have a collection policy and to have made that policy publicly available. However, how well academic libraries have implemented their collection development policy in their appraisal process and how much depth and breadth of information is being provided in the policy have not been much discussed in the literature. As university special collections are becoming recognized as a valuable scholarly and pedagogical resource in higher education institutions, it is worthwhile to examine such policies to better understand how appraisal affects the formation of collections.

METHODOLOGY

As a preliminary investigation, the sample of special collections from tier-one research universities in Texas was used in this study. Tier-one research universities were chosen because it is believed that they are likely to have a separate department for special collections and higher number of staff/funding, resulting in them being more likely to have established policies.

The initial search began by navigating to the university library special collections website and searching for a collection development policy. A further search was performed on Google using special collection* AND (collections development policy OR collection policy OR collecting policy OR appraisal policy) AND site: institution's URL. We found that some institutions have multiple policies that serve different types of materials in their special collection; in this case, different policies were combined into one document and treated/analyzed as a single policy for coding.

QDA miner was used for data organization and coding. Coding began by using the criteria originally established by Philips (1984), who developed the criteria to aid in the creation of a collections development policy for a manuscript

collection. In an iterative process, such criteria as initial categories were applied to the data, and then refined as new categories rose. Some of the subcategories, such as the different types of users, were combined for simplicity.

RESULTS

Of the 11 academic libraries targeted in this study, seven had a policy that was located.

Purpose of the institution and/or collection

Of a total of seven policies, a statement of purpose for the institution or collection was in 36% of the policies. The type of program supported, and the clientele served were stated in the same number of the policies. Most institutions emphasize the needs of their community users, including students, faculty, scholars, and other researchers. Such needs include research, instructional, and reference purposes.

Priorities and limitations of the collections

Overall, collection development policies examined in this study addressed the breadth and depth of the collection's focus. All the policies located mentioned either the type of the items they collect or exclude. The levels of collection intensities (e.g., minimal, basic, instructional, research, comprehensive) appeared in 28% of the policies.

The subject areas collected were described in 36% and were broad from Mexican Americans and LGBTQ communities to Texas history and military history. The languages collected were described in 28% of the policies; most institutions extensively collect English and Spanish language resources. Chronological periods and geographic areas covered by the collection were mentioned in 28% of the policies respectively. The emphasis of those policies is often on the city where the institution is located, Southwest United States, and Mexico, but other geographic areas may be considered, especially if items have significant research value. Statements about the inclusion of university archives, which typically include both official university records and materials regarding the university's history donated from students, faculty, and alumni, appeared in 28% of the policies.

Formats collected in

The format items were categorized into several groups. The most common statement about the format was an indefinite statement occurring in 85% of the policies. These indefinite statements took the form of any and other; this is a general list but not limited to these formats. As for format type, still image, such as photographs and maps, as well as text, such as digitized rare books and oral history transcripts, are popular formats collected in academic library special collections as they appeared in all the policies examined in this study. Sound records and moving images were included in 85% and 71% of policies, respectively. Physical object, such as memorabilia, appeared in 57% of the policies. Interactive resource like websites was located in 42%.

Acquisition, deaccession, and outreach

Almost half of the policies described the acquisition method, either through accepting donations or purchasing items. Deaccession policy, which describes how items will be removed from the collection, was detailed in 28% of the policies. Only one has a statement about cooperative agreements with other institutions. Not mentioned in any of the policies were statements of resource sharing, procedures affecting the policy, and procedures for monitoring the policy.

DISCUSSION

This study found that the collection development policies varied widely in detail. It is clear that the components of a collection development policy are not standardized. The lack of standardization makes comparisons and evaluations of appraisal decisions difficult. Similarly, the terms used for formats varied dramatically. More importantly, the format lists appear partial, and the inclusion of indefinite terms does not appear to truly limit the formats collected. Before categorization, many of the formats were not actually formats, such as ephemera, or were vague, such as removable electronic media. With such variance in the policies, are archivists' actual practices as dissimilar?

This study also found four institutions do not have publicly accessible collection development policies; it was assumed they may not have a policy or could be using a policy for internal purposes only. However, the number of policies found online in this study is more than in previous studies (e.g., Marshall, 2002; Wink, 2010). Additional studies will expand upon the number of university special collections examined and contact universities without policies online.

CONCLUSION

This study sought to explore the appraisal practices of university special collection archivists by examining their collection development policies posted online. There were considerable variations between the number of policy criteria listed within the collection policies of the university special collections. The results from this study have important implications for literature regarding the role of collection development policy in appraisal practice and process.

REFERENCES

- American Library Association (2017). *Special collections*.
<http://www.ala.org/tools/challengesupport/selectionpolicytoolkit/special>
- Dooley, J. M., & Luce, K. (2010). Taking our pulse: The OCLC Research survey of special collections and archives. Dublin, OH: OCLC Research. <https://doi.org/10.25333/gj2r-8172>
- Marshal, J. A. (2002). Toward common content: An analysis of online college and university collecting policies. *The American Archivist*, 65(2), 231-256. <https://doi.org/10.17723/aarc.65.2.d14g7x2615270j61>
- Phillips, F. (1984). Developing collecting policies for manuscript collections. *The American Archivist*, 47(1), 30-42. Society of American Archivists. (n.d.). Collecting policy. In *Dictionary of archives terminology*.
<https://dictionary.archivists.org/entry/collecting-policy.html>
- Society of American Archivists. (2017). *Collection development policy survey*.
https://www2.archivists.org/sites/all/files/aasection_collection_development_policy_2016survey.pdf
- Wink, T. (2010). Archival collection development policies: A study of their content and collaborative aims.
https://cdr.lib.unc.edu/concern/masters_papers/nc580r466?locale=en

Predicting Search Task Difficulty through a Discrete-Time Action Log Representation on Spectrum Kernel

Gacitúa, Daniel

Universidad de Santiago de Chile, Chile | daniel.gacitua@usach.cl

González-Ibáñez, Roberto

Universidad de Santiago de Chile, Chile | roberto.gonzalez.i@usach.cl

ABSTRACT

Predicting perceived difficulty on a web search task is an open problem in the interactive information retrieval field. A common approach to tackle it, is through features obtained from full search sessions, which are then used to train classification models. In this poster we attempt to predict perceived task difficulty at different stages of the search process. To do so, we use the spectrum kernel for support vector machine (SVM) classification. Our preliminary results suggest that by using behavioral data from the first query segment, it is possible to provide timely classifications of whether a search task is perceived as hard or easy.

KEYWORDS

Search task difficulty; search behaviors; web search

INTRODUCTION AND BACKGROUND

Task difficulty is a well-known facet of search tasks (Li & Belkin, 2008). Due to its subjective nature (i.e., it depends on the searcher), it has been studied through users' behavioral data. Some have faced the problem as a prediction task (e.g., Arguello, 2014; Liu et al., 2014) and have tackled it through machine learning models such as logistic regression and decision trees. As inputs for training such models, feature vectors of aggregated numeric variables (e.g., action counts, means, dwell times, etc.) have been used. A major problem with aggregated data to express a given feature is the loss of dynamic aspects of the search process. For instance, the number of clicks does not indicate when and where each individual click was performed. Likewise, query length does not reveal how the query was formulated in terms of pauses and query editing. Unlike previous approaches, in this poster, we attempt to predict task difficulty by using a representation of users' actions capable of expressing the dynamics of the search process itself. More specifically, we use the spectrum kernel for support vector machines (SVM) (Leslie et al., 2002). To use this kernel, all user actions are expressed as discrete-time character sequences (we refer to such sequences as action strings). Such micro-level actions could act as proxies of cognitive and metacognitive processes (Mostafa & Gwizdka, 2016).

METHOD

We conducted a user study following a rotational design. We recruited 63 participants (68.25% men, and 31.75% women). Their ages ranged between 18 and 63 years old ($M=22.02$, $SD=6.73$). All participants in this study signed a consent form. The study was performed online using a system called LETICIA. This system offers a simulated web environment, a search engine, and a controlled collection of Web documents. In addition, the system controls the study protocol and logs users' actions (e.g., keystrokes, mouse actions, visited links, queries, etc.). Within the system, a user can formulate queries, explore search results, and bookmark those that are found to be relevant for a given task. Regarding the latter, participants faced three tasks contextualized in different scenarios. In each task participants were required to search for information to answer the following questions: (Task 1) How does quantum computing work? and what are its potential uses?, (Task 2) Can black holes "swallow the Universe"? and what happens inside a black hole?; and (Task 3) What happens when we dissolve salt in hot water (compared to cold water)?

After showing the search task description to participants, they filled out a pre-task questionnaire to self-report their perceived task difficulty according to five dimensions (Capra et al., 2015): (1) Building search queries, (2) understanding search results, (3) establishing if results were useful, (4) determining when to stop searching, and (5) overall difficulty. Answers were given on a 6-point Likert scale. Then, the perceived difficulty for each user and task was determined by the rounded mean of the answers. Finally, the resulting score was coded as follows: scores of 3 or less were coded as "easy", whereas those higher than 3 were coded as "hard." After the pre-task questionnaire, participants were instructed to search for information, and bookmark at least three relevant documents in order to continue with the following search tasks. Overall, participants were given 25 minutes to complete all search tasks.

DATA PROCESSING

Action logs were segmented by user and query segment. A query segment includes all user actions from the start of a query formulation up to before the next query formulation. We coded each user action as an individual character (Table 1). Then, with the coded data we built action strings for each query segment. Each action string had two variants: Standard (considering all actions) and simplified (multiple consecutive occurrences of the same action

were replaced by a single character). For instance, for the following standard action string “ECMCCFCWWWWWWW_WWWQ” its simplified version would be “ECMFCW_WQ”.

S	Scroll Page	B	Bookmark Page
C	Mouse Click	T	Timeout Triggered
E	Page Enter	R	Finish Search Task
X	Page Exit	P	Long Pause Between Keystrokes (>1.2 s)
P	Use SERP Pagination	p	Short Pause Between Keystrokes (0.5-1.2 s)
F	Query Box Focus	_	Space Keystroke
f	Query Box Blur	Y	Enter or Tab Keystroke
Q	Search Query Sent	D	Backspace Keystroke
H	Search Result Click	d	Delete Keystroke
M	Show Task Description	W	Alphanumeric Keystroke

Table 1. User actions logged during the study and their corresponding character codes

Next, we added to each action string the corresponding class (hard or easy) according to the perceived task difficulty obtained during the pre-task questionnaire. Since the resulting dataset was unbalanced (115 query segments came from an “easy” task and 84 came from a “hard” one), an oversampling technique was applied (Mohammed et al., 2020). Then, this dataset was used to train an SVM model with the spectrum string kernel, using 10-fold cross-validation. The area under the ROC curve (AUC) was used as the evaluation measure for classification performance.

RESULTS

Classification results are described in Table 2, considering both the complete session (all) and the first query segment (1qs) for each task performed by the participants.

Dataset	AUC (unbalanced)	AUC (oversampled)
1qs-simplified	0.605	0.758
1qs-standard	0.564	0.621
all-simplified	0.586	0.607
all-standard	0.506	0.572

Table 2. Classification results of action strings using the spectrum string kernel

CONCLUSION AND FUTURE WORK

The results obtained from the classification models show that the best performance is obtained at the first query segment with a simplified action string. It is also worth noting that the oversampling technique enhanced predicting performance. Implications of this preliminary work may offer new insights into the timely support that can be offered to searchers depending on their task difficulty perception. Such support could be expressed in various forms including search assistants, reorganization of search results, and query suggestions, to name a few. Future work will be focused on fine-tuning these models and analyzing the search patterns formed in the action strings.

ACKNOWLEDGMENTS

This work is partially supported by the National Agency for Research and Development (ANID) Scholarship Program Doctorado Nacional no. 2019-21192162 and the TUTELAGE project funded by ANID (FONDECYT Regular, grant no. 1201610).

REFERENCES

- Arguello, J. (2014). Predicting search task difficulty. In *European Conference on Information Retrieval* (pp. 88-99). Springer, Cham.
- Capra, R., Arguello, J., Crescenzi, A., & Vardell, E. (2015). Differences in the use of search assistance for tasks of varying complexity. In *Proceedings of the 38th international acm sigir conference on research and development in information retrieval* (pp. 23-32).

- Leslie, C., Eskin, E., & Noble, W. S. (2001). The spectrum kernel: A string kernel for SVM protein classification. In *Biocomputing 2002* (pp. 564-575).
- Li, Y., & Belkin, N. J. (2008). A faceted approach to conceptualizing tasks in information seeking. *Information processing & management*, 44(6), 1822-1837.
- Liu, C., Liu, J., & Belkin, N. J. (2014). Predicting search task difficulty at different search stages. In *Proceedings of the 23rd ACM international conference on conference on information and knowledge management* (pp. 569-578).
- Mohammed, R., Rawashdeh, J., & Abdullah, M. (2020). Machine learning with oversampling and undersampling techniques: overview study and experimental results. In *2020 11th international conference on information and communication systems (ICICS)* (pp. 243-248). IEEE.
- Mostafa, J., & Gwizdka, J. (2016). Deepening the role of the user: Neuro-physiological evidence as a basis for studying and improving search. In *Proceedings of the 2016 acm on conference on human information interaction and retrieval* (pp. 63-70).

COVID-19 Tweet Links: A Preliminary Investigation of Type and Relevance

Goh, Dion Hoe-Lian Nanyang Technological University, Singapore | ashlgoh@ntu.edu.sg
Lee, Chei Sian Nanyang Technological University, Singapore | leecs@ntu.edu.sg
Zheng, Han Nanyang Technological University, Singapore | han019@ntu.edu.sg
Theng, Yin Leng Nanyang Technological University, Singapore | tyltheng@ntu.edu.sg

ABSTRACT

We conducted an exploratory study of the links found in Twitter tweets. Our results showed that the largest category of tweet links was social media platforms followed by alternative news sites. Government agencies and educational institutions were under-represented. In terms of relevance, about 75% of the links were related to COVID-19 but disappointingly, only 40% of the links were directly related to their respective tweets' topics.

KEYWORDS

COVID-19, pandemic, Twitter, tweets, links, type, relevance

INTRODUCTION

During the COVID-19 pandemic, users have generated an immense quantity of social media content. Twitter tweets, in particular, have often been used for analysis because they are regarded as a valuable source of data in the surveillance of diseases (e.g. Sinnenberg, et al., 2017). More importantly, Twitter is in active use by governments, organizations and individuals for the sharing of COVID-19 information.

The restricted length of a tweet is a double-edged sword. On the one hand, information is packaged into small chunks that can be quickly read. On the other, there is a limit to the amount of information each tweet can convey. Consequently, links are typically embedded, presumably to lead users to Websites with elaborated content or as a citation source. However, it is evident in many contexts that not all links serve these purposes. For example, links may be clickbait or serve other malicious purposes (Jamison, Broniatowski, & Quinn, 2019).

In the COVID-19 context, tweet links play an important role in the dissemination of public health information. However, if misused, fake news and other forms of disinformation may be spread instead. Surprisingly, there is, to our knowledge, little current work that has investigated the types and relevance of tweet links found about COVID-19. The present study thus aims to answer the following questions: (1) What types of Websites do the tweet links about COVID-19 point to? (2) Are these Websites relevant to COVID-19 and their respective tweet topics?

METHODOLOGY

Data were drawn from a project that collected COVID-19-related tweets (Chen, Lerman, & Ferrara, 2020). We selected a date range of one week, between 9 to 16 November 2020, comprising 189,071 tweets. This period was chosen as it represented one week after Pfizer and BioNTech announced the first vaccine efficacy results (Business Wire, 2020), providing hope, controversy, and plenty of online chatter. Next, tweets without links were discarded, resulting in a remainder of 56,577. Following this, a random sample of 4001 tweets was selected. This random sampling approach is consistent with prior public health research using tweets such as Cavazos-Rehg et al. (2016). Finally, for each of the 4001 tweets, its link was followed by three researchers independently. The link was checked to see if it was still accessible. If so, the type of Website was noted. Next, the link's content (Website) was read and ascertained if it was relevant to the tweet's topic. Once all the tweets were analyzed, the researchers discussed their respective results to resolve differences until consensus was reached.

RESULTS

Our analysis showed that the largest category of tweet links was social media platforms such as Facebook, accounting for 2032 (51%) links. Next came alternative news sites at 614 (15%) links, followed by mainstream news sites at 531 links (13%). At the other end of the spectrum, there were only 62 government Websites (1.5%), with online stores and educational institutions having the smallest share at 27 (0.7%) links. Approximately 7% or 285 links were not accessible. Table 1 shows the distribution of links by type of Website.

To answer the second question on relevance, the final row of Table 1 shows that of the 4001 links in our dataset, 3010 (75%) were relevant to COVID-19. However, there were only 1597 (40%) links that were directly relevant to their respective tweets' topics, meaning that the majority of links were not topically relevant.

Delving deeper into each category, the top three link categories that were most relevant to COVID-19 were mainstream news, alternative news and educational institution Websites. The three least relevant categories were

commercial Websites, online stores, and those that were not classifiable. Link categories that were most relevant to their respective tweets' topics were educational institutions, mainstream news and government-related Websites. Conversely, those that were not relevant were online stores, databases and those that were not classifiable.

Type	All Tweets		Relevant to COVID-19		Relevant to Tweet	
	Number	Percentage	Number	Percentage	Number	Percentage
Social media	2032	50.79	1597	78.59	760	37.40
Alternative news	614	15.35	554	90.23	353	57.49
Mainstream news	531	13.27	506	95.29	309	58.19
Commercial	113	2.82	72	63.72	31	27.43
Government	62	1.55	52	83.87	36	58.06
Non-profit	52	1.30	40	76.92	18	34.62
Databases	43	1.07	34	79.07	10	23.26
Personal website	30	0.75	24	80.00	17	56.67
Online store	27	0.67	7	25.93	4	14.81
Educational institution	27	0.67	24	88.89	16	59.86
Other	185	4.62	100	54.05	43	23.24
Not accessible	285	7.12	-	-	-	-
Total	4001	100	3010	75.23	1597	39.92

Table 1. Distribution of links by type of Website

DISCUSSION AND CONCLUSION

Our findings show that while there was a variety of link types posted in our dataset of COVID-19 tweets, their distribution was uneven. In particular, we were surprised to see that slightly more than half of the links pointed to social media content. Unfortunately, sites where credible information could presumably be found, such as mainstream news, governmental organizations and educational institutions were not well-represented. Collectively, these three categories comprised only about 15% of all links analyzed. Our results reinforce the prevailing notion that social media is a major source of news.

While it was heartening to see that the majority of tweet links (approximately 75%) were relevant to COVID-19, this also meant that around a quarter were unrelated, which is not an insignificant number. Alarming, when we checked if the links were both relevant to COVID-19 and its respective tweet topic, slightly less than 40% met this criteria. Taken together, the under-representation of certain categories of information sources together with the large proportion of non-relevant links may mean that people who use social media as a platform for information seeking may at best, not meet their information needs effectively, and at worst, fall prey to misinformation.

One important implication from the user's perspective is the need to be vigilant about links in online content, such as tweets. Users may exploit a hot topic, such as COVID-19, to advance an agenda, whether it is spam, advertising, or other deleterious aims. Admittedly, this may be challenging as URL shorteners are often used, masking the actual Website being referenced. The relative lack of links from official or more credible sources such as government agencies, educational institutions and mainstream news is another concern. Possible reasons include a lack of interest in posting such links by users, low levels of engagement by these organizations, or there are simply more people keen on posting other types of links. This requires further investigation.

There are limitations in the present study that warrant future work. Due to the large volume of data, we did not investigate the quality of the Websites being linked. As well, we were not able to analyze all the tweets during the study's date range. Additionally, more recent time periods may yield different outcomes. Future work may also consider analyzing the links of other social media platforms.

ACKNOWLEDGMENTS

The authors would like to thank Tongxin Hou, Ying Li, Fei Tang, Jinghua Wang, Jingxia Xu, Yanxin Zhan for their assistance with data collection and analysis.

REFERENCES

- Business Wire. (2020, November 9). Pfizer and BioNTech Announce Vaccine Candidate Against COVID-19 Achieved Success in First Interim Analysis from Phase 3 Study. <https://www.businesswire.com/news/home/20201109005539/en/%C2%A0Pfizer-and-BioNTech-Announce-Vaccine-Candidate-Against-COVID-19-Achieved-Success-in-First-Interim-Analysis-from-Phase-3-Study>
- Cavazos-Rehg, P. A., Krauss, M. J., Sowles, S., Connolly, S., Rosas, C., Bharadwaj, M., & Bierut, L. J. (2016). A content analysis of depression-related tweets. *Computers in human behavior*, *54*, 351-357.
- Chen, E., Lerman, K., & Ferrara, E. (2020). Covid-19: The first public coronavirus twitter dataset. *arXiv preprint arXiv:2003.07372*.
- Jamison, A.M., Broniatowski, D.A., & Quinn, S.C. (2019). Malicious actors on Twitter: A guide for public health researchers. *American Journal of Public Health*, *109*(5), 688-692.
- Sinnenberg, L., Buttenheim, A.M., Padrez, K., Mancheno, C., Ungar, L., & Merchant, R.M. (2017). Twitter as a tool for health research: A Systematic Review. *American Journal of Public Health*, *107*(1), e1-e8.

Sensitively Describing Objects Made to Represent Humans: A Case Study of Dolls at the Strong National Museum of Play

Graf, Ann Simmons University, USA | ann.graf@simmons.edu

ABSTRACT

This poster will present early findings on a study of 188 records for dolls at the Strong National Museum of Play in Rochester, New York, examined for evidence of descriptive practices applied to objects representing non-white humans. The work represented herein is preliminary in nature and will provide a model for further, in-depth research on how catalogers approach description of objects that are not human, yet that represent humans real or imaginary. While much research has been carried out on culturally insensitive description and subject headings applied to racial, cultural, and other human groups (Drabinski, 2013; Roberto, 2008; Watson, 2021), there is a lack of research that examines the same kinds of description applied to objects that represent humans.

KEYWORDS

Museums; cataloging; knowledge organization; cultural sensitivity

INTRODUCTION

While much research has been carried out on culturally insensitive description and subject headings applied to racial, cultural, and other human groups ((Drabinski, 2013; Roberto, 2008; Watson, 2021), there is a lack of research that examines the same kinds of description applied to objects that represent humans, for instance dolls. Dolls depicting people of color were much less common in the United States during the first half of the last century, especially from commercial manufacturers. When available commercially, they were nearly always designed from a Western and white point of view, with a sense of exception to the white norm and an often othering and even demeaning presentation (see Ideal Manufacturing Company, 1896). Dolls representing people of color continue to be produced in seemingly smaller numbers than white dolls, though this gap is narrowing. I have examined historical commercial product catalogs from a few of the most popular and commercially successful doll manufacturers in the United States to assess availability of non-white dolls and how they were historically described. I have also interviewed a prominent black doll collector and online museum curator in the United States and a curator of dolls at a children's museum in Boston to discuss further issues surrounding their description, including the challenges inherent in describing non-white dolls in collection databases without furthering historical insensitivities while providing accurate description for recordkeeping, retrieval, and research purposes.

It must be stated at the outset that the challenges of creating sensitive and useful catalog records for manufactured objects, whether representing humans or not, is complex. There will be the necessity of recording the names given to objects by their creators or manufacturers, or taken from historical catalogs, that may not have been seen as offensive at the time by those in a position to control such things. These historical positions of power of course did not often respect the voices of the cultures that were represented by or produced the objects (Turner 2015, Brown 2010). An example of this comes from a conversation with a curator at the Boston Children's Museum while examining a black cloth doll from the late 19th to early 20th century. The doll was given to the museum and documentation that came with the doll used the word "mammy" to describe the person portrayed by the doll. This term implies a caricature of black women from the Civil War and post-Civil War era onward. This word may be found in a free text descriptive field in the record for this doll, but would not be the chosen term used by catalogers today to describe it [Farkas, personal communication]. Museums seek to preserve the history and provenance of an object, and any related textual information, yet also to apply appropriate and culturally sensitive descriptive terminology that will make the object findable in modern discovery systems. The purpose of this research is to examine the state of description of human-like objects and to create awareness within the museum and archives community of any sensitivity issues surrounding in dealing with cultural artefacts, with hope for responsive change if warranted.

THE STRONG NATIONAL MUSEUM OF PLAY

These considerations led me explore the records of a large museum collection of dolls, which I found available at the Strong National Museum of Play (hereinafter, the Strong) in Rochester, New York. The "Dolls" collection at the Strong is searchable online and currently contains 4,879 objects that are named as dolls (Dolls n.d.). The Strong's collection website states:

The Strong owns and cares for the world's most comprehensive collection of toys, dolls, board games, video games, other electronic games, books, documents, and other historical materials related to play. This

unprecedented assemblage offers a unique interpretive and educational window into the critical role of play in human physical, social, and intellectual development and the ways in which play reflects cultural history (Dolls n.d.).

This large collection includes dolls from the 19th century to the present. The size of the collection and the availability of complete catalog records makes this an ideal test bed for an initial examination of descriptive practices applied to objects that represent non-white humans.

THE DATA

Working with a curator at the Strong, I was provided with descriptive catalog records for dolls in their collection that are part of their Diversity & Inclusion Initiative (188 records) and/or were tagged with the term “ethnicity” (923 records). Records were limited in this way for initial examination due to the large number of dolls in their collection. For this preliminary study, I have chosen to focus on the 188 records in the Diversity & Inclusion Initiative. These were downloaded in CSV format and imported into an Excel workbook for easier manipulation. All records were examined to determine trends and idiosyncrasies in descriptive terminology applied in free text description fields, title fields, and subject fields.

CONCLUSION

In this poster, I will discuss museum description of dolls over roughly the last century-and-a-half in the United States with particular attention to 188 non-white dolls within the Diversity & Inclusion Initiative at the Strong. I will focus on the challenges that institutional curators face when designing and maintaining appropriate, efficient, and sensitive records for these objects often created as playthings for children and that represent actual humans. The results of this methodology and analysis will be shared in the poster, as well as a response to the effectiveness of the methodology applied to the first, smaller group of doll records at the Strong.

Testing the methodology described herein on a smaller collection of records will afford valuable insight to modify processes as needed before applying analysis to the second, much larger collection of records. This work is valuable to the field of library and information science by revealing how a historically dominant culture may easily perpetuate already well-researched cultural insensitivities in documentation practice for humans onto objects representing humans. It is hoped that the research will be valuable to create awareness of possible issues and to inform broader, inclusive, and respectful cataloging practices going forward.

REFERENCES

- Brown, B. (2010). Objects, Others, and Us (The Refabrication of Things). *Critical Inquiry*, 36(Winter), 183-207.
- Dolls. (n.d.). The Strong National Museum of Play. Retrieved from (https://onlinecollection.museumofplay.org/ArgusNet/Portal/Public.aspx?_gl=1*r4judo*_ga*MTk1NDQxMzg2Ni4xNjUzNTk3NjY5*_ga_NNDS4KF8SX*MTY1NDcyMDcwNy4zLjEuMTY1NDcyMDcxMy4w&_ga=2.6279542.475446128.1654720708-1954413866.1653597669&lang=en-US).
- Drabinski, E. (2013). Queering the Catalog: Queer Theory and the Politics of Correction. *Library Quarterly: Information, Community, Policy*, 83(2), 94-111.
- Ideal Manufacturing Company. (1896). *Supplement to Catalogue B of Ideal Toys*. Detroit, Mich.: Ideal Manufacturing Company. Retrieved from (https://museumofplay.access.preservica.com/uncategorized/IO_0945a115-e0c8-4a89-be6b-797d199e5433/).
- Pilgrim, D. (2012). The Mammy Caricature. Ferris State University. Retrieved from (<https://www.ferris.edu/HTMLS/news/jimcrow/mammies/homepage.htm>).
- Roberto, K.R., ed. (2008). *Radical Cataloging: Essays at the Front*. Jefferson, NC: McFarland & Company.
- Turner, H. (2015). Decolonizing Ethnographic Documentation: A Critical History of the Early Museum Catalogs at the Smithsonian’s National Museum of Natural History. *Cataloging & Classification Quarterly*, 53(5/6), 658-676.
- Watson, B.M. (2021). Advancing Equitable Cataloging. In *Vol. 8 (2021): Proceedings from North American Symposium on Knowledge Organization*. Retrieved from (<https://journals.lib.washington.edu/index.php/nasko/article/view/15887>).

Uncover Marginalized Narratives of Japanese American Incarceration: An Annotation Scheme for Natural Language Processing and Data Analytics

Han, Yingying University of Illinois, Urbana-Champaign, USA | yh17@illinois.edu
Maganti, Roopesh University of North Texas, USA | roopeshmaganti@my.unt.edu
Chen, Jiangping University of North Texas, USA | jiangping.chen@unt.edu
Chen, Haihua University of North Texas, USA | haihua.chen@unt.edu

ABSTRACT

Around 120,000 people of Japanese ancestry was forced to remove into internment camps in the United States during World War II. Densho Digital Repository curates a collection of oral histories, which mainly includes 904 filmed interviews about “Japanese American incarceration experience from those who lived it”. This study uses web scraping techniques to collect 904 narrators’ bio information and analyzes their demographic information, including race, age, and location. Nisei counts the most of the Japanese American narrators (78.82% of all narrators). 279 (39.35%) of narrators were minors (born between 1926 and 1942) in the year 1942, when the incarceration happened, with over half of them (58.78%) being teenagers (age between 12 and 17). The majority of narrators (82.09% of the total) are from the west coast, namely, Washington State, Oregon State, or California State. This study also constructs a scheme for incarceration oral history, based on which the incarceration historical texts can be manually or/and automatically processed. The annotation scheme includes seven categories: (1) pre-war background to the incarceration; (2) government’s decision to remove ethnic Japanese; (3) life after removal and during the incarceration; (4) military services; (5) returning of ethnic Japanese after WWII; (6) legal challenges; (7) redress movement.

KEYWORDS

Japanese American incarceration; Digital archives; Natural language processing; Annotation scheme

INTRODUCTION

Ten weeks after the Japanese attack on Pearl Harbor on December 7, 1941, President of the United States, Franklin D. Roosevelt, issued Executive Order 9066, which resulted in the incarceration of approximately 120,000 people of Japanese ancestry. Two-thirds of those forced to live in the military camps were U.S. citizens. The government investigation in the 1980s concluded that people of Japanese ancestry in the United States had suffered a grave injustice. As memories of those who live the history fade, a nonprofit organization, Densho, was founded to preserve the history of Japanese American Incarceration, primarily through videotaped oral history testimonies from Japanese Americans (Densho, n.d. a). Densho developed Densho Digital Repository (DDR) to “educate, promote and advance the ideals of democracy and encourage civic engagement” (Beckman & Froh, 2018). DDR curates a collection of oral histories, including 904 filmed interviews about the “Japanese American incarceration experience from those who lived it” (Densho, n.d. b). All the interviews have been fully transcribed and segmented into sub-topics for ease of viewing. Meanwhile, the collection also provides demographic information of narrators. More researchers are turning to computation approaches such as text mining, network analysis, and data visualization to advance their studies and look for invisible patterns within large-scale datasets.

A High-quality and domain-specific corpus for computation analysis requires a semi-structured initial dataset with a well-defined annotation scheme (Chen, 2022). Corpus annotation, which can be done manually or automatically, is a process of adding interpretive information into a collection of texts to enrich a corpus (Hovy & Lavid, 2010). An annotation scheme is a codebook that defines the annotation categories and the annotation guidelines (Hovy & Lavid, 2010). A feasible strategy to computationally process and mine the texts, according to Chen (2022), is first to adopt an annotation scheme and annotate a small amount of high-quality initial data, and then augment data using a semi-supervised learning algorithm on the large-scale unlabeled data. This study uses DDR’s oral history collection as our corpus and has two contributions: (1) To uncover the marginalized history of Japanese American Incarceration, this study first explores the background of those who have lived through the history and shared their narratives. Specifically, this study analyzes the narrators’ demographic information, including their nationality, race, age, and geographic location. (2) We propose a fine-grained annotation scheme for Japanese incarceration oral history, based on which the incarceration historical texts can be manually or/and automatically processed.

DATA PREPARATION AND ANALYSIS

We use web scraping techniques to extract 904 oral history transcripts from the DDR website. Considering the variety of data elements in each web page, such as paragraph and list elements, we use the python Beautiful Soup

85th Annual Meeting of the Association for Information Science & Technology | Oct. 29 – Nov. 1, 2022 | Pittsburgh, PA. Author(s) retain copyright, but ASIS&T receives an exclusive publication license.

package to extract the data. We create data soup for each narrator's web page, after which we iterate, fetch all interview transcripts and store them in JSON format. We also extract the race, year of birth, location of the birth of the narrator using spacy phrase matching technique from narrators' bio page. We then analyze the text data obtained by splitting the tokens. Not every bio includes all the demographic information. We extract the race information from 846 narrators, the year of birth information from 709 narrators, and the geographic location information from 832 narrators.

FINDINGS

Three of the 850 narrators self-identify themselves as Japanese, and 847 narrators self-identify as Americans. Among the 847 Americans, one narrator does not indicate her race. 797 (94.21% of all American narrators) are Asian Americans, and 49 (5.79% of all American narrators) are white. Among all the 797 Asian American narrators, 796 are Japanese Americans, and Nisei (second-generation Japanese American) counts the majority of the Japanese American narrators (670, 78.82% of all the narrators, 84.17% of Japanese American narrators). 709 narrators indicate their year of birth. 389 (54.87%) narrators were adults in 1942, when the Japanese American Internment happened. 41.47% were young adults (age 18-25), and 13.4% were adults (aged between 26-50, born between 1893-1925). Meanwhile, 279 (39.35%) of narrators were minors (born between 1926 and 1942) in the year 1942, with over half of them (58.78%) being teenagers (age between 12 and 17). 832 narrators indicate their hometown information. 683 (82.09%) of them are from the west coast, namely, Washington State, Oregon State, or California State, and 42 (5%) are from Hawaii.

We create an annotation scheme with seven categories, which is adapted from the contents of two classic historical books about Japanese American incarceration. The books are Wendy Ng's *Japanese American Internment During World War II: a History and Reference Guide* and Greg Robinson's *A Tragedy of Democracy: Japanese Confinement in North America*. Both Ng and Robinson demonstrate the history of Japanese American internment from the following perspectives: (1) Pre-war background to the incarceration; (2) Life after removal and during the incarceration; (3) Military services; (4) Legal challenges; (5) Redress movement. We include these five categories in our scheme. Ng's book also introduces the returning of ethnic Japanese from camps after WWII, and we include it as a category of our scheme. Robinson's unique perspective on how the government decided to remove ethnic Japanese is also included in our scheme. Figure 1 presents a more detailed explanation of each category with definitions. We also demonstrate a transcript example of each category, which will help future researchers to understand how each category can be annotated using the oral history transcripts.

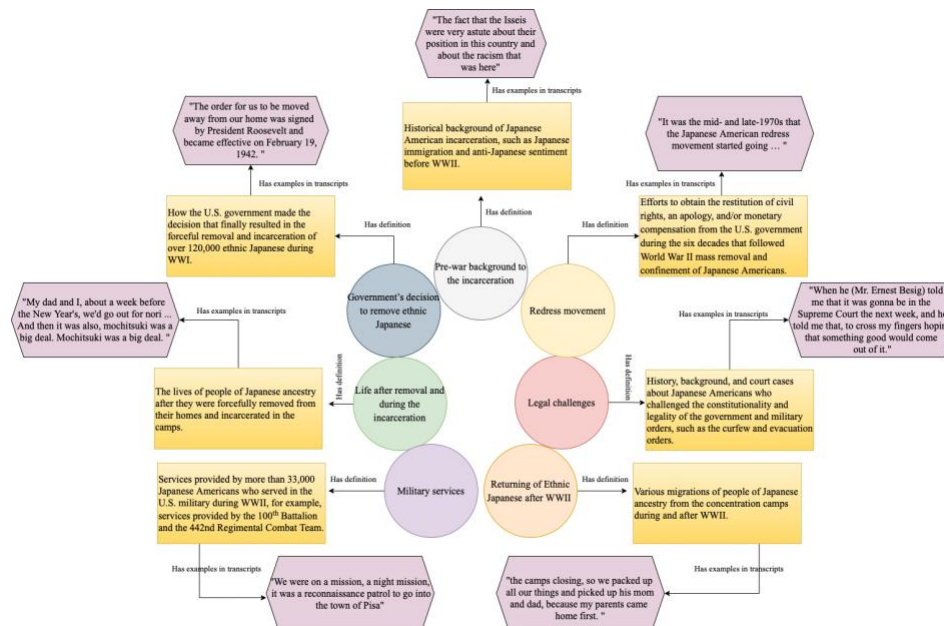


Figure 1: Annotation scheme for the history of Japanese American incarceration

CONCLUSION

This study analyzes an oral history collection of 904 narrators' interviews about Japanese American incarceration. Specifically, we analyze the demographic information of 904 narrators. Meanwhile, we also construct an annotation scheme based on which the incarceration historical texts can be manually or/and automatically processed. Future work will include annotating the interviews with the scheme and building machine learning models to automatically process the textual oral history.

REFERENCES

- Beckman, S & Froh, G. (2018). Densho Digital Repository: Preserving Community Memory. Retrieved from (http://downloads.alcts.ala.org/ce/20180426_Densho_Digital_Repository_Slides.pdf)
- Chen, H. H. (2022). *Data quality evaluation and improvement for machine learning*. [Doctoral dissertation, North Texas University].
- Densho. (n.d. a). About Densho. Retrieved from (<https://densho.org/about-densho/>).
- Densho. (n.d. b). Narrators. Retrieved from (<https://ddr.densho.org/narrators/>).
- Hovy, E., & Lavid, J. (2010). Towards a 'science' of corpus annotation: a new methodological challenge for corpus linguistics. *International journal of translation*, 22(1), 13-36.

Measuring Author Similarity by Topic Networks

Hsiao, Tsung-Ming
Chen, Kuang-hua

Dept. of Library and Information Science, National Taiwan University | f03126010@ntu.edu.tw
Dept. of Library and Information Science, National Taiwan University | khchen@ntu.edu.tw

ABSTRACT

Researchers have applied multiple methods, like co-authorship and author co-citation, to measure the relationships between authors and author similarity. These methods play a critical role in various applications of informetrics. A more delicate approach to gauging author similarity improves the related applications. By applying the method developed by the multiplex network, this study investigated how to identify author similarity by the mesoscale structures of author topic networks, which represent an author's research topic by organizing the interactions of the topic terms of their publications. Compared with the co-authorship relationship and topic-term similarity results, identifying author similarity by the mesoscale structure of author topic networks provides a better result reflecting how close authors' research topics are.

KEYWORDS

Informetrics; Multiplex Network; Author Similarity; Mesoscale Structure

INTRODUCTION

During the development of informetrics, multiple methods have been proposed to measure the relationships between authors, e.g., co-author, author co-citation, and author bibliographic coupling. Recently, the advance in multiplex networks provides alternative approach to gauging the relationships between authors based on their research topics.

Iacovacci et al. (2015) measured the similarity between research domains based on the mesoscale structure of the collaboration networks of these domains. The proposed method analyzes a network where "the nodes are connected by different types of interaction" (Iacovacci et al., 2015, p.12). We applied their method and used the topic terms of an author's publications to form a network, coined as author topic network (ATN), to represent their research topics, gauged the authors' similarity by analyzing the mesoscale structure of their ATNs, and compared the results with two other methods, co-author relationships and topic-terms similarity.

RELATED WORKS

Multiplex networks describe the interactions of numerous complex systems whose nature is different (Iacovacci et al., 2015). The complex systems are everywhere in the daily life. Currently, multiplex networks have been applied to numerous domains, including neuroscience, molecular biology, ecology, economy, transportation networks, infrastructures, and climate (Bianconi, 2018). The multiplex network is used to characterize the interactions of different elements. Some studies integrated the information from different networks into a new network and investigated the correlation between networks (Bianconi, 2018; Baccini et al., 2022). Scholars also applied these methods to simulate the diffusion of information in different contexts and explore what kinds of nodes affect the process of diffusion most (Li et al., 2019). Additionally, researchers have measured the relationships between nodes by the similarity of the networks characterizing the nodes (Iacovacci et al., 2015).

Various methods of informetrics aims at measuring the relationships between authors. These methods rely on cooperation relationships or co-occurrence frequency. Multiplex network shows alternative approach based on authors' research topics. Authors' research topics can be represented as a network composed of the topic terms of their works. This network, ATN, records topic terms related to an author's works and how the topic terms interact. The similarity based on the mesoscale structure of ATN will consider both the common terms and how topic terms relate to each other. Hence, it may measure the similarity between authors' research interesting and identify their relationship better. This study investigated the potential of this approach by comparing the results of the similarity based on ATN and the other two methods, co-author frequency and topic-term similarity.

RESEARCH DESIGN

In this study, we chose the authors of informetrics as the research target because our expertise could help us evaluate the results. Specifically, the authors who published at least 20 articles in three journals from 2010 to 2019 were considered in this study. The three journals were Journal of the Association for Information Science and Technology, Scientometrics, and Journal of Informetrics. The bibliographic data, whose data type was limited to "Article", were downloaded from Web of Science (WoS). Accordingly, 5,312 records were downloaded. We extracted the authors' full names from the "AF" field in these records, uppercased each author's name, and calculated number of publications for each author. Finally, 36 authors and 1,036 articles in total are considered in this study.

Then, we counted the co-author frequency between the 36 authors and measured their co-author relationships. We also measured the relationships between authors by checking their topic terms. The topic-term collection of an

author was built by aggregating their articles' KeywordPlus provided by WoS. For each pair of authors, their topic-term similarity was equal to the number of the common terms divided by the number of all terms in two collections. Both the results of co-author relationships and topic-term similarity were compared with that of ATN similarity.

For each author, the ATN was built with the KeywordPlus of their publications. ATN_{α} represent the ATN of author α . ATN_{α} can be represent as (V, E_{α}) , where V is the set of nodes $\{n_1, n_2, \dots, n_k\}$, k is the number of all distinct terms in the including articles, and E_{α} is a set of edges that shows connection between nodes. If an article contains three topic terms n_x, n_y , and n_z , it creates three edges, namely (n_x, n_y) , (n_y, n_z) , and (n_x, n_z) . After completing all authors' ATNs, the method of Iacovacci et al. (2015) was used to measure the ATN similarity between them.

RESULTS

The procedures mentioned above created three author similarity matrices. The agglomerative hierarchical clustering provided by SciPy was applied to categorizing them, and the distance between the authors and clusters was decided by Ward variance minimization algorithm. The subgraphs in Figure 1 show the three clustering results.

First of all, the authors who collaborated with each other frequently were put together in all three methods, e.g., Ludo Waltman and Nees Jan van Eck. However, the processes of forming the large clusters were different. For example, a noticeable difference is the process of categorizing the eight authors: Waltman, van Eck, Maisano, Franceschini, Fdez-Valdivia, Garcia, Rodriguez-Sánchez, and Prathap. They were divided into four sub-clusters, including (1) Waltman and van Eck, (2) Maisano and Franceschini, (3) Prathap, and (4) Gracia, Fdez-Valdivia, and Rodriguez-Sánchez. In Figure 1(A), these sub-clusters merged into one cluster without including the other authors. In Figure 1(B) and 1(C), however, these sub-clusters were merged in different order. A large cluster included Prathap and the other authors first. Then, this cluster included (1) and (2) sequentially. The authors of sub-cluster (4) were not be merged until all other authors were put into one cluster.

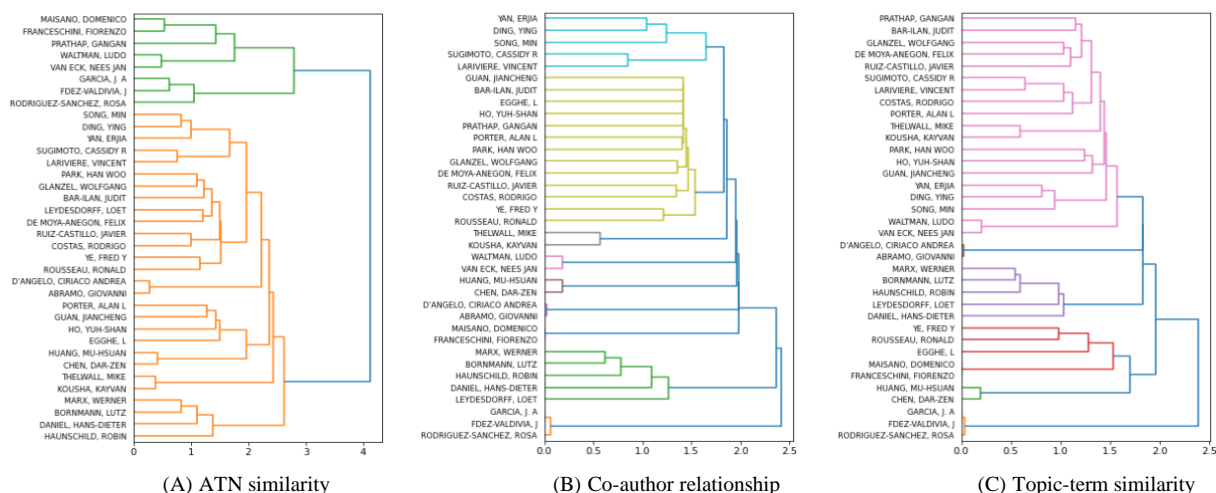


Figure 1. The results of agglomerative hierarchical clustering

We further examined the authors of yellow-green group in Figure 1(B) and found that these authors usually did not collaborate with the other included authors. ATN similarity and topic-term similarity further differentiated them.

The examination showed that ATN similarity may better reflect the similarity based on authors' research topics. For example, the included publications of de Moya-Anegón are regarding scientific development in Cuba and Latin American, scientific collaboration, and academic ranking. The including publications of Glänzel are in terms of subject classification and bibliometric-indicator issues, e.g., h-like index, distributions, and altmetrics. Based on our examination of the including publications, the topics of Glänzel are closer to those of Park and Bar-Ilan than de Moya-Anegón. Similarly, the topics of Bar-Ilan are more similar to Park than those of Prathap.

CONCLUSION AND FUTURE RESEARCH

The preliminary results of this study support that measuring authors' relationships based on their ATN may reflect the similarity of their research topics better. In the future study, we will enlarge the research scale to examine the current results and explore whether this method can be used to measure the relationships of different entities, like works and subjects.

ACKNOWLEDGMENTS

This work was partially supported by Ministry of Science and Technology, Taiwan under Grant No. MOST110-2410-H-002 -232 -MY2 and by the Universities and Colleges Humanities and Social Sciences Benchmarking Project under Grant No. 111L9A002.

REFERENCES

- Baccini, F., Barabesi, L., Baccini, A., Khelifaoui, M., & Gingras, Y. (2022). Similarity network fusion for scholarly journals. *Journal of Informetrics*, *16*(1), 101226
- Bianconi, G. (2018). *Multilayer networks: Structure and function*. Oxford.
- Iacovacci, J., Wu, A., & Ginestra, B. (2015). Mesoscopic structures reveal the network between the layers of multiplex data sets. *Physical Review E*, *92*(4), 042806.
- Li, W., Tian, L., Gao, X., & Pan, B. (2019). Impacts of information diffusion on green behavior spreading in multiplex networks. *Journal of Cleaner Production*, *222*, 488-498.

An Exploratory Study on Chinese Parents' Concerns about Preschoolers' Use of Digital Devices and Expectations for Digital Literacy Education

Huang, Ruhua School of Information Management, Wuhan University, China | rhhuang@whu.edu.cn
Shi, Leyi School of Information Management, Wuhan University, China | 1445296030@qq.com

ABSTRACT

It is necessary to carry out digital literacy education (DLE) for preschoolers to ensure they can participate in digital life in a safe and healthy way. This poster reports results from a qualitative study of Chinese parents' expectations of DLE for preschoolers. Through semi-structured interviews with seven parents whose children are aged from 3 to 6, it is found that all parents thought it necessary to carry out DLE for preschoolers. Chinese parents have expectations for the educational practices in kindergartens, but the digital literacy teaching activities in kindergartens are rare at present. In terms of teaching content, Chinese parents pay more attention to how their children could identify bad information and protect personal privacy.

KEYWORDS

Parents; Preschoolers; Digital literacy education; China

INTRODUCTION

In China, a preschooler refers to someone who is aged between 3 to 6 and is in the transition education period from kindergarten education to elementary education (The Ministry of Education of the People's Republic of China, 2016). Nowadays, an increasing proportion of preschoolers are accessing digital devices and Internet (CNNIC, 2022). However, Internet literacy of minors in China is quite limited, and some serious social problems such as cyber bullying and bad information are hard to ignore (Ji et al., 2021). Therefore, digital literacy education (DLE) is essential. It is noted that DLE is an essential approach to help young children to avoid network infringements (UNESCO, 2020). Moreover, researchers believe that young children who enter formal schooling with stronger digital skill may be better equipped to learn (Hurwitz & Schmitt, 2020).

Parental mediation has proven to be one of the most important ways to improve children's online behavior in China (CNNIC, 2021). However, parental supervision is not a perfect remedy, because minors could access the Internet not only at home, but also somewhere else without parents' supervision (Wang, 2018). There have been many related studies on parents' perspectives about preschoolers' digital life (Radesky et al., 2016; Brito et al., 2017; Sergi et al., 2017) and influence on children's development of digital skills (Livingstone et al., 2015; Livingstone et al., 2017; Scott, 2022), which further confirms that parents are one of the most critical "stakeholders" in children's use of digital devices and DLE. However, there is a lack of relevant research on parents' expectations for DLE for preschoolers, especially those focused on Chinese context. Therefore, we present an exploratory interview study to understand Chinese parents' viewpoints about preschoolers' use of digital devices and expectations of DLE for preschoolers, aiming to answer the following research questions:

RQ1: What viewpoints did parents have about preschoolers' use of digital devices?

RQ2: What expectations did parents have about DLE for preschoolers?

METHODS

We recruited 7 participants from relatives, friends and colleagues of the researchers. The sample requirements were limited as: (1) raise at least one child aged 3-6 at present; (2) whose children have used digital devices, such as iPad, smart phone and computer; (3) have some basic understanding on DLE or other relevant literacies education. Four mothers and three fathers were selected, aged from 28 to 37, who all have college degrees or above. Because of the COVID-19 pandemic, all interviews were conducted through telephone calls rather than face-to-face interviews.

We used semi-structured interviews to collect data. The interviews lasted from about 10 minutes to nearly 30 minutes. Interview questions are shown in Table 1. We recorded and transcribed all the interviews. Due to the small amount of data and commonality of interviewees' responses after preliminary analysis of transcribed texts, a coding scheme was developed based on the collected data and qualitative coding was carried out manually, looking for patterns and themes related to our research questions.

Parents' viewpoints towards preschoolers' use of digital devices	Parents expectations of DLE for preschoolers
What concerns do you have about your child's use of digital devices?	Besides yourselves, what kind of participant do you expect that could provide relevant teaching activities for your child?
Do you think it necessary to educate your child to participate in digital life safely and healthily?	What kind of teaching content do you prefer to provide for your child?

Table 1. Interview Questions

FINDINGS

The main concerns of parents regarding preschoolers' use of digital devices include bad information content, privacy disclosure and internet addiction. All participants mentioned that they worried about side effect of bad information content. In our research, bad information not only contains information that violates Chinese laws (The State Council of the People's Republic of China, 2011), such as pornographic and violent information, but also includes content that does harm to moral standards and correct values, such as soft porn, celebrity culture, rude words and superstition culture. M1, M2 and F6 pointed out that they had encountered children imitating some rude words after browsing social media APPs, and F7 pointed out that he had heard his child sharing misinformation. M4 said: *bad information is pervasive. Some uploaders are so good at disguising the bad content in order to escape from supervision. Even if I open the "Teen Mode" when my kids using social media platforms, there is still some soft porn information.* More than half of the participants mentioned privacy disclosure (M2, M4, F5, F7). M2 said: *my daughter has already had her own Wechat and Tencent QQ accounts. I am worried whether she might have posted anything personal on these platforms.* Some participants also raised concerns about internet addiction (M1, M3, F5, F7), and thought using digital devices for a long time may cause diminution of vision of preschoolers (M1, M2, M3, M4, F7).

With these concerns, all participants thought it necessary to educate preschoolers to use digital devices safely and healthily. Some participants' answers implicated that relevant DLE would develop preschoolers' self-discipline and self-management ability when using digital devices, which might be helpful to save parents time and energy which would be otherwise spent on monitoring their children's digital activities (M1, M2, M3, M4, F6).

Parents thought that informal education settings may be more anticipated. All participants pointed out that as a formal education setting, kindergarten had few practices in digital literacy education. Some participants also pointed out that teaching methods were too simple in kindergartens, while some informal education settings, like public libraries and Children's Palaces, might carry out better education activities (M2, M3, F5, F6). However, more than half of the participants thought that they lack channels of understanding and access of related teaching activities of informal education organizations. M1 said: *(I hope that) kindergarten teachers and community workers can provide parents with more information about related activities carried out by different institutions.* M2 said: *(I hope that) all types of institutions can come into our community and hold relevant activities for young children initiatively.*

As for teaching content, what parents care about most is to teach preschoolers how to screen out bad information and protect personal privacy. All participants mentioned that it was necessary to teach preschoolers to identify bad information, in order to help them distinguish online information and online behaviors between good and bad clearly. More than half of the participants thought it important to teach preschoolers protect online privacy (M2, M4, F5, F7). F5 said: *I really hope that my daughter could learn how to protect her online security...I think it is more important for girls to be educated about protecting online privacy and safety.* Actually, related research has pointed out that a significant number of girls were engaging in risky activities including disclosing personal information and sending personal photos to online acquaintances (Berson & Berson, 2005).

CONCLUSION AND DISCUSSION

Through interviews with seven parents, it is found that Chinese parents have concerns on preschoolers' use of digital devices, especially on harmful information and privacy disclosure. They believe that DLE for preschoolers is urgent. These concerns are closely related to parents' expectations of DLE for preschoolers. Firstly, Chinese parents expect to access more DLE opportunities from informal educational settings. Secondly, Chinese parents expect to prioritize improving preschoolers' skills about identifying bad information and protecting privacy.

However, the sample was relatively small and selective, and the interview questions was relatively simple. Therefore, future studies should consider expanding sample size and designing more comprehensive questionnaires and interview questions to conduct more in-depth analysis. We hope that by promoting future studies, we can provide a more feasible reference for development of DLE for preschoolers in China and even the world.

REFERENCES

- Berson, I. R., & Berson, M. J. (2005). Challenging Online Behaviors of Youth: Findings From a Comparative Analysis of Young People in the United States and New Zealand. *Social Science Computer Review*, 23(1), 29–38. <https://doi.org/10.1177/0894439304271532>
- Brito, R., Francisco, R., Dias, P., & Chaudron, S. (2017). Family dynamics in digital homes: The role played by parental mediation in young children's digital practices around 14 European countries. *Contemporary Family Therapy*, 39(4), 271–280. <https://doi.org/10.1007/s10591-017-9431-0>
- China Internet Network Information Center (CNNIC). (2022). *The 49th Statistical Report on Internet Development in China*. Retrieved from (<http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/>).
- China Internet Network Information Center (CNNIC). (2021). *Research Report on Internet use of Minors in China (2020)*. Retrieved from (http://www.cnnic.cn/hlwfzyj/hlwxzbg/qsnbg/202107/t20210720_71505.htm).
- Hurwitz, L. B., & Schmitt, K. L. (2020). Can children benefit from early internet exposure? Short- and long-term links between internet use, digital skill, and academic performance. *Computers & Education*, 146. <https://doi.org/10.1016/j.compedu.2019.103750>
- Ji, W., Shen, J., Yang, Y., & Ji, L. (2021). *Blue Book of Teenagers: Annual Report on the Internet Use of Chinese Minors (2021)*. Beijing, China: Social Sciences Academic Press (CHINA).
- Livingstone, S., Mascheroni, G., & Staksrud, E. (2015). Developing a framework for researching children's online risks and opportunities in Europe. London School of Economics, London: EU Kids Online.
- Livingstone, S., Ólafsson, K., Helsper, E. J., Lupiáñez-Villanueva, F., Veltri, G. A., & Folkvord, F. (2017). Maximizing opportunities and minimizing risks for children online: The role of digital skills in emerging strategies of parental mediation. *Journal of Communication*, 67, 82-105. <https://doi.org/10.1111/jcom.12277>
- Radesky, J. S., Kistin, C., Eisenberg, S., Gross, J., Block, G., Zuckerman, B., & Silverstein, M. (2016). Parent Perspectives on their mobile technology use: The excitement and exhaustion of parenting while connected. *Journal of Developmental and Behavioral Pediatrics*, 37(9), 694–701. <https://doi.org/10.1097/DBP.0000000000000357>
- Scott, F. L. (2022) . Family mediation of preschool children's digital media practices at home. *Learning Media and Technology*, 47(2), 235-250. <https://doi.org/10.1080/17439884.2021.1960859>
- Sergi, K., Gatewood, R., Elder, A., & Xu, J. (2017). Parental perspectives on children's use of portable digital devices. *Behaviour & Information Technology*, 36(11), 1148-1161. <https://doi.org/10.1080/0144929X.2017.1360941>
- The Ministry of Education of the People's Republic of China. (2016). *Working Regulations of Kindergarten*. Retrieved from (http://www.moe.gov.cn/srcsite/A02/s5911/moe_621/201602/t20160229_231184.html).
- The State Council of the People's Republic of China. (2011). *Measures for the Administration of Internet Information Services*. Retrieved from (http://www.gov.cn/zhengce/2020-12/26/content_5574367.htm).
- United Nations Educational, Scientific, and Cultural Organization. (2020). *Global Education Monitoring Report, 2020: Inclusion and Education: All Means All*. Retrieved from (<https://unesdoc.unesco.org/ark:/48223/pf0000373718>).
- Wang, P. (2018). Research on Internet Security Utilization of Minors in Foreign Countries: Progress and Enlightenment. *Information and Documentation Services*, (1): 89-96.

Analysis Of Acknowledgments of Libraries in the Journal Literature Using Machine Learning

Hubbard, David Texas A&M University, United States | hubbardd@library.tamu.edu
Laddusaw, Sierra University of Arkansas - Fort Smith, United States | sierra.laddusaw@uafs.edu
Tan, Qiaoyu Texas A&M University, United States | qytan@tamu.edu
Hu, Xia Rice University, United States | xia.hu@rice.edu

ABSTRACT

Increasing emphasis is being placed on research impact and it has prompted scholars to explore contributions beyond traditional research impact metrics. Acknowledgments, which are formal statements of indebtedness and contribution, within the journal literature provide an additional means to assess impact. This study examines contributions of libraries to the scholarly literature within acknowledgments using a combination of machine learning and manual methods to quantify and characterize acknowledgments.

KEYWORDS

Acknowledgment, Bibliometric analysis, Libraries, Machine Learning, Web of Science

INTRODUCTION

Acknowledgment studies have been used to identify contributions to publications with a focus on funding, credit, and genre (Desrochers et al., 2017). Recently, acknowledgment of the contribution of libraries and librarians in the publication process has been explored (Finnell, 2014; Hubbard & Laddusaw, 2020; Hubbard et al., 2018; Scrivener, 2009; Stigberg et al., 2015). These studies were conducted through manual extraction and classification of acknowledgments in limited sample sizes. Web of Science's Funding Text Field is increasingly being used to identify acknowledgments beyond just funding (Costas & van Leeuwen, 2012; Paul-Hus & Desrochers, 2019; Paul-Hus et al., 2017). However, there is a challenge when scaling up the sample size due to the sheer number of publications in bibliographic databases and ambiguities of language. This study looked to overcome those challenges using machine learning to identify relevant library acknowledgments.

METHODS

The acknowledgments to libraries and librarians were obtained from Web of Science Core Collection (WOS). This was accomplished by searching the term *librar** in the Funding Text (FT) to locate mentions of librarians, libraries, or library (i.e., library acknowledgment). Results were refined by year (2008-2020) and Document Types (Articles and Review Articles).

A model architecture was developed to determine whether the acknowledgment found was truly an acknowledgment about libraries and librarians versus other types of libraries (e.g., DNA library). The overall architecture of the model is shown in Figure 1. The model contains four components: input data format, data processing, representation learning, and classification model. The input used was the entire acknowledgment text from WOS, which was then processed in two steps to remove stop words using the NLTK tool and extract a fixed-length sub-sentence composed of a set number of words to the left or right of the keyword (i.e., *librar**). A number of fixed-length sub-sentences were tried from 1 to 10 words on either side of the keyword [i.e., window size (*w*)]. The extracted sub-sentence was classified using machine learning models that first represented the input data into embedded vectors using word2vec from Google. After the embedding layer, each acknowledgment text is represented by an embedding vector. In the final classification step, the Support Vector Machine (SVM) was used as the base classifier (Hearst et al., 1998); we exploit the SVM API provided by scikit-learn package in this project. The extraction and classification were developed in Python and is available as an open source project on GitHub (https://github.com/qiaoyu-tan/Acknowledgement_analysis). The model was first trained on a training set of acknowledgments (a subset) that were manually classified as being "true" acknowledgments to a library, libraries, and librarians. After the SVM was trained, we employed it to classify the entire set of acknowledgments downloaded from WOS. To accelerate the evaluation process, we randomly sample 380 acknowledgments from that set and manually assess the accuracy of the model. Of those that were true library acknowledgments within the random sample of 380, further *manual* classification was performed by acknowledgment category (i.e., facilities, people, resources, services, and general) as used in previous studies (Hubbard & Laddusaw, 2020; Hubbard et al., 2018), WOS Categories, first author affiliation by country, funding, and whether the acknowledgment was to a librarian(s) affiliated with an author's institution (local) or not (other).

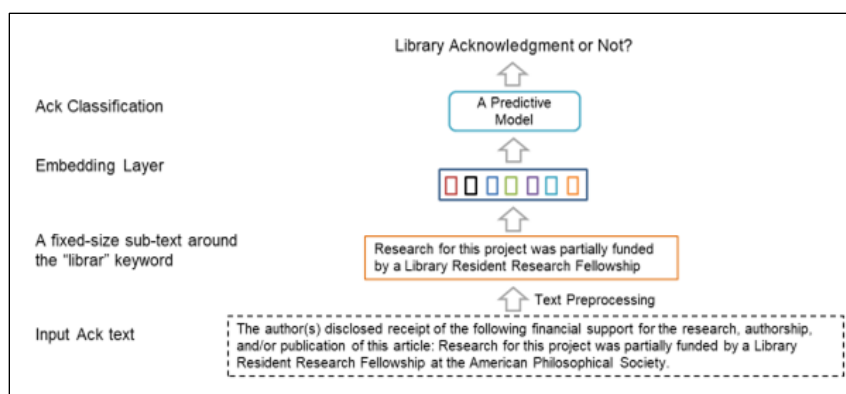


Figure 1. Model Architecture for Acknowledgment Processing

RESULTS AND DISCUSSION

There were 31,329 articles (Article or Review) found in WOS with the term *librar** for the years 2008-2020, which is 0.25% of the 12,679,964 acknowledgments indexed in WOS. There was a small number of acknowledgments (41 or 0.13%) that could not be predicted by the model since they do not contain valid keywords. Of the 380 test acknowledgments randomly sampled, 150 (39%) were determined to be actual acknowledgments to libraries or librarians. This was determined by manual inspection by two co-authors. The other 230 (61%) were determined to be false hits (e.g., a DNA library). Table 1 shows the accuracy of the machine learning prediction of the 380 test acknowledgments for window sizes of 1, 5, and 10. The accuracy was calculated using a confusion matrix (Ting, 2011). The smaller window size yielded a higher accuracy.

Window Size (w)	Precision (%)	Recall (%)	Accuracy (%)	F1 Score
1	88	62	82	0.73
5	97	49	79	0.65
10	95	35	74	0.51

Table 1. Precision, Recall, Accuracy, and F1 Score for Machine Learning Prediction Model by Window Size

Among the 150 library acknowledgments, authors acknowledged the libraries associated with their institutions slightly less (46%) than libraries not associated with their institution (51%), with a small percentage undeterminable (3%). Only 42 (or 28%) of the library acknowledgments were associated with funding from libraries.

The 150 acknowledgments that mentioned libraries or librarians were manually classified into the following categories: facilities (11 or 7%), people (73 or 49%), resources (80 or 53%), services (51 or 34%), and general (33 or 22%). In some cases, more than one category was assigned. For example, “I want to thank the librarians who have given me access to their manuscript and book” was assigned to the categories of people, resources, and services. Most of the acknowledgments were for use or access to a particular resource (e.g., archives, books, databases, etc.), though that category was closely followed by people that included both named individuals or generally (e.g., “librarians” or “staff”).

The 150 acknowledgments were found in journal articles associated with 114 of the 252 WOS Categories. The WOS categories associated with 5 or more journal articles included: History & Philosophy of Science (11), Multidisciplinary Sciences (7), Pharmacology & Pharmacy (6), Environmental Sciences (6), History (6), Humanities, Multidisciplinary (5), Medicine, General & Internal (5), and Public, Environmental & Occupational Health (5). The authorship of the 150 acknowledgments were affiliated with 36 different countries, with the United States accounting for over one third of the authorship.

CONCLUSION

Acknowledgments to libraries and librarians within the WOS funding acknowledgments was found to be small (< 1%); though these are only those identified as WOS funding acknowledgments. Further research is needed to understand why it is not higher, though it may be that such resources and services are taken for granted. The machine learning was effective at identifying over 80% of the library acknowledgments when compared to the manually classified random sample. Contrary to previous work that found more library acknowledgments in the biological and environmental sciences (Hubbard & Laddusaw, 2020), this larger study found more frequent library acknowledgments associated with the humanities and medical sciences based on WOS Categories.

ACKNOWLEDGMENTS

This work was supported by the T3: Texas A&M Triads for Transformation seed-grant program.

REFERENCES

- Costas, R., & van Leeuwen, T. N. (2012). Approaching the “reward triangle”: General analysis of the presence of funding acknowledgments and “peer interactive communication” in scientific publications. *Journal of the American Society for Information Science and Technology*, 63(8), 1647–1661.
- Desrochers, N., Paul-Hus, A., & Pecoskie, J. (2017). Five decades of gratitude: A meta-synthesis of acknowledgments research. *Journal of the Association for Information Science and Technology*, 68(12), 2821–2833.
- Finnell, J. (2014). Much obliged: Analyzing the importance and impact of acknowledgements in scholarly communication. *Library Philosophy and Practice*, 1229. <https://digitalcommons.unl.edu/libphilprac/1229>.
- Hubbard, D. E., & Laddusaw, S. (2020). Acknowledgment of libraries in the journal literature: An exploratory study. *Journal of Data and Information Science*, 5(3), 178-186.
- Hubbard, D. E., Laddusaw, S., Kitchens, J., & Kimball, R. (2018). Demonstrating library impact through acknowledgment: An examination of acknowledgments in theses and dissertations. *The Journal of Academic Librarianship*, 44(3), 404–411.
- Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and their Applications*, 13(4), 18-28
- Paul-Hus, A., & Desrochers, N. (2019). Acknowledgements are not just thank you notes: A qualitative analysis of acknowledgements content in scientific articles and reviews published in 2015. *PLOS ONE*, 14(12), e0226727.
- Paul-Hus, A., Díaz-Faes, A. A., Sainte-Marie, M., Desrochers, N., Costas, R., & Larivière, V. (2017). Beyond funding: Acknowledgement patterns in biomedical, natural and social sciences. *PLOS ONE*, 12(10), e0185578
- Scrivener, L. (2009). An exploratory analysis of history students’ dissertation acknowledgments. *The Journal of Academic Librarianship*, 35(3), 241–251.
- Stigberg, S., Guittar, M., & Morse, G. (2015). Capturing qualitative data: Northwestern university special libraries’ acknowledgments database. *portal: Libraries and the Academy*, 15(4), 571–585.
- Ting, K. M. (2011) Confusion matrix. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of machine learning* (p. 209). Springer.

Google Scholar Metrics: Assessing LIS Researchers' Scholarship in a Developing Country

**Islam, Md. Anwarul
Howlader, Ahmadul Islam
Roy, Prodip**

University of Dhaka, Bangladesh | anwar@du.ac.bd
Nuclear Power Plant Company Bangladesh Ltd, Bangladesh | ahmadul4206@gmail.com
RMIT University, Australia | prodip.roy@rmit.edu.au

ABSTRACT

The objective of this study is to examine and analyze the use of Google Scholar metrics by a cohort of Bangladeshi library and information science (LIS) researchers. This study is a follow-up to Islam and Roy (2021)'s study, which identified publications published by Bangladeshi LIS researchers in the Scopus and Web of Science (WoS) databases. We collected data purposively from the GS Profiles of LIS researchers working in different public and private universities. Only 32 LIS researchers in Bangladesh have a GS profile, with a total of 646 publications and 5752 citations listed in their GS profiles. All LIS researchers have an h-index, and our study identified strong correlations between the GS Citation metrics (publications, citations, h-index). However, the overall scholarship visibility is low. This is the first time an attempt has been made to examine publication metrics of GS in the context of Bangladeshi LIS researchers. These findings would help policymakers and researchers integrate GS Profile into the university website to display their scholarship activities.

KEYWORDS

Digital scholarship, Visibility, Google Scholar, Citation metrics, Bangladesh, Developing country

INTRODUCTION

For measuring the scholarly output of researchers in different disciplines, publication metrics have been used for many years in academia. Researchers are often evaluated by the number of publications and citations they produce, which are used to inform the hiring, tenure, and promotion committee to assess the faculty members' academic works (Yang and Meho, 2007). Many researchers and organizations have widely adopted the best-known citation databases Web of Science (WoS), Scopus, and Google Scholar (GS). GS feature such as a free, web-based database with facilities to search across many disciplines and explore related works, citations, authors, and publications help to evaluate and compare the research activities. For universities that cannot afford the subscription expenses of WoS or Scopus, GS can be an alternative tool (Jacso, 2012; Lund et al., 2021). Despite the limitations, e.g., false ranking, repetition, manipulation of citation, poor quality of sources, many studies found that GS Citations is one of the preferred sources used by the bibliometric community (Delgado Lopez-Cozaret et al., 2018; Martín-Martín et al., 2018). Many researchers use GS for evaluating and comparing scholarship activities since they do not have access to commercial databases. They assert that the GS profile can be used as one of the evaluation tools for researchers in developing countries who do not have access to commercial databases such as WOS and Scopus.

OBJECTIVES OF THE STUDY

The objective of this study is to examine and analyze the use of GS metrics by a cohort of Bangladeshi LIS researchers. To achieve this objective, we have come up with some RQs.

- How many Bangladeshi LIS researchers have GS profiles?
- How well are their positions in the context of Google Scholar metrics e.g. h-index?
- Is there any correlation among GS Citations metrics?
- How similar are citation counts in GS, Scopus and WoS?

METHODOLOGY

The target population of this study is LIS researchers, including faculty members, researchers and practitioners in Bangladesh serving in the universities and research organizations. For the present study, we set to analyze the GS profiles and examine the GS metrics of Bangladeshi LIS researchers. Firstly, this is a follow-up to the previous study by Islam and Roy (2021), which examined the extent to which Bangladesh-based LIS researchers have published in leading LIS journals indexed by WoS and Scopus. Secondly, for compiling the lists of LIS professionals whose profile is available in GS, the authors relied on the list of Islam and Roy (2021)'s study. Finally, the authors manually checked the GS profiles of LIS researchers in the respective universities and institutes. The respective library websites were searched based on the universities' names. From the library staff listing, the librarians' names were gathered and collated. Then, the author checked their names with their institutions in the GS. In the cleaning stage, the authors found many duplications of names like Islam, M. S or Islam, S. Hence, the authors avoided duplicating those names. Finally, the authors found 32 LIS researchers whose profiles are listed in the GS. Authors noticed that many LIS researchers have published in the LIS journals but did not have their GS profiles.

Hence, the name of those LIS researchers and their profiles could not be discovered. Data were collected from December 1 to December 30, 2021.

FINDINGS AND DISCUSSION

The thirty-two researchers had 646 publications, with a range between 1 and 60 publications in a single profile. Most LIS researchers had publications ranging from 1 to 10 in the GS profiles. The LIS researchers of Bangladesh have an average of 20.19 publications in the GS profiles. The 646 publications received a total of 5752 citations; the average citation per publication is 8.90. Of the thirty-two LIS researchers, all of them have an h-index. For those with an h-index score, the values ranged from 1 to 19, and the average h-index was 3.25. We found that most researchers (9) have an h-index of 3 (Figure 1). A Pearson correlation with a value of $r=0.813$ indicates a

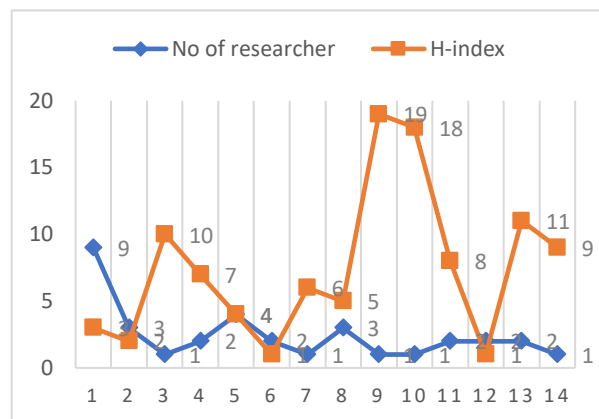


Figure 1. h-index (n=32)

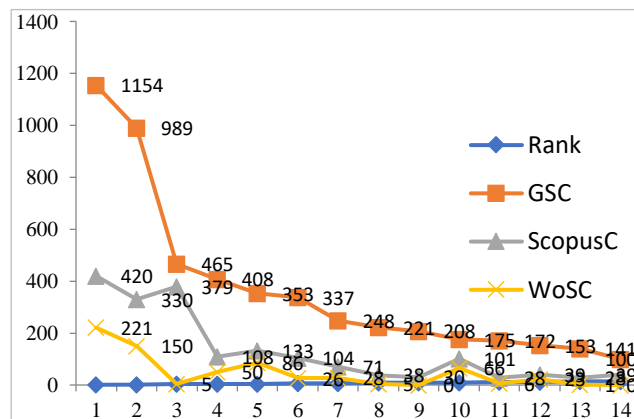


Figure 2. Comparison of citation analysis (n=32)

strong association between the number of publications and citations. This also presents the association between citations and the h-index. A Pearson correlation with an $r=0.947$ indicates a strong association between the number of citations and the h-index. We compared the top ten LIS researchers' citations found in Google Scholar and Scopus with those found in WoS and determined their differences in citation counts. Most LIS researchers have more citations in GS than in other citation databases (Figure 2).

The RQs have been analyzed and discussed systematically. We found that only a few LIS researchers have a GS profile, and many LIS researchers are unaware of this. For example, in Islam and Roy (2021)'s study, authors found that Islam, M. S., who has the third highest number of LIS publications in the Scopus and WoS, does not have a GS profile. The top five LIS researchers accounted for 224 (34.67%) of the total publications (646) in the Google Scholar profile. Of the top five LIS researchers, four faculty members are from two public universities and one from the private university library. We found that 646 publications received 5752 citations, and the top five researchers counted 336 (58.57%) of all citations. This finding supported Islam and Roy (2021)'s study where they found Ahmed, S.M.Z, Islam, M.A and Islam, M.S had the highest number of publications and citations in the Scopus and WoS databases. Compared to Scopus and WoS, the top ten LIS researchers have more citations in GS. For example, Ahmed has 974 citations in GS, 327 in Scopus and 230 in WoS. The results suggest that GS citation data is a superset of Scopus and WoS, with substantial extra coverage in all areas.

In general, citation metrics of GS should not be used alone for any assessment as there are other tools and processes to assess the scholarship activities of researchers. The other indicators like seminal work, research grants, labs being run, PhD supervision, etc. (which are not a part of citation analysis) help to measure the scholarship activities. However, in developing countries, while researchers do not have access to most of the subscribed databases and have less funding support for research, metrics analysis of GS can be a way to assess the scholarship activities. The other way GS profile helps to visualize the scientific output of countries. For example, AD Scientific Index (Alper-Doger Scientific Index) ranks individual scientists, universities and countries based on the productivity of scientific publications. Using the total and last five years' values of the i10 index, h-index, and citation scores in GS, this system ranks the researchers, institutions and countries. This system ranks institutions based on the scientific characteristics of affiliated scientists (AD Scientific Index, 2022). By integrating the researcher GS profile in the university/organization website, researchers and universities can leverage their improved visibility to find potential new audiences and collaborators (Ale Ebrahim et al., 2014). We suggest appropriate policies and guidelines set by authorities to create GS profiles in organizations that could help display their scholarship activities. For example, every university library and library professional in Bangladesh can help Bangladeshi researchers display their publications by advising on best practices in post-publication strategies, including creating a compelling academic profile.

REFERENCES

- AD Scientific Index (2022). World Scientist and University Rankings 2022, retrieved from <https://www.adscientificindex.com/>
- Ale Ebrahim, N., Salehi, H., Embi, M.A., Habibi, F., Gholizadeh, H. and Motahar, S.M. (2014). Visibility and citation impact, *International Education Studies*, 7(4), 120-125
- Delgado López-Cózar, E., Robinson-García, N., and Torres-Salinas, D. (2014). The Google scholar experiment: How to index false papers and manipulate bibliometric indicators, *Journal of the Association for Information Science and Technology*, 65(3), 446-454.
- Islam, M. A and Roy, P.K (2021). Scholarly Productivity of Library and Information Science Research in Bangladesh: A Bibliometric Study from 1971 – 2020, *DESIDOC Journal of Library and Information Technology*, 41(3), 213-225
- Jacso, P. (2012). Google Scholar metrics for publications: The software and content features of a new open-access bibliometric service, *Online Information Review*, 36(4), 604–619.
- Lund, B. D., Wang, T., Shamsi, A., Abdullahi, J., Awojobi, E. A., Borgohain, D. J., ... and Yusuf, A. O. (2021) Barriers to scholarly publishing among library and information science researchers: International perspectives, *Information Development*, doi: 10.1177/02666669211052522.
- Martín-Martín, A., Orduna-Malea, E., Thelwall, M., and López-Cózar, E. D. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories, *Journal of informetrics*, 12(4), 1160-1177.
- Yang, K., and Meho, L. I. (2007). Citation analysis: a comparison of Google Scholar, Scopus, and Web of Science. *Proceedings of the American Society for information science and technology*, 4(1), 1-15.

Sustainable Development Goals and Public Libraries: An Exploratory Study in a Developing Country

Islam, Md. Anwarul

University of Dhaka, Bangladesh | anwar@du.ac.bd

Sultana, Rafia

Department of Public Libraries, Bangladesh | rafiasultanasabiha@gmail.com

Widen, Gunilla

Åbo Akademi University, Finland | Gunilla.Widen@abo.fi

ABSTRACT

The objectives of this study is to investigate how public libraries in Bangladesh work to achieve some of the UN SDGs and the strategies they adopt to work in line with the mission of UNSDGs. Data were gathered from 59 public librarians working in public libraries of Bangladesh who provided qualitative answers to three open-ended questions in a survey questionnaire. SDG#1, 2, 3, 4 and 17 are the areas where public libraries are doing well and summarized their opinions how public libraries can do well further for other goals. Finally, authors put some recommendations to achieve other goals as well.

KEYWORDS

UNSDGs, Public library, Bangladesh, Access to information, Community engagement, Information services, Qualitative approach

INTRODUCTION

For building an inclusive society public libraries offer a local gateway of knowledge, support lifelong learning and formal education, helping access to information, strengthening early literacy, promote awareness of cultural heritage, fostering ICT use and support other social issues for the community people (IFLA/UNESCO, 1994). The manifesto of public libraries mostly falls with the essence of the United Nation Sustainable Development Goals (UNSDGs) which came into action in 2016. The seventeen goals and one hundred sixty nine targets work to ensure 'Leave no one behind' which is the essence of public libraries to build an inclusive society (UNDP, 2015; Scott, 2011). For taking a bottom up approach to development, public libraries can work as one of the key-partners and active contributors. For doing this, public libraries need to promote libraries' role as development agents, namely by gathering evidences and evaluating their contribution to SDGs implementation (Pinto & Ochôa, 2017). Considering this, many public libraries across the world are designing community services, creating need-based content, fostering inclusive environment and providing services for the marginalized and ethnic groups (Bradley, 2016; McCook & Phenix, 2007; Edwards & Edwards, 2010).

In 2018, International Federation of Library Associations and Institutions (IFLA) has created a map of libraries and provide performance metrics of all types of libraries in all regions of the world. IFLA demonstrated 34 countries with 56 stories and showed how libraries and access to information contribute to improved outcomes across all SDGs. We found that public access to information, training in new skills, and services that give people opportunities to improve their lives lead to support the SDG 1 # 'no poverty' in Kenya, South Africa, Chile and India (IFLA, 2018). Many international studies support that public libraries support to achieve SDGs by many ways. Some of these studies are; public libraries support SDG 1 "no poverty" and SDG 2 "zero hunger" by converting spaces as shelters, food distribution centers, etc. (Tyler, 2019); Public libraries in the USA are providing COVID-19 trusted sources (ALA, 2020); German public libraries help refugees by offering training, education, and health information (Wang & Lund, 2020); public libraries in Scotland support health and well-being (Tyler, 2019) and public libraries in Australia are helping social inclusion (Hider, *et.al.*, 2022).

Public libraries in Bangladesh run under the Directorate of public libraries (DPL). There are seventy one government public libraries run across the country and these libraries offer wide range of services. However, these services mostly fall under the reading room service, book lending, ICT service and some extension services ((Islam and Islam, 2010; Chowdhury, Islam and Islam, 2012). While there are some studies done in the context of public libraries in Bangladesh, very few studies focus on SDGs with public libraries. Alam (2020) discussed how public libraries in Bangladesh can help to achieve some SDGs and make people more media and information literate. For bringing out the evidences and taping the examples under the SDGs and public libraries in Bangladesh, the present study demonstrates how public libraries are working well toward the UNSDGs.

OBJECTIVES OF THE STUDY

The objectives of this study are to know the evidence gathering of Bangladeshi public libraries' contribution to the UN 2030 Agenda. To address this, we have come up with a number of research questions (RQs)

RQ1: What are public libraries in Bangladesh doing to meet the SDGs?

RQ2: Are there any SDGs projects running in the public libraries?

RQ3: How do public libraries can help to achieve some of the SDGs?

METHODOLOGY

The present study is a part of a larger quantitative survey of public librarians working in public libraries across the country. For the larger study, a web-based survey questionnaire with link was sent out to the seventy one public librarians and fifty-nine responded to the survey which was 83.09 percent of all population. The objective was to investigate the SDGs awareness, perceived understanding and identifies the challenges affect the public library adoption to SDGs in Bangladesh. Under the same questionnaire, there were three open ended questions and the qualitative analyses of the three open-ended responses are the research focus of this study. We have inserted the three open-ended responses to a separate worksheet, edited, coded and finalized it.

DATA ANALYSIS

After analyzing the open and axial coding of the responses, we came up with selective categories for each question to synthesize the findings (Corbin and Strauss, 1990). There were 38 responses stating what set their libraries to progress towards SDGs and we mapped their answers under the SDGs (RQ1). The numbers within brackets indicate the sum total for all responses in that category. Selective responses are summarized which falls under some of the goals. **SDG#4 Quality Education** - *'We are doing our best with quality education. Being a self-literary center for all, our public library provides access to all community people'*; *'We have extra focus for kids to expand creativity with toy bricks, coding with kano computers, storytelling sessions, poem recitation etc'*; *'We are working to provide quality education for all ages group. We are working on keeping teens aware of the climate goal and providing a range of study materials such as books on science and maps'* (16); **SDG#1 No poverty & SDG# 2 Zero Hunger**- *'Access to information, reduced inequality by providing free computer, internet and training facilities to empower'*; *'We offer career guideline session for the job seeker'* (12); **SDG#17 Partnerships to achieve the Goal** - *'We are collaborating with other national and international organizations to support the community'*; *'We are jointly conducting projects and training programs with the ICT ministry, Education ministry and respective authorities'* (10). Under the **RQ2**, there were 29 responses stated that they run some projects. Some of the selected leading projects run by the public libraries in Bangladesh are 'Libraries Unlimited' 'Bangobondhu o Muktijuddho Corner project' 'Establishment of Chittagong Muslim Institute Cultural Complex Project'. For example, 'Libraries Unlimited' a collaborative project run by British Council and Department of Public Libraries (DPL) seek to improve access to information and knowledge, develop partnership with government, train public library staff, developing a shared vision of public libraries to embed the country's commitment to the Sustainable Development Goals. In addition, this project runs coding workshop for the school going kids to learn programming language and 'do your: bit' 'Robot Olympiad' workshop for the kids. Some of these activities fall under the goals of **SDG #1, #2, #3'Good Health & Well-being'** and **#17** . Under the **RQ3**, there were 22 responses make some recommendations to achieve the SDGs. *'To attain more focus towards SDGs, awareness of librarians and policy makers is the primary requirement'*; *'Policy can be made with certain goals and specific tasks can be assigned to public libraries that are compatible with the goals'* (8) *'Funding is very important'*; *'There should be some extra budget along with the tasks towards SDG's'*(5)"; *'Public library needs a complete action Plan on UN SDG's'*; *'Proper guidelines should be made'*; *'Goal oriented task should be introduced'*; *'DPL should think out of box to be engaged with SDGs more'*(5) *'By providing the related information among the specific people, Public libraries can achieve some of the goals'*; *'PL can create awareness in the community with the 17 goals'*(4).

DISCUSSION

Even though the sample size was small, this exploratory study can draw some significant findings to discuss. While the public librarians are working well with the SDG#1, 2, 3, 4 and 17, the other goals are not identified in our analysis. It might be the reason that public librarians in Bangladesh are not well aware of the other SDGs. Under the lens, 'providing access to information' public libraries can support the other goals also. For example, public libraries can offer programmes for women and girls to access information about their rights and health SDG #5: Gender equality. Public libraries can organize and increase awareness programs among adolescent girls on menstrual hygiene where one of the third girls in Bangladesh had heard about menstruation before menarche. In a large scale, public libraries can support SDG#10 Reduce inequality by creating need-based contents and offering those contents countrywide through Internet. The evidences of public libraries support to SDGs should be collected and compiled to convince the government and policymakers. Public libraries do not have any silver bullet to cover all the SDGs equally. Considering the capacity and local context, public libraries in Bangladesh can design the community-based information services to make people aware about SDGs and can look into what other public libraries are doing well globally. Later, public libraries can start to work on certain goals which will systematically follow targets, indicators and methods to achieve the SDGs. Finally, based on the national priority, public libraries can create a SDG mapping tool and can align with the 'SDG Tracker'- a set of thirty nine indicators prepared by Bangladesh government to achieve UNSDGs.

REFERENCES

- Alam, A (2020) "Achieving Sustainable Development Goals (SDGs) in Bangladesh: The Role of Public libraries" Paper presented at the 1st Multidisciplinary International Conference on the Social and Life Sciences: Gender, Health, Information and the Environment (ICSL 2020), February 27-29, 2020 <http://iikm.ewubd.edu/icsl/> (accessed 10 May 2022).
- American Library Association (ALA) (2020) Public Libraries Respond to COVID-19: Survey of Response & Activities, Retrieved from (<http://www.ala.org/pla/issues/covid-19/march2020survey>)
- Bradley F (2016). A world with universal literacy: The role of libraries and access to information in the UN 2030 Agenda, *IFLA Journal*, 42(2), 118–125
- Chowdhury, S., Islam, S., & Islam, A. (2012) The information literacy education readiness of Central Public Library (CPL) in Dhaka of Bangladesh, *International Journal of Information Science and Management*, 9(2), 23-44.
- Corbin, J. and Strauss, A. (1990). Grounded theory research: procedures, canons, and evaluative criteria, *Qualitative Sociology*, 13(1), 3-2
- Edwards, J. B., & Edwards, S. P. (Eds.) (2010), 'Beyond Article 19: Libraries and social and cultural rights' Library Juice Press, LLC.
- Hider, P., Garner, J., Wakeling, S., and Jamali, H. R. (2022) Serving Their Communities: An Analysis of Australian Public Library Mission Statements, *Journal of Library Administration*, 62(2), 190-205
- IFLA (2018) Libraries and the sustainable development goals: a storytelling manual, Retrieved from (<https://www.ifla.org/wp-content/uploads/2019/05/assets/hq/topics/libraries-development/documents/sdg-storytelling-manual.pdf>)
- IFLA/UNESCO (1994) IFLA/UNESCO Public Library Manifesto 1994, Retrieved from (<https://www.ifla.org/wp-content/uploads/2019/05/assets/public-libraries/publications/PL-manifesto/pl-manifesto-en.pdf>)
- Islam, M. A & Islam, M.M (2010) Community Information Services Through Public Libraries in Bangladesh: Problems and Proposals', *Library philosophy and practice*, Retrieved from (<https://digitalcommons.unl.edu/libphilprac/320/> i)
- Pinto, L. G., & Ochôa, P. (2017). Public libraries' contribution to Sustainable Development Goals: gathering evidences and evaluating practices, Retrieved from (<http://library.ifla.org/id/eprint/1946/>)
- McCook, K. D. L. P., & Phenix, K. J. (2007) Public libraries and human rights, *Public Library Quarterly*, 25(½), 57-73
- Scott, R. (2011). The role of public libraries in community building, *Public Library Quarterly*, 30(3), 191-227.
- Tyler, A. (2019) Health on the shelf: health and Well-being in Public Libraries in Scotland, Retrieved from (<https://scottishlibraries.org/media/3008/health-on-the-shelf.pdf>)
- United Nations Development Programme (2015) World leaders adopt Sustainable Development Goals. Retrieved from (<http://www.undp.org/content/undp/en/home/presscenter/pressreleases/2015/09/24/undp-welcomes-adoption-of-sustainable-development-goals-by-world-leaders.html>)
- Wang, T. & Lund, B. (2020) Announcement information provided by United States' public libraries during the 2020 COVID-19 pandemic, *Public Library Quarterly*, 39(4), 283-294

Student Internships in the Time of Crisis: How IT Companies Cope with New Challenges and Opportunities

Jakopec, Tomislav

Faculty of Humanities and Social Sciences Osijek, Croatia | tjakopec@gmail.com

Aparac Jelušić, Tatjana

Faculty of Humanities and Social Sciences Osijek, Croatia | taparacjelusic@ffos.hr

Szombathelyi, Donata

Faculty of Humanities and Social Sciences Osijek, Croatia | dszombathelyi@ffos.hr

ABSTRACT

This poster presents findings from a small-scale study which investigated student internship (hybrid and virtual) during the COVID-19 lockdown. The study was conducted in 18 Croatian IT companies using semi-structured interviews. The study tried to answer the following research questions: RQ1: How did local IT companies organise student internships during COVID-19 lockdown; RQ2: What are advantages and disadvantages of online student internship. The study was conducted within the framework of Erasmus + project DECriS (Digital Education for Crisis Situations: Times When There is No Alternative, 2021–2023). Results of this study showed that the advantages of virtual internships are flexibility, saving time, the possibility for the interns to dedicate themselves to their private and academic obligations, comfort, easier work organization, the possibility of self-organization and reduced risk of COVID-19 infection. However, other respondents indicated the following negative sides of virtual internship: the lack of communication among students and mentors, students' lack of opportunity to learn about organizational, business and ultimately the human side of the company.

KEYWORDS

Cooperation with IT sector; Croatia, IT companies, (L)IS education, Student internships; Virtual internships

INTRODUCTION

Higher education institutions (HEIs) across the globe strive to educate young people not only for the current job market, but also to make them aware of various paradigms in the development of civilization and culture in order to better understand changes in today's digital age, and especially in crisis situations such as COVID-19.

This study was carried out as a part of one of the intellectual outcomes of the Erasmus + project DECriS (Digital Education for Crisis Situations: Times When There is No Alternative, 2021–2023) which is focused on the European context and has three main goals: to explain the terminological differences in describing the work of Higher Education (HE) students in public information institutions and private enterprises as part of the curriculum (obligatory or optional); to point out the results of some studies taken before and during COVID-19 pandemic in order to compare the approaches, models and outcomes; and to analyse the results of interviews with owners/directors/Human resource management employees of private computer companies. After this pilot study (which was conducted in Osijek, Croatia) similar studies will be conducted at other partner HEIs (Barcelona, Spain; Sofia, Bulgaria; Hildesheim, Germany and Zagreb, Croatia).

PREVIOUS RESEARCH

The benefits of library internships and field experiences have been well documented by international authors (Juznic and Pymm, 2011; Bird, Chu and Oguz, 2011; Martínez Arellano and Ortega, 2012). During the COVID-19 pandemic, some authors (Juarez and Blackwood, 2022; Bayir, 2021) investigated the approaches taken by LIS schools which aimed to minimise the obstacles in the organisation of virtual internships. From the point of view of the hosting institutions, the most important issue relates to the availability of adequate, reliable and flexible software and fast communication networks.

RESEARCH METHODOLOGY

Aim, preparation and sample

In Croatia, more specifically in the eastern part of the country where the city of Osijek is located, a number of successful IT companies started emerging in 2012 (Jakopec, 2020) and over the course of years they established a fruitful cooperation with the Department of Information Sciences at the University of Osijek. In this study authors wanted to learn how these local IT companies organised student internships during COVID-19 lockdown and what they detected as advantages and disadvantages of the virtual internship programs. It was decided that in such a competitive environment, as IT sector, online questionnaire would not yield relevant results. Therefore, a semi-structured interview was chosen as a research method. Company directors, branch managers or human resources managers in 20 IT companies were invited to participate in the study. A total of 18 companies responded to the

invitation. The interviews were conducted *in vivo* in March, 2022. The data were audio recorded, transcribed and analysed.

Data analysis

With two set research questions authors wanted to learn how the local IT companies organised internships during COVID-19 lockdown and what they detected as advantages and disadvantages of the online/remote internship. We analysed the following:

- a) companies profile including information about their experiences from before the COVID-19 pandemic (regarding remote working, student scholarships and internships) and preferred way of offering students' internship programs (on site, hybrid, remote),
- b) respondents' opinion about the students' internship in the context of the availability of resources (people and time) required to supervise practical training activities as part of internship programs, before the COVID-19 pandemic, during and after the lockdown,
- c) respondents' perceptions and experiences regarding technologies, tools and working processes during internships.

Results of this study showed that the advantages of virtual internships are flexibility, saving time, the possibility for the interns to dedicate themselves to their private and faculty obligations, comfort, easier organization, the possibility of self-organization and reduced risk of COVID-19 infection. However, some respondents see more negative sides of the implementation of virtual internship among these lack of communication, missing the opportunity to learn about organizational, business and ultimately the human side of the company.

DISCUSSION AND CONCLUSION

The pandemic emphasized our dependence on technology and thus enabled the business world in the IT sector to gain even stronger dominance. Although students and educators do not know will education look like after the pandemic, most agree that it will be different. Virtual internship programs, which were offered during the recent lockdown, proved to be a solid alternative to the traditional onsite internships. In the period when physical contact was limited, it enabled the successful implementation of educational programs which in most cases include obligatory student internships. Some respondents in this study indicated that their companies as hosting institutions might consider developing future internship programs based on the perceived advantages of virtual internships and that the perceived disadvantages can help them design an effective hybrid model. Before setting up the model which could suit HEI sector the best, it is necessary to conduct the same or similar surveys in other public and/or business environments in order to be able to take into account all relevant facts and set up a comprehensive model of organizing internships at times when there is no alternative.

REFERENCES

- Bayır, D. (2021). Impacts of the COVID-19 Pandemic on Experiential Learning: Development of a Community-Grounded Online Internship Program. *UTP University of Toronto Journals. Advance Access Article*. Published Online: December 21. <https://www.utpjournals.press/doi/abs/10.3138/jelis.2020-0108>
- Bird, N.J., Chu, C.M., and Oguz, F. (2011) Four "I"s of Internships for the New Information Society: Intentional, Interconnected, Interdisciplinary and International. 77th IFLA General Conference and Assembly, San Juan, Puerto Rico. Retrieved from: <https://core.ac.uk/download/pdf/149239145.pdf>
- Jakopec, T. (2020). Razvojni put diplomiranih studenata informacijskih znanosti Filozofskog fakulteta u Osijeku zaposlenih u IT tvrtkama. *Vjesnik bibliotekara Hrvatske* 63(1/2), 683–702. doi:10.30754/vbh.63.1-2.670.
- Juarez, D. and E. Blackwood (2022). Virtual Undergraduate Internships: One COVID-19 Side Effect That Academic Libraries Should Keep. *Portal: Libraries and the Academy*, 22(1), 81–91. 10.1353/pla.2022.0010.
- Juznic, P. and Pymm, B. (2011). Students on Placement: A Comparative Study. *New Library World*, 112 (5/6), 248–260.
- Martínez Arellano, F. F. and J. R. Ortega (2012). Linking Theory and Practice in LIS Education in Latin America and the Caribbean. IFLA Conference paper. <https://www.ifla.org/past-wlic/2012/213-arellano-en.pdf>

Children's Interest, Search, and Knowledge: A Pilot Analysis of a STEM Maker Workshop

Jung, Yong Ju The University of Oklahoma, USA | yongju@ou.edu

Liu, Jiqun The University of Oklahoma, USA | jiqunliu@ou.edu

ABSTRACT

When understanding information seeking, individuals' knowledge construction and interest development are critical to be considered because of their active interactions with each other. Especially for children, these three dimensions—interest, search, and knowledge—can constantly influence each other and support meaningful learning experiences. However, there is a lack of empirical investigation to understand this interplay between the dimensions, particularly in children's real-life environments. Building upon the previous literature on the conceptual demonstration of interest, search, and knowledge construction, we used a qualitative, video-based research approach to conduct a pilot analysis of a child's experiences from a maker workshop at a science museum. Our pilot analysis provides a basis for empirical investigation of children's authentic information seeking and learning.

KEYWORDS

Children's search; interest; knowledge construction; makerspace; STEM learning

INTRODUCTION AND THEORETICAL FRAMEWORK

Information seeking and use is an essential component of individuals' learning, especially related to expanding knowledge (Rieh et al., 2016). When examining information seeking and/or knowledge construction, it has been common to highlight the cognitive process involved. However, the importance of their affective aspect is getting more attention; interest development, which embraces both cognitive and emotional attributions, has tight relationships with learning and knowledge acquisition (Renninger & Hidi, 2016). Characterizing and predicting the interplay of interest development, knowledge construction, and information seeking are critical for understanding the role of information seeking in promoting knowledge construction and ultimately cultivating new interests and supporting learning experiences. Practically, it is also useful for designing interactive systems and educational programs to facilitate children's learning in real-life environments.

To represent the interplay of the three dimensions discussed above, in a previous study (Liu & Jung, 2021), we developed an interest-search-learning (ISL) model that characterized transitions in the states of interest, search, and learning and the temporal interactions among them. According to this model, meaningful learning experiences allow individuals to find new interests in certain areas or maintain existing interests, and their learning and interests also motivate them to actively engage in broader information-seeking episodes. Then, the new information collected and used in information-seeking tasks could further enhance learning and reinforce interests behind motivating the tasks.

However, we did not provide empirical evidence that could clarify the interactions and state transition patterns defined in the ISL model. Besides, how children seek and utilize new information in their learning and interest development, especially in real-life information institution settings (e.g., museums, libraries, makerspaces), remains a critical but open question. Inspired by the conceptual explorations under ISL, our study takes a step forward by empirically exploring the interactions among online information seeking, interest development, and knowledge construction in learning, in particular, of children's search behaviors and activities. Specifically, this late-breaking work aims to answer one **research question**: how do a child's interest development, information seeking, and learning activities interact with each other in a real-life, collective learning environment?

METHODS

Setting and Data Source

To understand the multidimensional transition of children's interest development, search, and knowledge construction in a real-life setting, we investigated a set of video data from a family STEM workshop at a science museum. This workshop included two parts of activities (first, exploring the science museum for 2-30 minutes, and second, making artifacts using circuits and doughs for 3-40 minutes). This pilot analysis only focused on the second activity (i.e., maker workshop) of one dyad of mother and child (female, 10 years old). For the activity, materials such as Play Doh, Squishy Circuits (circuits that could be connected to tangible doughs), and a tablet (i.e., iPad mini) were provided. The workshop was an authentic—rather than a controlled or experimental—educational environment as it was how a workshop for children's STEM learning would normally occur. A brief instruction about how to use Squishy Circuits was provided by a facilitator, but basically, the activity was driven by the family. This family was collaboratively engaged in the activity, leading to collective search and learning. This family created multiple artifacts such as a necklace and a car. Using the tablet was not mandatory, so their search behaviors with it were driven by their needs; this family used the tablet multiple times to search for information.

85th Annual Meeting of the Association for Information Science & Technology | Oct. 29 – Nov. 1, 2022 | Pittsburgh, PA. Author(s) retain copyright, but ASIS&T receives an exclusive publication license.

Pilot Analysis

We conducted a qualitative video-based interaction analysis (Jordan & Henderson, 1995; Nassauer & Legewie, 2021) to examine observable experiences and interactions of the family during the workshop. Our pilot analysis was performed using V-note, software for video analysis, through the following steps: First, we used a coding framework to identify observational evidence of each dimension—interest development, knowledge construction, and online information search. The coding framework was adapted and modified building upon previous literature (e.g., Hidi & Renninger, 2006; Jung et al., 2019; Liu & Jung, 2021; Scardamalia & Bereiter, 1991). It included observable evidence of newly-triggered short-term (situational) interest (e.g., emotional reactions such as exclamation and curiosity about a new subject) and maintained situational interest (e.g., repeated attention on the subject after situational interest was triggered); that of the states of search (e.g., exploration, exploitation, known-item, and synthesis/evaluation); and that of knowledge construction (e.g., knowledge acquisition, use, building). Due to the page limit, the full coding framework was not included in this paper. Second, to establish the trustworthiness of this coding, the first and second authors had a meeting, reviewed the transcript and video recording together, and made a full consensus about the coded episodes. Third, after identifying such evidence, we repeatedly revisited the coded episodes, especially when evidence of all three dimensions were identified at the same moment, to explore in what situations such interplay happened and map them into the series of episodes.

FINDINGS

In Episode 1 (Figure 1), the mother started using the tablet to search for examples of Squishy Circuits artifacts and found an example of a car. She showed it to the child, which was evidence of exploitation search as well as knowledge acquisition. This sparked the daughter's new interest in "building a car."

Part of the transcript from Episode 1:

Mom: (*Mom uses the iPad to search for topically relevant information {exploitation search} and shows the child the iPad screen*) They are showing you can build a car {new knowledge about the car artifact}.

Child: (*Reading the iPad together {new knowledge}*) Oh, that's really cool! {triggered situational interest} But I can't build, so. I mean I could try after I am done with this.

After their interest in car artifacts was sparked, they continued making. However, in Episode 2, they ended up experiencing a challenging situation where the LED light connected to their circuit was not working. Despite the challenge, the child maintained her situational interest as she kept trying, gathered new information, and paid attention to her artifact. The child used the tablet to check the example and further information to solve their specific problem and built knowledge with her mother by making constant discussions to try different methods.

Part of the transcript from Episode 2:

(*Making is continued {maintained situational interest}. Both the child and mother look at the information from the iPad to figure out why the circuit is not working {known-item search}.*)

Mom: The long end has to be on the red side [of the LED light legs].

Child: Yeah it...it was...do we, do we have to use this? {knowledge building}

Mom: Yeah. Yeah, it has to be somewhere.

Child: Oh, we have to make wheels, I'll make the wheels {information use}.

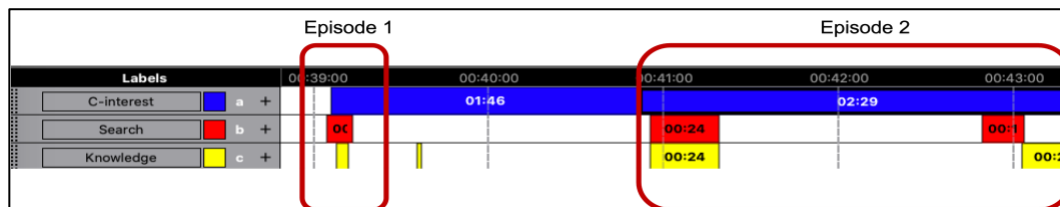


Figure 1. Screenshot of the Coding Timeline for Episodes 1 & 2

CONCLUSION

We conducted a pilot analysis of the interplay of interest, search, and knowledge—particularly from a child's real-life learning experiences during a maker workshop at a science museum. Our study serves as an initial attempt to empirically examine the ISL model (Liu & Jung, 2021) that conceptualizes the interplay of interest development, search, and knowledge construction. Also, our pilot analysis shows the potential of analyzing video data to understand the multidimensional transitions of children's interests, search, and knowledge. Particularly, our attempt can help scholars and practitioners from real-life information institutions, such as libraries and museums. Our coding methods and conceptual framework could also be reused in other learning and information-seeking environments. Limitations include that this study used a small-scale, single case and fully relied on video data. In future, we will conduct a larger-scale study with more participants, multiple data sources, and diverse learning settings.

REFERENCES

- Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41*(2), 111-127.
- Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *Journal of the Learning Sciences, 4*(1), 39-103. https://doi.org/10.1207/s15327809jls0401_2
- Jung, Y. J., Zimmerman, H. T., & Land, S. M. (2019). Emerging and developing children's situational interests during tablet-mediated biology education activities at a nature center. *Science Education, 103*, 900-922. doi:10.1002/sce.21514
- Liu, J., & Jung, Y. J. (2021). Interest development, knowledge learning, and interactive IR: Toward a state-based approach to search as learning. In *Proceedings of the 2021 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '21)* (pp. 239-248). ACM. <https://doi.org/10.1145/3406522.3446015>
- Nassauer, A., & Legewie, N. M. (2021). Video data analysis: A methodological frame for a novel research trend. *Sociological Methods & Research, 50*(1), 135-174.
- Renninger, K. Ann, & Hidi, S. E. (2016). *The power of interest for motivation and engagement*. Routledge.
- Rieh, S. Y., Collins-Thompson, K., Hansen, P., & Lee, H. J. (2016). Towards searching as a learning process: A review of current perspectives and future directions. *Journal of Information Science, 42*(1), 19-34.
- Scardamalia, M., & Bereiter, C. (1991). Higher levels of agency for children in knowledge building: A challenge for the design of new knowledge media. *Journal of the Learning Sciences, 1*(1), 37-68. https://doi.org/10.1207/s15327809jls0101_3

Supporting Prosocial Behaviour in Online Communities through Social Media Affordances

Kelly, Dominique	University of Western Ontario, Canada dkelly48@uwo.ca
Liu, Yifan	University of Western Ontario, Canada yliu4369@uwo.ca
Mayhew, Alex	University of Western Ontario, Canada amayhew@uwo.ca
Chen, Yimin	University of Western Ontario, Canada ychen582@uwo.ca
Cornwell, Sarah E.	University of Western Ontario, Canada scornwel@uwo.ca
Delellis, Nicole S.	University of Western Ontario, Canada ndelelli@uwo.ca
Rubin, Victoria L.	University of Western Ontario, Canada vrubin@uwo.ca

ABSTRACT

Affordances are action possibilities that emerge from the relationship between the properties of an object and an interacting agent's capabilities. This poster examines how one type of affordance— anonymity— is enabled or constrained by six features of social media platforms. Our work is one step in a broader agenda to: (1) identify affordances that influence users' behaviour in online communities; (2) outline the social media features that enable or constrain those affordances; and, (3) experimentally determine whether certain affordances, or combinations of affordances, support prosocial behaviour. Prosocial behaviour in information and communication technologies (ICTs) is broadly viewed here as benefiting other individuals in online communities, which in turn could increase the subjective well-being (SWB) of the community's members. SWB is characterized by high positive affect, low negative affect, and high life satisfaction.

KEYWORDS

Anonymity; Affordances; Social Media Design; Human-Computer Interaction; Prosocial Behaviour

INTRODUCTION

The term *affordance* was coined by ecological psychologist James Gibson to describe the action possibilities available in one's environment (Bucher & Helmond, 2018; Evans et al., 2017). Cognitive scientist Don Norman later popularised the concept of affordances in the fields of design and human-computer interaction (HCI). The presence of an affordance—that is, a possibility for action—is jointly determined by the relationship between the properties of an object and the capabilities of an interacting agent (Norman, 2013). For example, a chair affords lifting only if the interacting agent is physically capable of picking it up (Norman, 2013).

Research on social media affordances tends to focus on the “new dynamics or types of communicative practices and social interactions that various features afford” (Bucher & Helmond, 2018, p. 239). Evans et al. (2017) distinguish between the material *features* of an object, the *affordances* that features enable, and an affordance's possible *outcomes*. Whereas features are static, affordances are dynamic and emerge from the relationship between the user, the object, and its features (Evans et al., 2017; Treem & Leonardi, 2013). Moreover, unlike features, affordances “are not binary”: they vary in degree (Evans et al., 2017, p. 40). Lastly, affordances “invite behaviors and other outcomes” (Evans et al., 2017, p. 40). A social media feature could afford searchability and visibility of content, which could in turn allow a user to locate another user's profile photo (Evans et al., 2017).

Prior work has linked the digital features of social media platforms to the affordances that each enables. For example, the affordance of *persistence*—broadly defined as information remaining accessible rather than expiring—could be enabled by a feature that displays the past activity of individuals on a platform (Treem & Leonardi, 2013).

OBJECTIVES

A growing body of work has applied concepts from positive psychology to the design of information and communication technologies (ICTs) (Desmet & Pohlmeier, 2013; Riva et al., 2012). To our knowledge, however, no research has specifically examined the relationship between the affordances of social media platforms and users' subjective well-being (SWB). Our work addresses this research gap by investigating how social media affordances could support prosocial behaviour in online communities, thereby increasing the SWB of the community's members.

Prosocial behaviour is “any action that provides benefit to others” (Olson et al., 2008, p. 462). It is a broad category of behaviour that “encompasses helping, but it also includes actions that are not necessarily intended to assist others, such as following the rules in a game, being honest, and cooperating with others in social situations” (Olson et al., 2008, p. 462). Prosocial behaviour has been shown to be related to SWB, a construct typically defined as high positive affect, low negative affect, and high life satisfaction (Heintzelman & Tay, 2017, p. 10). Gherghel et al. (2021) found that “[...] engagement in prosocial behavior was positively associated with empathic concern and

subjective well-being” (p. 4446). The positive relationship between prosocial behaviour and SWB may also be mediated by competence, autonomy, and relatedness (Feng & Zhang, 2021).

We aim to identify affordances that influence users’ behaviour in online communities and outline social media features that enable or constrain each affordance, as illustrated in the literature. The following section introduces one such affordance—anonymity—and explains how six social media features enable or constrain it.

EXAMPLE AFFORDANCE: ANONYMITY

Anonymity has been defined as “the degree to which a communicator perceives the message source is unknown and unspecified” in communication research (Anonymous, 1998, p. 387). According to Evans et al. (2017), anonymity meets the criteria necessary to be considered an affordance. Anonymity could lead to a variety of possible outcomes (Evans et al., 2017), such as open and intimate conversations and experimentation with new ideas (Bernstein et al., 2011). Indeed, empirical research shows that anonymity has “broad, varied, and inconsistent behavioural effects,” perhaps due in part to differences in how the concept has been operationalized (Burkell, 2006, p. 189). Furthermore, anonymity exists on a continuum: rather than fully anonymous or fully identified, a message source may be partially anonymous (Anonymous, 1998) and “[t]here is great variability in the degree of anonymity that particular objects afford” (Evans et al., 2017, p. 41).

In the context of social media, Correa et al. (2015) distinguish between the protection of online and offline identities. On Facebook, for instance, users are expected to post from a single online identity (i.e., their account) that matches their offline identity (i.e., their real name, which serves as their username) (Correa et al., 2015). Together, these features afford Facebook users little to no anonymity. Meanwhile, features could increase a user’s degree of anonymity by allowing them to hide their offline identity (e.g., through the adoption of pseudonyms) and/or hide their online identity (e.g., by not requiring the creation of user accounts, so that posts are unlinked to any single identity, pseudonymous or otherwise).

Following Treem and Leonardi’s (2013) approach of linking social media features to particular affordances, Table 1 presents a non-exhaustive list of features that increase or decrease online users’ anonymity.

Social media feature	Description	Impact on anonymity*	Illustration in literature
Profile photos	Users are allowed (and encouraged) to upload profile photos that reflect their corporeal selves. For example, the default profile photo icon could be a silhouette of a face, suggesting that a selfie is an appropriate kind of photo to upload.	-	(Cirucci, 2015)
Pseudonyms	Users can adopt a pseudonym as their username instead of entering their real name.	+	(Bernstein et al., 2011; Correa et al., 2015; Hogan, 2013)
No profile pages	Users cannot create profile pages. Consequently, other users cannot navigate posts linked to any particular username.	+	(Correa et al., 2015)
Shared usernames	Multiple users can share the same username.	+	(Bernstein et al., 2011; Correa et al., 2015)
Non-persistent usernames	Users can change their username at any time.	+	(Bernstein et al., 2011; Correa et al., 2015)
No user accounts	Users cannot (or are not required to) create accounts.	+	(Bernstein et al., 2011)

Table 1. Social Media Features Impacting Users’ Anonymity

CONCLUSION

Our work builds upon prior research that has examined the affordances of social media platforms. We intend to: (1) identify additional affordances that influence users’ behaviour in online communities; (2) outline the social media features that enable or constrain those affordances; and, (3) experimentally determine whether certain affordances, or combinations of affordances, support prosocial behaviour and consequently increase the SWB of an online community’s members. As a result of our research, we aim to recommend features that could be implemented in social media platforms to encourage prosocial behaviour. These recommendations could contribute to the development of ethical Mass Influence Systems (MISs) that promote human flourishing (Mayhew et al., 2022).

* “-” indicates decrease, and “+” indicates increase

ACKNOWLEDGMENTS

This work has been funded by the University of Western Ontario Strategic Support for the Canadian Sciences and Humanities Research Council (SSHRC) Success grant, awarded to Dr. Rubin in fall 2021. The grant was entitled "Rewiring for Happiness: Computing for Kinder, Healthier, and Compassionate Life Experiences with Information Commutation Technologies (ICTs)."

REFERENCES

- Anonymous. (1998). To reveal or not to reveal: A theoretical model of anonymous communication. *Communication Theory*, 8(4), 381-407. <https://doi.org/10.1111/j.1468-2885.1998.tb00226.x>
- Bernstein, M., Monroy-Hernández, A., Harry, D., André, P., Panovich, K., & Vargas, G. (2011). 4chan and /b/: An analysis of anonymity and ephemerality in a large online community. *Proceedings of the International AAAI Conference on Web and Social Media*, 5(1), 50-57. <https://ojs.aaai.org/index.php/ICWSM/article/view/14134>
- Bucher, T., & Helmond, A. (2018). The affordances of social media platforms. In J. Burgess, T. Poell, & A. Marwick (Eds.), *The SAGE Handbook of Social Media* (pp. 233-253). SAGE Publications.
- Burkell, J. (2006). Anonymity in behavioural research: Not being unnamed, but being unknown. *University of Ottawa Law & Technology Journal*, 3(1), 189-204. <https://ssrn.com/abstract=999692>
- Cirucci, A. M. (2015). Facebook's affordances, visible culture, and anti-anonymity. *Proceedings of the 2015 International Conference on Social Media & Society*, 1-6. <https://doi.org/10.1145/2789187.2789202>
- Correa, D., Silva, L., Mondal, M., Benevenuto, F., & Gummadi, K. (2015). The many shades of anonymity: Characterizing anonymous social media content. *Proceedings of the International AAAI Conference on Web and Social Media*, 9(1), 71-80. <https://ojs.aaai.org/index.php/ICWSM/article/view/14635>
- Desmet, P. M. A., & Pohlmeier, A. E. (2013). Positive design: An introduction to design for subjective well-being. *International Journal of Design*, 7(3), 5-19. <https://ijdesign.org/index.php/ijdesign/article/view/1666/595>
- Evans, S. K., Pearce, K. E., Vitak, J., & Treem, J. W. (2017). Explicating affordances: A conceptual framework for understanding affordances in communication research. *Journal of Computer-Mediated Communication*, 22(1), 35-52. <https://doi.org/10.1111/jcc4.12180>
- Feng, L., & Zhang, L. (2021). Prosocial tendencies and subjective well-being: The mediating role of basic psychological needs satisfaction. *Social Behavior and Personality: An International Journal*, 49(5), 1-10. <http://dx.doi.org/10.2224/sbp.9986>
- Gherghel, C., Nastas, D., Hashimoto, T., & Takai, J. (2021). The relationship between frequency of performing acts of kindness and subjective well-being: A mediation model in three cultures. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues*, 40(9), 4446-4459. <http://dx.doi.org/10.1007/s12144-019-00391-x>
- Heintzelman, S. J., & Tay, L. (2017). Subjective well-being: Payoffs of being happy and ways to promote happiness. In D. S. Dunn (Ed.), *Positive Psychology: Established and Emerging Issues* (1st ed., pp. 9–28). Routledge. <https://doi.org/10.4324/9781315106304>
- Hogan, B. (2013). Pseudonyms and the rise of the real-name web. In J. Hartley, J. Burgess, & A. Bruns (Eds.), *A Companion to New Media Dynamics* (1st ed., pp. 290-307). Blackwell Publishing Ltd. <https://doi.org/10.1002/9781118321607.ch18>
- Mayhew, A., Chen, Y., Cornwell, S. E., Delellis, N. S., Kelly, D., Liu, Y., & Rubin, V. L. (2022). Envisioning ethical mass influence systems. *The 85th Annual Meeting of the Association for Information Science and Technology (ASIS&T2022): "Information for a Sustainable World: Addressing Society's Grand Challenges,"* October 29 - November 1, 2022, Pittsburgh, PA.
- Norman, D. (2013). *The design of everyday things*. Basic Books.
- Olson, J. M., Breckler, S. J., & Wiggins, E. C. (2008). *Social psychology alive* (1st Canadian ed.). Thomson Nelson.
- Riva, G., Baños, R. M., Botella, C., Wiederhold, B. K., & Gaggioli, A. (2012). Positive technology: Using interactive technologies to promote positive functioning. *Cyberpsychology, Behavior, and Social Networking*, 15(2), 69-77. <https://doi.org/10.1089/cyber.2011.0139>
- Treem, J. W., & Leonardi, P. M. (2013). Social media use in organizations: Exploring the affordances of visibility, editability, persistence, and association. *Annals of the International Communication Association*, 36(1), 143-189. <https://doi.org/10.1080/23808985.2013.11679130>

A Preliminary Analysis between Available and Shared Resources for the Lebanese Job Crisis

Khoury, Christy

Syracuse University, USA | cjkhoury@syr.edu

ABSTRACT

The Port of Beirut, Lebanon explosion on August 4, 2020, forced increased emigration of the Lebanese people, as they sought stability outside of their economically and politically volatile country (Farhat, 2021). Thus, families remaining in Lebanon are more dependent than ever on remittances. However, the ability of families to receive remittances is contingent on the success of those emigrating from Lebanon to gain employment. Job seekers often turn to the internet for career information from communication platforms like Reddit and dedicated job forums, like Jobs for Lebanon, among other institutional resources. This study focuses on these online institutional resources' hyperlink network, as hyperlink networks can represent a collective public good (Fu & Shumate, 2016). Data collection resulted in a network of 148 institutions. Yet with the plethora of resources available, the Lebanese community is still undergoing ineffective job searching. This preliminary research explores the barrier between institutionally presented job resources and Lebanese job seekers by comparing the structure of the existing institutional resource network to resources shared in a leading online community. A network analysis of 148 institutions reveals: (1) unproductive job information exchange between major institutions, and (2) an opportunity for institutions to better support Lebanese job seekers.

KEYWORDS

Lebanon; Job Seeking; Network Analysis; Information Diffusion; Information Exchange

INTRODUCTION

On August 4, 2020, one of the largest non-nuclear explosions in history pulverized the Port of Beirut and damaged over half the city (Helou et al., 2021). The explosion activated global crisis response efforts to aid the hundreds killed and thousands injured or displaced from the damaged city infrastructure. Now in 2022, the Lebanese economy and people are still hurting as hyper-inflation increased the poverty rate to 78% (Human Rights Watch, 2021). Corrupt government officials, political instability, and the exponentially worsening overall situation is provoking increased emigration as Lebanese people seek a better future for themselves and their families. Given Lebanon's high dependence on remittances, emigrants' success is pertinent to the livelihood of Lebanese citizens (Awdeh, 2014). Many institutions and non-governmental organizations (NGOs) have observed this vulnerability and created or supported organizations in response. These institutions include 'Jobs for Lebanon' or 'Hire Lebanese', which are job search sites specific to the Lebanese job seekers. With a majority of Lebanese people still unsuccessful in their search (Farhat, 2021), their declining morale indicates that these resources are not reaching or fitting the intended audience. This informs the following critical questions that this preliminary study aims to explore:

- 1) What institutional information resources are *available* online for Lebanese people seeking job opportunities?
- 2) What gap, if any, is revealed when comparing (1) to job resources *shared* in online communities?

RELATED WORK

This work is related to information exchange under the larger information behavior umbrella. Popularly studied through social network analysis, information exchange studies the relationships built by exchanging intangible resources like information or social support (Haythornthwaite, 1996). The actors exchanging the resources can be institutions, like NGOs, or individuals in online communities. Analyzing patterns of information exchange in a network indicates the likelihood of actors being exposed to, and thus using, resources (Haythornthwaite, 1996). There is prior work on information and resource exchange, yet none focused on the ongoing Lebanese crisis. Given the context described in the previous section, the importance of effective job resource exchange cannot be undermined. This work expands earlier scholarship by using a transformative mixed methods framework, which enables researchers to advocate for people or a community who have experienced oppression and/or discrimination from race/ethnicity, immigrant status, political conflicts or other harm (Creswell & Creswell, 2018). By applying a transformative framework to the Lebanese community, which has openly discussed the difficulties of job seeking after the Port of Beirut explosion, I investigate and bring voice to a crucial concern at the intersection of information exchange and crisis studies.

METHODS

There were two primary methods facilitated in this preliminary study: network analysis of the hyperlink network and a content analysis of Reddit page r/Lebanon.

Network Analysis

A network analysis was conducted to understand how the network of NGOs informs the overall institutional structure of the Lebanese job-seeking resource space. Data for the institutional hyperlinks were collected manually, beginning with the most popular Lebanese Job resource “Jobs for Lebanon” (JFL). For each site (i.e., resource), the partners and sponsors were denoted as targets. Each target’s website was then used as a source, and the process repeated until no new institutions were identified. If a sponsor was not an NGO, or not dominantly relevant to the research question (e.g., Facebook), the sponsor would not be listed as a target. This was done to keep the network boundary refined to the scope of the study. This manual website crawling process resulted in an edge list of 148 nodes used to analyze different measurements of centrality.

Content Analysis

A content analysis was conducted to compare the institutions collected via the website crawl against the information shared amongst the Lebanese online community via Reddit’s subreddit r/Lebanon. This online community has over 80,000 members exchanging information about a variety of topics including job opportunities, career advancement tips and job seeking resources. To collect the public data from r/Lebanon, a Python-based API (subreddit_downloader.py) was used to gather the last 3 months of data. Posts were manually reviewed by the author for any resource used to answer a job inquiry which was then denoted as a shared resource (Krippendorff, 1989).

RESULTS

The hyperlink analysis, conducted by metrics discussed next, revealed dispersed information resources for Lebanese professional development or job seeking sites. Network density reflects the overall strength and interconnectedness between actors in a network (Sohn, 2009). Closeness centrality is a metric that informs how long the information will take to spread between vertexes in the network, while betweenness centrality measures the importance of a node in connecting other nodes in the network (Bloch et al., 2021). Lastly, eigenvector centrality measures the influence of a node in a network with a relative score (Bloch et al., 2021). The low overall network density of 0.0095 indicates an opportunity for better connected information resources to enable quicker and more widespread information exchange. A more connected network allows for a more effective and rapid job information diffusion process, enabling an improved search process for the Lebanese community. JFL unsurprisingly had the highest betweenness centrality signaling significance of connecting other nodes. However, JFL’s low closeness centrality of 0.112 demonstrates a slow information flow impacting the efficacy of job information diffusion to other resources.

Furthermore, these findings demonstrate how resources not focused on career advancement or job postings – such as Beirut Digital District (BDD) and Outerpond – are also central in the network when considering their eigenvector centrality scores. BDD ranked first for eigenvector centrality, while Outerpond followed JFL by ranking third. BDD provides eco-friendly working environments and event spaces for professionals, while the Outerpond is e-bridge platform promoting cross-border collaboration with Lebanese businesses. Both institutions demonstrate organizations whose sole focus is not disseminating career advancement opportunities for Lebanese but do in fact play a large role in diffusing job information. A similar finding was discovered when analyzing betweenness centrality scores, as BDD, Outerpond, Chamber of Commerce Lebanon-France and ESA Business School had comparable scores to JFL. This noteworthy finding begs for two follow-up question: (1) how can institutions not solely focused on career advancement better support Lebanese job seekers? And, (2) are these institutions aware of their assistive role in the job crisis, and what observable efforts can be seen?

When comparing the institutions with the highest centrality scores in the network to the institutions shared via Reddit’s r/Lebanon, there is only an approximate 6% overlap. In other words, a vast majority of the resources that are prominent in the hyperlink network are not shared in one of the largest Lebanese online communities. Additionally, job seeking resources that had very low centrality in the network were shared in the subreddit due to their focus on hiring in Lebanon and the greater Middle East. Examples include Naukri Gulf and Hire Lebanese, which were signaled on Reddit as helpful job sites by “pinning” the resources but calculated to each have low closeness centrality scores of approximately 0.0081 compared to JFL’s 0.112.

CONCLUSION

Conducting the preliminary analysis on the hyperlink network of Lebanese job resources revealed a poorly connected network structure. When comparing this network analysis to the dialogue on a popular Lebanese online community, r/Lebanon, a disconnect between available and used resources is revealed. This disconnect cannot be resolved without future participatory research to better understand why the central institutions in the network are not shared in online communities. This may be a case of inaccessibility, information overload, or another unobserved cause. Whatever the root issue may be, future investigative research must be conducted to better support and empower the Lebanese job seekers eagerly searching for stability. This research can then be transferred to better understand career information gaps found in other communities currently facing related disparities.

REFERENCES

- Awdeh, A. (2014). Remittances to Lebanon: Economic Impact and the Role of Banks. *ESCWA Publications*.
- Bloch, F., Jackson, M. O., & Tebaldi, P. (2021). Centrality Measures in Networks. *ArXiv:1608.05845 [Physics]*.
<http://arxiv.org/abs/1608.05845>
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (Fifth edition). SAGE.
- Farhat, S. (2021, December 15). *Battling Lebanon's talent exodus with remote work for global companies*. France 24.
<https://www.france24.com/en/middle-east/20211215-battling-lebanon-s-talent-exodus-with-remote-work-for-global-companies>
- Fu, S., & Shumate, M. (2016). Hyperlinks as Institutionalized Connective Public Goods for Collective Action Online. *Journal of Computer-Mediated Communication*, 21. <https://doi.org/10.1111/jcc4.12161>
- Haythornthwaite, C. (1996). Social network analysis: An approach and technique for the study of information exchange. *Library & Information Science Research*, 18(4), 323–342. [https://doi.org/10.1016/S0740-8188\(96\)90003-1](https://doi.org/10.1016/S0740-8188(96)90003-1)
- Helou, M., El-Hussein, M., Aciksari, K., Salio, F., Della Corte, F., von Schreeb, J., & Ragazzoni, L. (2021). Beirut Explosion: The Largest Non-Nuclear Blast in History. *Disaster Medicine and Public Health Preparedness*, 1–2.
<https://doi.org/10.1017/dmp.2021.328>
- Human Rights Watch. (2021). Lebanon: Events of 2021. In *World Report 2022*. <https://www.hrw.org/world-report/2022/country-chapters/lebanon>
- Krippendorff, K. (1989). Content Analysis. *International Encyclopedia of Communication*, 1, 403–407.
- Sohn, D. (2009). Disentangling the Effects of Social Network Density on Electronic Word-of-Mouth (eWom) Intention. *Journal of Computer-Mediated Communication*, 14(2), 352–367. <https://doi.org/10.1111/j.1083-6101.2009.01444.x>

Evaluating Visitor Engagement through Instagram: A Case Study of *One with Eternity: Yayoi Kusama in the Hirshhorn Collection*

Kim, Jinyi George Mason University, USA | jkim266@gmu.edu

Lee, Myeong George Mason University, USA | mlee89@gmu.edu

ABSTRACT

Museums invite visitors to actively engage in exhibitions. However, it is unclear how visitors experience and appreciate exhibitions, because (1) people's motivations and goals to visit museums vary significantly and (2) scholars have yet to standardize a typology of observable visitor experiences. This study explores the potential of social media postings to assess museum visitors' engagement. Images posted on Instagram at an exhibition, *One with Eternity: Yayoi Kusama in the Hirshhorn*, are examined as a case study to explore visitor engagement using social media. A qualitative analysis of the 964 Instagram images posted for the exhibition suggest that Instagram can surface visitors' experiences with artifacts presented in museums: they enjoy interactive and immersive experiences in the exhibition. This preliminary study implies that there are opportunities to extend this approach to facilitate visitor interactions with exhibitions through social media.

KEYWORDS

visitor engagement; Instagram image analysis; museum visitor experience

INTRODUCTION

Museums are sites of informal education (Falk & Deirking, 2002; Hein, 2002) where visitors are directed by their own interest (Rogoff et al., 2016). Engagement is a critical part of the informal learning process (Budge & Burness, 2018). Accordingly, museums create visual learning environments for visitors to establish their individual interpretative strategies (Hooper-Greenhill, 2000). Exhibitions are often designed for visitors to engage in personal meaning-making and facilitate learning. *Meaning-making* describes how people interpret their environment and understand their lives based on their individual knowledge and experience (National Research Council, 2000). Studies show that visitors construct meanings from exhibitions that may differ from what the artists intended (Adams et al., 2003). Social media provides museums with new opportunities to understand how visitors engage with exhibits and construct meanings from them (Budge, 2017). Previous research on museum visitors' use of Instagram has focused on communication (Weilenmann et al., 2013), place-making (Budge, 2018), aesthetic inspiration (Budge, 2017), and memory (Rhee et al., 2021). However, research addressing the relationship between social media and informal learning experiences remains limited. This study explores social media's potential to understand visitor experiences. To understand how museum visitors use social media and how the technology moderates their interaction with museum artifacts, we analyze 964 Instagram images from *One with Eternity: Yayoi Kusama in the Hirshhorn Collection* (Hirshhorn, n.d.). Qualitative analysis highlights visitors' interactions with the exhibition and their visual meaning-making. Analyzing volumes of Instagram images allows us to understand how visitors engage in exhibitions and what visitors value from their museum experiences. Specific questions we ask in this preliminary case study are:

1. What kinds of images do visitors post using the Instagram hashtag #eternalkusama?
2. What types of engagement can Instagram images reveal?

DATA

The selected exhibition is *One with Eternity: Yayoi Kusama in the Hirshhorn Collection*. Kusama's worldwide popularity, highly Instagrammable visuals, and the exhibition timeline accorded well with this research. The exhibition displayed five of Kusama's works: *Infinity Mirror Room—Phalli's Field* (2017) (hereafter "*Field*"), *Infinity Mirrored Room—My Heart is Dancing into the Universe* (2018) ("*Universe*"), *Pumpkin* (2016), *Overcoat* (1964) ("*Coat*"), and *The Hill* (1953) ("*Hill*") at the Hirshhorn Museum and Sculpture Garden, Washington, D.C. Between April 1 and May 30, 2022, 964 images under the #eternalkusama tag were collected through 4K Stogram Instagram downloader software (4K Download, 2022).

METHODS

For the first research question, we classified the Instagram images tagged #eternalkusama into any of the five categories of Kusama's works (defined by the artist) using Google AutoML Vision (Google, 2018). A single-label classification model was built by training 140 images of her work that were already classified by the artist (F1=.97). Then, the 964 Instagram images were predicted using the trained model. If the confidence level of the prediction

was higher than 0.65, the image was classified into one of the four categories (The Hill category was removed, because there were not enough training images). In cases where the confidence level was less than 0.65, they were classified as *Others* (denoting non-exhibition images). To answer the second question, we qualitatively coded the images by focusing on visitors' visual meaning-making practices. We generated themes in an inductive way to identify emerging themes from the photos (Namey et al., 2011).

FINDINGS

In the AutoML-based image classification analysis, images of exhibition objects are dominant, compared with the images of non-exhibition objects. However, this still demonstrates visitors' self-selection of subjects from the exhibition that curators designed. Of these categories, "objects" is the most frequent. Most object images feature interactive, visually and spatially impressive works rather than two-dimensional objects hung on the wall. For example, *Field*, *Universe*, and *Pumpkin* are frequently posted object images that are placed in the exhibition space where visitors can physically walk around them.

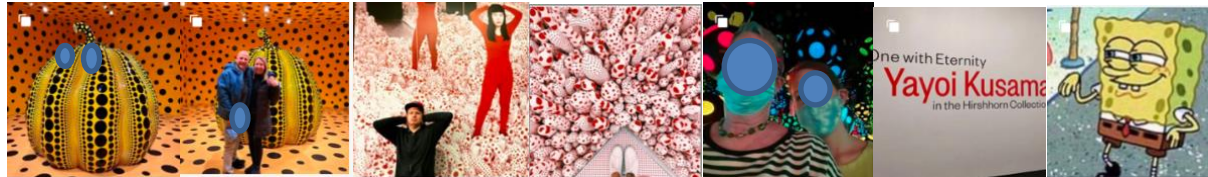


Figure 1. Codes: objects, standing, performing, re-curating, selfies, information, interpretation
(from left to right)

Emerging Themes of Visitors' Experiences

Codes that emerged are *objects*, *standing*, *performing*, *re-curating*, *selfies*, *information*, and *interpretation* (Figure 1). *Objects* refers to photos of artworks visitors take in the exhibition. *Standing* is distinguished from *performing* by how visitors contextualize themselves with exhibition objects; *performing* includes mimicry and interaction. *Re-curating* occurs when visitors re-frame the exhibition to express their aesthetics or creativity, such as taking a photo from a different angle, adding filters, or enlarging a particular portion of the work. *Selfies* are self-portraits focusing on an individual's face, making the exhibit a backdrop (Kozinets et al., 2017). *Information* includes exhibition titles, wall texts, snapshots of exhibition news articles, and exhibition spaces. *Interpretation* occurs when visitors produce images related to what they saw in the exhibition (e.g., fashion items or other objects featuring Kusama's iconic pattern) or recontextualize images external to the exhibition (e.g., SpongeBob).

Among the types of visitors' engagements with the exhibition defined in the image codes, standing is the second popular type of images after objects. Many visitors took photos standing in the exhibition. This suggests that visitors may self-record their visits for memory or social display. Re-curated images are unique way of experiencing the exhibition. They demonstrate visitors' meaning-making practices with the exhibition content. Most images coded *selfie* co-occur with *Universe* and *Field* images because of the visually pleasing backdrops. Museum selfies deliver an evident message: "I was there, I came, I saw, and I selfied." (Fei, 2016, 03:55). An image capturing the outside of an exhibition installation presents the atmosphere of before and after the exhibition. Visitors post images to communicate information with online users, and these posts can complement museums' official exhibition information. Visitors interpret the exhibition by posting personal responses to the exhibition as after-visit activities. Some images represent daily life activities associated with the exhibition objects or artists.

Overall, the Instagram image data reveal that visitors experience the exhibition by engaging with the artifacts in various ways, which are essential to informal learning. The identified characteristics of the *One with Eternity* exhibition Instagram images are that, (a) by taking and posting photos, visitors document their experiences and enact their meaning-making through re-curated social media images; (b) visitors prefer taking photos in immersed experiences? such as mirror room installations, rather than taking a third-person perspective for artifacts, (c) visitors post their responses and exhibition information to communicate with their social networks, (d) selfies in museums are not just self-photos of a person, but a self-photo about where they are, and (e) visitors enjoy creating aesthetic or artistic expressions through social media.

DISCUSSION & FUTURE WORK

These findings suggest that museums can utilize social media not only to promote exhibitions but also to understand visitors' engagement and preferences. From an informal learning perspective, we can see visitors use Instagram to engage with exhibitions and expand their interpretations through self-reflective postings. Although this pilot study has a dataset of only 964 images, its methodological implications will help museum professionals leverage social media data to inform strategic decision making. Future studies can extend this research with a mixed-method approach, such as combining Instagram image-text analysis and interviews with Instagram, furthering understanding about people's informal learning processes in museums and their technology-mediated experiences of exhibitions.

REFERENCES

- Adams, M., Falk, J. H., & Dierking, L. D. (2003). Things change: museums, learning, and research. In M. Xanthoudaki, L. Tickle, & V. Sekules (Eds.), *Researching visual arts education in museums and galleries* (pp. 15-32). Springer, Dordrecht.
- Budge, K. (2017). Objects in focus: Museum visitors and Instagram. *Curator: The Museum Journal*, 60(1), 67-85.
- Budge, K., & Burness, A. (2018). Museum objects and Instagram: agency and communication in digital engagement. *Continuum*, 32(2), 137-150.
- Budge, K. (2018). Encountering people and place: Museums through the lens of Instagram. *Australasian Journal of Popular Culture*, 7(1), 107-121.
- 4K Download. (2022). *4K Stogram* [Computer software]. <https://www.4kdownload.com/products/stogram/1>
- Falk, J., & Dierking, L. (2002). *Learning from museums: visitor experiences and the making of meaning*. AltaMira.
- Fei, J.J. (2016, March 2). *Art in the Age of Instagram* [video]. YouTube. <https://www.youtube.com/watch?v=8DLNFDQt8Pc>
- Google. (2018). *Google AutoML Vision* [Computer software]. <https://cloud.google.com/vertex-ai>
- Hein, G.E. (2002). *Learning in the museum*. Routledge. <https://doi.org/10.4324/9780203028322>
- Hirshhorn Museum. (n.d.). One with Eternity: Yayoi Kusama in the Hirshhorn Collection Apr 01–Nov 27, 2022. Retrieved June 1, 2022, from <https://hirshhorn.si.edu/exhibitions/one-with-eternity-yayoi-kusama-in-the-hirshhorn-collection/>
- Hooper-Greenhill, E. (2020). *Museums and the interpretation of visual culture*. Routledge.
- Kozinets, R., Gretzel, U., & Dinhopl, A. (2017). Self in art/self as art: Museum selfies as identity work. *Frontiers in Psychology*, 8, 731.
- Namey, E. E., MacQueen, K. M., Guest, G. (2011). *Applied Thematic Analysis*. SAGE Publications.
- National Research Council. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition*. National Academies Press.
- Rogoff, B., Callanan, M., Gutiérrez, K. D., & Erickson, F. (2016). The organization of informal learning. *Review of Research in Education*, 40(1), 356-401.
- Rhee, Pianzola, F., & Choi, G. (2021). Analyzing the Museum Experience Through the Lens of Instagram Posts. *Curator*, 64(3), 529–547. <https://doi.org/10.1111/cura.12414>
- Stylianou-Lambert, T. (2017). Photographing in the art museum: Visitor attitudes and motivations. *Visitor Studies*, 20(2), 114-137.
- Weilenmann, Hillman, T., & Jungselius, B. (2013). Instagram at the museum: communicating the museum experience through social photo sharing. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1843–1852. <https://doi.org/10.1145/2470654.2466243>

The Transitioning of the Term “Digital Library” in Response to Information Resilience: A Thematic Mapping of Literature From 2000-2021

Kumari, Madhuri

Central University of Gujarat, India | madhurik909@gmail.com

Gala, Bhakti

Central University of Gujarat, India | bhakti.gala@cug.ac.in

ABSTRACT

This study aims to explore the transition of the term “digital library” through a thematic mapping of published literature. To reach this purpose, we conducted a network analysis on the retrieved bibliographic dataset from the Scopus database. From 985 documents total of 77 high-frequency keywords have been selected. Co-occurrence of the top 20 keywords has been identified to get indicative themes. Among the top 20 high-frequency keywords, the term “human” is identified as an emerging term based on the average publication year 2017. The terms from the blue-colored cluster are “human”, “disease” “diagnosis” and many more to support information resiliency due to expanding areas in the context of the digital library.

KEYWORDS

Digital library; Information Resilience; Thematic mapping

INTRODUCTION

“Information is power,” said Francis Bacon. Providing the right information to the right person at the right time is an adage echoed by S R Ranganathan to library professionals. So, it is crucial to generate, manage, and make information easily accessible to a specific community or group of communities through digital libraries (IFLA, 2012; Xie & Matusiak, 2016). Yugandhar (2022) found that the development of digital libraries in India has been skewed, prompting the authors of this study to further explore this phenomenon. The study is conducted to identify two objectives. Firstly, to get the indicative terms by the co-occurrence of high-frequency keywords from the published literature in India with a time range from 2000 to 2021. Secondly, to know the community themes of interest and the emerging terms within the term “Digital Library”. The sustained emergence of the term “Digital Library” over a span of a decade among the community indicates information resilience. Findings from this study would guide the LIS professionals to set their narratives and expand the scope of research.

METHODOLOGY

Scopus database was used to retrieve the bibliometric data for analysis. As of 15th February 2022, the published literature was searched with the keywords “digital librar*” and “India” in English. “AND” Boolean operator has been used to get a precise result. A total of 985 documents were retrieved. The method of network analysis had been taken as a research design to identify leading and emerging topics. The formula given by Kauffman & Group (1969) was used to find the transition between high and low frequency among the frequency list of keywords (Guo et al., 2017). A total number of 6719 keywords were retrieved from the corpus by considering author and index keywords. Author keywords are those assigned by the authors of the retrieved papers, whereas Indexed keywords are those assigned by the database to the papers. Using Biblioshiny bibliometric software, the number of keywords retrieved having only one frequency was 2900 (I_1).

$$N = \frac{1}{2}(-1 \pm \sqrt{1 + 8I_1}) \quad (N = \text{Transition point}, I_1 = \text{Number of keywords with the frequency of one}).$$

Using the above formula, the researchers reached a value N equal to 77. Therefore, for further analysis, 77 keywords were considered for conducting the study using the software VOSviewer. To identify an emerging term, the average publication year of documents containing a keyword or term, published by a country is considered (Waltman et al., 2010). The smart local moving algorithm technique (SLM) (Waltman & Jan Van Eck, 2013) was used to demonstrate the clustered keywords.

DATA ANALYSIS

The top 20 indicative terms are listed in Table 1. The rank of keywords is based on their frequency of occurrence in published literature.

Rank	Keywords	Occurrences	Links	Total link strength	Avg. pub. year
1	digital libraries	645	334	3268	2012.879
2	digital library	132	172	418	2014.78
3	information retrieval	100	181	581	2011.93
4	information services	74	151	455	2011.878
5	libraries	74	170	434	2012.027
6	India	66	91	252	2009.955
7	search engines	59	136	387	2012.915
8	database systems	50	151	389	2007.88
9	semantics	40	106	248	2013.6
10	information technology	37	105	265	2005.892
11	world wide web	35	103	223	2007.2
12	human	33	111	342	2017.667
13	internet	32	89	180	2007.656
14	metadata	31	96	187	2010.226
15	computer software	30	87	169	2008.633
16	information management	30	96	191	2009.133
17	open-source software	30	76	169	2014.6
18	article	29	100	339	2015.931
19	societies and institutions	29	91	215	2007
20	e-learning	28	78	147	2015.357

Table 1. Top 20 High-Frequency Indicative Keywords in the Digital Library Research Field

Figure 1 shows the clusters of community themes derived from the co-occurrence of high-frequency keywords. The cluster shows the emerging themes in relationship with the term “digital library”. The terms “human”, “disease” “diagnosis”, “molecular docking”, “virtual reality”, “virtual screening”, “article”, and “priority journals” from this cluster, point to the extension of the scope of the term “digital library”.

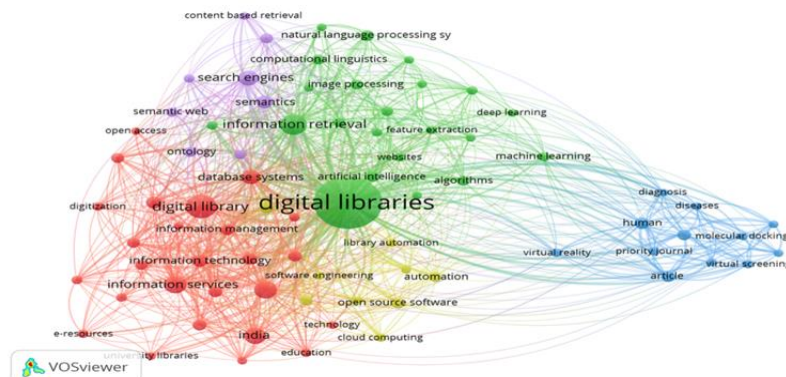


Figure 1. Visualization of Community Theme Co-Occurrence Network of High-Frequency Keywords

By using the clustering technique, the formation of a blue-colored cluster to the extreme right was observed. The term “human” from the blue-colored cluster was identified as a recent keyword based on the average publication year (2017).

DISCUSSION AND CONCLUSION

The study helps in enhancing our understanding of the trajectory of the term “digital library” towards humankind by the blue-color clustered terms (i.e., “human”, “disease”, “diagnosis”, “virtual reality”, “molecular docking”, article” and “priority journal”). These terms belong to journals like the *Journal of Molecular Structure*, *Computers in Biology and Medicine*, *Journal of Chemical Information and Modeling*, *Diagnostics*, etc. The readers of these journals play a significant role in the transition of information within the context of the digital library. This transition of the term ‘Digital Library’ indicates resilience in the evolution of digital library development in India. It also gives the initiation point to LIS professionals to be ready to adopt the emerging subject domain (Atkins et al., 2007). This finding is consistent with the world perspective too (Borgohain et al., 2021). A broader view is represented by high-frequency keywords of the digital library research thematic trends (Lis & Tomanek, 2020).

ACKNOWLEDGMENT

This collaboration is an outcome of the SIG-DL, HINDAS project.

REFERENCES

- Atkins, D., Fulker, D., Kornig, S., Ludascher, B., Diego, S., Center, S., Marlino, M., Savino, P., Seshagiri, N., Sumner, T. R., Wright, M., Walters, S., & Singhal, M. (2007). *Digital Libraries in Education, Science, and Culture*. <https://iite.unesco.org/pics/publications/en/files/3214660.pdf>
- Borghain, D. J., Zakaria, S., & Kumar Verma, M. (2021). Cluster Analysis and Network Visualization of Global Research on Digital Libraries during 2016–2020: A Bibliometric Mapping. *Science and Technology Libraries*. <https://doi.org/10.1080/0194262X.2021.1993422>
- Guo, D., Chen, H., Long, R., Lu, H., & Long, Q. (2017). A co-word analysis of organizational constraints for maintaining sustainability. *Sustainability (Switzerland)*, 9(10), 1–19. <https://doi.org/10.3390/su9101928>
- IFLA. (2012). IFLA Manifesto for Digital Libraries. *IFLA Journal*, 38(1), 88–93. <https://doi.org/10.1177/0340035211435775>
- Kauffman, S. a, & Group, N. P. (1969). Dispersion of Papers among Journals based on Mathematical Analysis of Two Diverse Medical Literatures. *Group*, 224, 177–178. <http://adsabs.harvard.edu/abs/1969Natur.224.177K>
- Lis, A., & Tomanek, M. (2020). Sport management: Thematic mapping of the research field. *Journal of Physical Education and Sport*, 20(April), 1201–1208. <https://doi.org/10.7752/jpes.2020.s2167>
- Waltman, L., & Jan Van Eck, N. (2013). *A smart local moving algorithm for large-scale modularity-based community detection*.
- Waltman, L., van Eck, N. J., & Noyons, E. C. M. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, 4(4), 629–635. <https://doi.org/10.1016/j.joi.2010.07.002>
- Xie, I., & Matusiak, K. K. (2016). Digital library collection development. In *Discover Digital Libraries* (pp. 37–58). Elsevier. <https://doi.org/10.1016/b978-0-12-417112-1.00002-8>
- Yugandhar, B. (2022). DIGITAL LIBRARIES IN INDIA: A REVIEW. In *Peer Reviewed and Refereed Journal* (Issue 1). <http://ijmer.in.doi./2022/11.04.05>

Differences in User Information Behavior between Official Media and Private Media during the COVID-19 Pandemic

Lei, Li Beijing Normal University, China | leili@bnu.edu.cn
Xuyan, Wang Beijing Normal University, China | wangxuyan0105@163.com
Shujun, Liu Beijing Normal University, China | shujun_liu@mail.bnu.edu.cn
Kun, Huang Beijing Normal University, China | huangkun@bnu.edu.cn

ABSTRACT

Weibo is a widely used social media platform showing all kinds of information related to the COVID-19 pandemic promptly in China. Official media and private media are two typical types of media on Weibo. Due to the different characteristics of these two media types, in the context of public health emergencies, it is worth exploring whether there are differences in the users' interactive behavior with information from these two types of media. This is of great significance to the integration and development of these two types of media. This study obtained data on the interaction behaviors of Weibo users with posts published by the two media types at various stages of the pandemic. Statistical analyses have confirmed significant differences in interaction behavior data between users and these two media types. In future research, based on the findings of this study, we will investigate the reasons behind these differences to provide relevant guidelines and suggestions for the release of different media in public health emergencies by conducting a deep dive analysis of user reviews.

KEYWORDS

User information behavior; COVID-19 pandemic; Official media; Private media; Weibo; China

INTRODUCTION

COVID-19 is attacking the world, inflicting havoc on human health and the world economy. In the current context, people are in urgent need of relevant pandemic prevention knowledge and information (Hughes, 2009). Social networking platforms provide people with the means of collecting and disseminating information. As one of the biggest social network platforms in China, with 511 million active users, Weibo offers dynamic and timely interactions (Feng, 2017). Groups of people come together on this platform to acquire and to interact with information (Reynolds, 2005). During this period, various media accounts on Weibo, including official media (supported by the government) and private media (supported by private corporations and individuals), play an essential role in spreading important information on pandemics and in guiding public opinion. However, the different nature of these two media types means that users may show different interaction behaviors in their published posts of the same event. We collected relevant data to verify the differences in user information behavior between the two different media types, and further distinguished the differences at different stages of the pandemic.

METHOD

Pandemic Stage Division

According to "China's Action against COVID-19," a white paper issued by The State Council Information Office of the People's Republic of China, we divided the research phase into five parts. Selection of time nodes for the first four stages from the white paper included, incubation period (27/12/2019 to 19/01/2020), outbreak period (20/01/2020 to 20/02/2020), controlling period (21/02/2020 to 17/03/2020), results appearing (18/03/2020 to 28/04/2020). The fifth phase of the recovery period starts and ends from April 29, 2020 to September 5, 2020. Since 31 provinces and municipalities in China achieved zero new local cases for 20 consecutive days on September 5, 2020, it means that the pandemic was temporarily effectively controlled on this day.

Event Selection

To ensure the richness of the acquired data, according to the popularity of events provided by Weibo, the representative events selected at each stage are the most concerned hot events in the current period. Starting at the second stage, we selected the top five popular events in each stage based on the pageviews of the events, e.g., Shuanghuanglian could inhibit COVID-19, this is Wuhan at 10 o'clock today, and Cinemas in low-risk areas will open for business on July 20. It is noteworthy that the first stage was a special time when most media platforms did not realize that the pandemic was spreading, so the number of relevant popular events was minimal. For this reason, only one representative event, four new COVID-19 cases in Wuhan, was selected at this stage.

Data Collection and Annotation

According to “China’s Action against COVID-19,” a white paper issued by The State Council Information Office of the People’s Republic of China, we divided the research phase into five parts. Selection of time nodes for the first four stages from the white paper included, incubation period (27/12/2019 to 19/01/2020), outbreak period (20/01/2020 to 20/02/2020), controlling period (21/02/2020 to 17/03/2020), results appearing (18/03/2020 to 28/04/2020). The fifth phase of the recovery period starts and ends from April 29, 2020 to September 5, 2020. Since 31 provinces and municipalities in China achieved zero new local cases for 20 consecutive days on September 5, 2020, it means that the pandemic was temporarily effectively controlled on this day.

Data Analysis

In order to eliminate the influence brought by the number of followers on its posts, we calculated the 1/10,000 percentage of comments, thumbs-up, and reposts of each blog post for the number of fans of each account, thereby obtaining the comment rate, thumbs-up rate, and repost rate for each post. We analyzed the differences in these rates between the official media and the private media at various stages of the pandemic. Because the dimensions of the comment rate, thumbs-up rate, and repost rate for these two media types are different, the rates were normalized. Normalized data does not follow the normal distribution. Therefore, the Mann-Whitney U test was used to analyze the differences.

RESULTS AND DISCUSSION

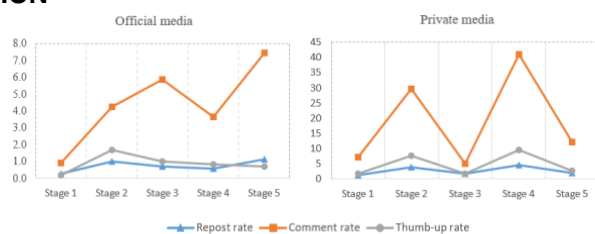


Figure 1. Average comment, thumbs-up, and repost rate of official and private media posts at each stage

Official media has a large number of fans. Therefore, the interaction ratio of official media is much smaller than that of private media. For a clear comparison of the interactions between users and these two types of media at different stages of the pandemic, Figure 1 separately shows the average comment rate, thumbs-up rate, and repost rate of posts published by the two media types at each stage. Figure 1 shows that the peaks of the interaction of the two types of media are on the comment rate, but the peaks of the comment rate of the two types of media are at different stages. The official media took the lead in the fifth stage, the recovery stage. The comment rate obtained by the official media from the first stage to the third stage gradually increased, and it dropped suddenly in the fourth stage, but peaked in the fifth stage. In the private media, the comment rate dropped to a minimum in the third stage, reaching its peak in the fourth stage, the results appearing period, but plummeted in recovery stage.

Stage of development	Repost rate	Comment rate	Thumbs-up rate
Stage 1	0.779	0.678	0.429
Stage 2	0.272	0.081	0.016
Stage 3	0.982	0	0.138
Stage 4	0	0	0
Stage 5	0	0	0.215
Total	0	0.151	0

Table 1. The p-values of the Mann-Whitney U test on the two types of media for each rate at each stage

Table 1 shows the significant value of the difference between the comment rate, thumbs-up rate, and repost rate of each stage for the two types of media. There are significant differences between the official media and private media in repost and comment rates as a whole. Specifically, there are significant differences in these rates in the fourth stages of the pandemic. There are significant differences in the comment rate and the repost rate in stages five. Stage 3 only has a significant difference in comment rate, and stage 2 only has a significant difference in thumb-up rate.

To sum up, it can be seen that users are more willing to comment on official media during the recovery period, and comment on private media during the results appearing period. The difference in the comment content of the same event under different media types may be an important factor leading to the difference in comment rate. The next research plan is to combine the expressions of different media types on the same event, compare the differences in user comment topics, emotions, etc., and analyze the factors that affect users’ behavior differences under different media types from the perspective of content characteristics.

CONCLUSION

This study found that at various periods in the development of the COVID-19 pandemic, there were significant differences in the comments, thumbs-up, and repost behaviors of Weibo users to posts published by official media and private media. The differences in user comments and bloggers' posts need further analysis.

REFERENCES

- Feng L. (2017). Maximizing micro-blog influence in online promotion. *Expert Systems with Applications*, 70, 52–66.
- Hughes, A. L. (2009). Twitter adoption and use in mass convergence and emergency events. *International Journal of Emergency Management*, 6(3-4), 248–260.
- Reynolds, B. (2005). Crisis and Emergency Risk Communication as an Integrative Model. *Journal of Health Communication*, 10(1), 43-55.

Intertwining Search and Non-Search Activities during Cross-Session Search Tasks

Li, Yuan University of North Carolina at Chapel Hill, USA | yuanli@live.unc.edu

Capra, Rob University of North Carolina at Chapel Hill, USA | rcapra@unc.edu

ABSTRACT

Users often conduct cross-session searches to look for information over time, especially for complex tasks. Cross-session search activities may be intermixed with periods of direct engagement to implement the work task (potentially based on information found during search). Much prior work has focused on analyzing the changes in search activities between sessions and on trying to predict task continuation. Less consideration has been paid to the non-search task activities and their influence on search sessions. We conducted a diary study to investigate cross-session search behavior in real-life tasks. Our preliminary analysis identified eight types of non-search task activities people conducted in-between their cross-session search processes. We also observed three different patterns regarding how users' search sessions intertwined with non-search activities. Our findings provide evidence-based insights about cross-session search process in real-world task contexts and provide insight into the underlying reasons for users' search behaviors during cross-session search.

KEYWORDS

Cross-session search; Information use; Non-search task activity; Information seeking process

INTRODUCTION

Cross-session search (XSS) is common in our everyday work and life. It describes a series of information search sessions that people conduct at different time points to complete a task or a project. Previous studies mainly focus on users' *search activities* during the XSS process. For example, research on supporting XSS has investigated: topics, queries, session length, search devices, information keeping methods, and task continuation prediction (Agichtein et al., 2012; Han, He & Chi, 2017; Kotov et al., 2012; MacKay & Watters, 2009; Morris et al., 2008; Wang et al., 2013; Wu et al., 2018). Researchers have also studied the reasons that lead people to search across multiple sessions, including the reasons why they resume a search session and why they stop during a cross-session search process (Lin and Belkin, 2005; Spink et al., 2002, MacKay and Watters, 2009).

Work tasks and task processes have an important effect on users' information search behavior (Capra, Arguello, and Zhang, 2017; Liu et al., 2012; Vakkari and Hakala, 2000). However, less attention has been paid to the influence of and intertwining of work task processes and search behaviors during an XSS. Users may intertwine actions of (1) working on the main work task, with (2) searching for needed information, (3) applying found information, and (4) iterating these steps. From this perspective, users' information search activities and their information use activities could affect each other as the work task moves forward. This study aims to investigate how information search activities interweave with other non-search task activities for real-world work tasks that involve XSS.

METHOD

We conducted a diary study of users' information behavior for work tasks that involved XSS. We recruited twenty-five participants (9 undergraduate students, 7 graduate students, and 9 non-student staff) from our university. Participants ranged in age from 18 to 62 years and 60% were female. Every participant had a self-selected task that was expected to involve multiple information search sessions. During the study period, each participant used a customized Google account to search online for their selected task. They could search at any time and their search histories were recorded. Participants completed questionnaires before and after each search session and kept a daily review diary about their task activities.

To identify different types of information behaviors and task activities, we extracted users' descriptions of task activities from their responses to four open-ended questions: (1) a question about activities they did before they started the current search session in the *pre-session questionnaire*, (2) a question about their search process in the *post-session questionnaire*, (3) a question about any activities they did besides searching during a search session in the *post-session questionnaire*, (4) a question about their work process for the day in the *daily review diary* (this inquired about search sessions and other task activities). Following Berg's (2017) qualitative analysis workflow, we grouped users' activities into two categories: (1) Information search activities (coded as Search, S), referring to activities when participants interacted with online search engines to look for information related to their tasks; and (2) Non-search activities (coded as Doing, D), referring for any other physical and mental activities participants conducted to complete their tasks *without* using search engines. Through the open-coding process, we further classified the non-search activities into eight different types (explained in next section).

RESULTS & DISCUSSION

Our analysis identified eight types of *non-search task activities* that participants conducted for their tasks, including: (1) setting up/adjusting task scope and goals, for instance, “I’ve had two conversations with my sister about what she wants to do with her baby shower.” (p107); (2) creating task work plans, for instance, “I made a list of people and search terms that I wanted to focus on” (p105); (3) working on the task using prior knowledge, for instance, “[I] wrote a partial visual analysis of [topic] briefly from memory” (p113); (4) applying the found information to address (sub)task problems, for instance, “I have read thoroughly the articles that I found in previous search sessions and drafted three slides about their content” (p131); (5) exchanging ideas with collaborators, for instance “I texted [name] every time I saw an activity that sounded fun.” (p107); (6) consulting others for opinions and inspirations, for instance, “I texted some of my friends to ask if they’d heard of [topic]” (p101); (7) monitoring task progress, for instance, “[I] reviewed my notes to determine how I can complete my task” (p120); and (8) taking notes about the found information and ideas, for instance, “I took detailed notes with this search...I also made notes of questions and other searches I may need to do” (p116). Previous research and models of information behavior have included some of these activities (e.g., formulating goals, consulting others, collaborating, note-taking) (Ellis & Haugan, 1997; Morris et al., 2008; Wilson, 2000). Here, we note that, in general: (a) the purpose of these non-search task actions was not to find specific information required by the task; and that (b) participants separated these actions from their interactions with search engines or other information systems.

To investigate how the information search and non-search activities were intertwined and related to each other, we first abstracted all the non-search activities as *doing*, and then analyzed their relationship with the search activities. We identified three different patterns: SDS, DSDS, and SSD. We note that there are two main differences among these modes: (1) whether the process starts with a *search* activity or a *doing* activity, and (2) whether the search sessions are mixed with working sessions through the task process.

SDS was the most common pattern in our data (n=13). It describes a process when a person starts a work task by searching for information (S) (one session or multiple sessions) and then applies the found information to task (D). They may repeat this process until they have found all the information needed to complete their task. For instance, participant p108 wanted to create a dance show. She described that she started with a search session (S) in which she “searched [for] a lot of music, poems, literature and lyrics that I may be using in my piece.” Then, she engaged in a doing phase (D) where she “drafted [an] idea on [based on her] research.” Later she searched again (S) to find “some new songs” and “book and lyric quotes.” Finally, she (D) “recorded text and lyrics for project use” and “rehearsed my piece from start to finish.”

DSDS was the next most common mode (n=9). It describes a process when a person starts to work on a task without searching. Instead, they begin the task based on their prior knowledge (D). As they move forward, they find that they need to look for more information, so they search online (S). As they find enough information needed for the current stage, they go back to work on the task again (D). After this the process may iterate similar to the SDS mode—they keep switching between searching and working on the task till they achieve the task goals. For instance, participant p102 was helping his friends look for an apartment. Before he started to search online, he (D) “discussed [with my friends] and gave... my initial impressions...we specifically discussed [the apartment area] where I currently live.” Then he started his search (S) to “look up apartment complexes...for more details.” As he found information about different apartments, he texted his friends (D) to discuss “the complexes I’ve found and we narrowed it down to a list that sounded of most interest.”

SSD was a less commonly observed pattern (n=3). It refers to the process when a person collects *all* the needed information across multiple sessions and then uses the found information to complete their task *without* further search. For instance, participant p101 who wanted to write a screenplay conducted three search sessions over 11 days. She searched for information and took notes about the background, local legends, and myths, but did not start to write. For the last search session, she described, “I wanted to do a... last sweep... to make sure I had all the links I needed before I start writing.”

This study identified eight different types of non-search task activities involved in cross-session search and three patterns of intertwining of search and non-search activities. These initial results can help us further understand how users’ information search activities can be affected by other task activities. Conversely, these results also give insight into how non-search activities influence future search activities in cross-session search processes. Furthermore, the results also provide a new dimension for search results customization and recommendations.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. IIS-1552587. We thank our study participants for their time and the UNC Institutional Review Board.

REFERENCES

- Agichtein, E., White, R., Dumais, S., & Bennett, P. (2012). Search, Interrupted: Understanding and Predicting Search Task Continuation. *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '12*, 315–324.
- Lune, H., & Berg, B. L. (2017). *Qualitative research methods for the social sciences*. Pearson.
- Capra, R., Arguello, J., & Zhang, Y. (2017). The Effects of Search Task Determinability on Search Behavior. *ECIR2017*, 108–121.
- Chou, C.-W. J., & Lo, S.-C. S. (2015). Exploring Information Use Behavior in the Context of Knowledge Construction. *Journal of Library and Information Studies*, 132(2), 1–27.
- Ellis, D. and M. Haugan (1997). Modelling the information seeking patterns of engineers and research scientists in an industrial environment. *Journal of Documentation*, 53, 384-403.
- Han, S., He, D., & Chi, Y. (2017). Understanding and modeling behavior patterns in cross-device web search. *Proceedings of the Association for Information Science and Technology*, 54(1), 150–158.
- Mackay, B., & Watters, C. (2009). Understanding and Supporting Multi-session Web Tasks. *Understanding and Supporting Multi-Session Web Tasks*, 45(1), 1–13.
- Morris, D., Morris, M. R., & Venolia, G. (2008). SearchBar: a search-centric web history for task resumption and information re-finding. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI'08* (pp. 1207–1216). Florence, Italy: ACM.
- Kotov, A., Bennett, P. N., White, R. W., Dumais, S. T., & Teevan, J. (2011). Modeling and analysis of cross-session search tasks. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information - SIGIR '11* (p. 10).
- Liu, C., Liu, J., Cole, M., Belkin, N. J., & Zhang, X. (2012). Task difficulty and domain knowledge effects on information search behaviors. In *Proceedings of the ASIST Annual Meeting* (Vol. 49).
- Spink, A., & Cole, C. (2006). Human Information Behavior: Integrating Diverse Approaches and Information Use. *Journal of the American Society for Information Science and Technology*, 57(1), 25–35.
- Wang, H., Song, Y., Chang, M. W., He, X., White, R. W., & Chu, W. (2013). Learning to extract cross-session search tasks. In *WWW 2013 - Proceedings of the 22nd International Conference on World Wide Web* (pp. 1353–1363).
- Wilson, T.D. (2000). *Human information behaviour*. *Informing Science*, 3(2), 49–56.
- Wu, D. (2018). Identifying and Modeling Information Resumption Behaviors in Cross - Device Search. In *SIGIR'18* (pp. 1189–1192). Ann Arbor, MI.: ACM.
- Vakkari, P., & Hakala, N. (2000). Changes in relevance criteria and problem stages in task performance. *Journal of Documentation*, 56(5), 540–562.

A Preliminary Study on App-Switching during the Mobile Search

Liang, Shaobo School of Information Management, Wuhan University, China | liangshaobo@whu.edu.cn
Wei, Ziyi School of Information Management, Wuhan University, China | 2019301040048@whu.edu.cn
Zou, Ning School of Computing and Information, University of Pittsburgh, USA | niz19@pitt.edu

ABSTRACT

Users' mobile app switching behavior has recently received attention, particularly in the context of mobile search. This poster used Amazon MTurk to recruit 86 participants and used structured diaries and interviews to collect data on app switching behavior during the mobile search in daily scenes. We summarized the various behaviors after switching apps during the mobile search process. We also concluded why people switch apps based on information, personal characteristics, and mobile app features.

KEYWORDS

APP switching; User behavior; Information behavior; Mobile app; Mobile users

INTRODUCTION

Smartphones have recently become increasingly important in human communication and interaction (Miranda et al., 2015). A mobile application (app) is a key component of smartphones that may be used for almost any purpose, including entertainment and socializing (Jesdabodi and Maalej, 2015). Böhmer et al. (2011) evaluated mobile app consumption behavior in a large-scale research study and contributed significantly to the subsequent research. Ferreira et al. (2014) focused on app micro-usage, particularly the duration and context of use.

Users' mobile app switching behavior has recently received attention, particularly in the context of mobile search; it refers to the movement from one app to another throughout the mobile search process (Carrascal and Church, 2015). Turner et al. (2019) confirmed that app-switching behavior has invariant traits and continues in an underlying commonality in earlier investigations. Jeong et al. (2020) used quantitative analysis to collect vast log data from users to analyze app switching behavior. Users frequently migrate between mobile apps in different circumstances, according to Roffarello and De Russis (2022). They also presented RecApps (Recommending Apps), an interactive floating widget to support mobile app switching.

However, there is currently a scarcity of studies on app switching in the context of mobile search, particularly employing quantitative and qualitative research approaches. Furthermore, the actions after an app switching may represent the users' more complicated search intentions, but these follow-up behaviors have received insufficient attention. We used Amazon MTurk to recruit 86 participants, using structured diaries and interviews to collect data on app switching behavior during the mobile search in daily settings. During the mobile search process, we want to describe the various behaviors that occur after switching apps. We also looked into why users switch to other apps.

STUDY DESIGN

In this study, we used mixed methods to collect research data. The structured diary collected users' daily mobile search experience, including the app used, search query, search experience, and search context information. Eighty-six users, including 52 females and 34 males, were recruited for this study. Their average age was 23, and each respondent was labeled *Ri*. After the interview, each respondent was paid \$30 to contribute to this study.

We ask these users to provide us with the most impressive app switching experience during the mobile search at least once a day and recall as much data as possible. This study collected 342 app switching behavior during the mobile search through structured diaries. After that, interviews were also used to analyze the reasons behind the characteristics of structured diary data. We used NVivo 11 to analyze the content of the interview text using the content analysis method.

After the interview, the researchers transcribed all the interview records into text and used NVivo 11, qualitative analysis and research software, to analyze the content of the interview text. The three authors of this article were the coders to ensure the coding result was reasonable and adequate.

FINDINGS

Activities after app-switching

During a mobile search, users' app switching behavior is seen in Figure 1. The goal of app switching is to meet needs. Users execute transactional tasks by completing cross-app operations. Some of them are sequential tasks, meaning the actions on the many apps are all directed toward the same goal. Take a selfie using the camera and share it on Instagram, for example (*R13*). Discrete tasks, on the other hand, are a collection of unrelated tasks.

85th Annual Meeting of the Association for Information Science & Technology | Oct. 29 – Nov. 1, 2022 | Pittsburgh, PA. Author(s) retain copyright, but ASIS&T receives an exclusive publication license.

There are two ways if someone selects to continue search after switching an information-seeking task, i.e., search the same thing and refine the search. The latter is mainly through query reformulation (Jansen et al., 2009). Modifying the previous search query could lead to better results (Huang and Efthimiadis, 2009). It can be seen from the interview content that some would reformulate search terms after switching to another app. As *R41* said that he would "change up search terms a little bit," another user (*R64*) was more specific: "Change up my search terms to include or delete certain words." Query reformulation is regarded as a modification of research behavior based on the earlier search results and cognition. Before decision-making, users will compare the results they got from different apps and decide whether to adopt them. Adoption behavior is widely perceived to affect the use of information technology (Grover et al., 1998; Karahanna et al., 1999). Contrary to Non-adoption Behavior, adoption behavior reflects the acceptance of the user, which may be expressed by purchasing (*R1: Buy something*) and downloading behaviors (*R42: Download games.*). Neutral behavior is the third option, neither rejection nor acceptance, like "write a comment (*R20*)".

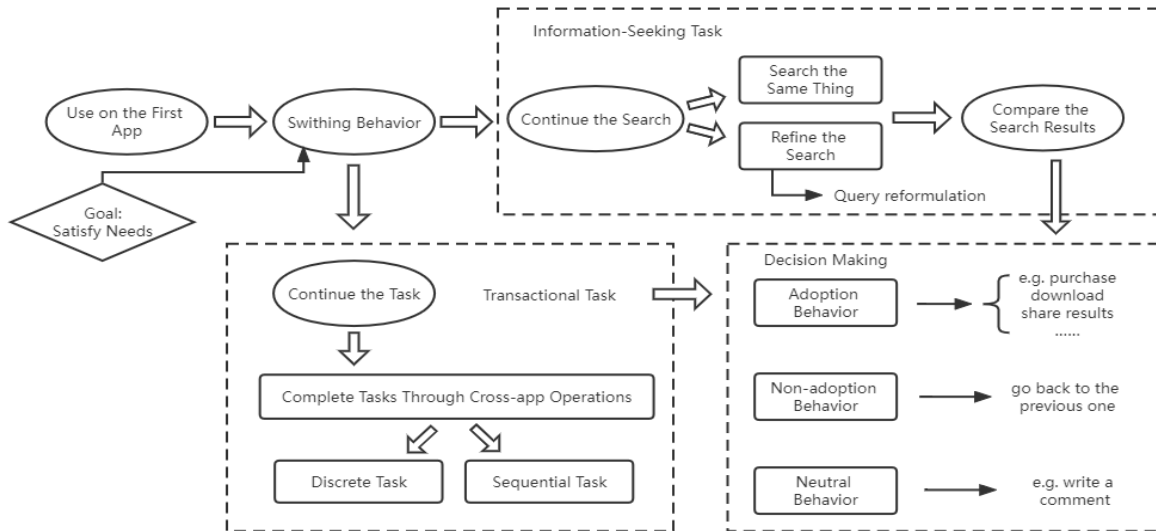


Figure 1. App-switching behavior path during the mobile search

Reasons for app-switching

After seeing that several apps were regularly linked in the process of switching, Roffarello and De Russis (2022) proposed the theory that there are numerous reasons why users consistently switch between apps. Larivière et al. (2013) claimed that mobile devices provide customer value, including information, entertainment, emotion, convenience, monetary value, etc. According to our interviews, these values were considered when users transferred between applications. We seriously discussed and finally categorized them through content analysis. Verbatim analysis of interview content was conducted before conceptualizing the natural language sentences. Twenty-four subcategories were extracted after open coding and then clustered into seven categories in axial coding. Finally, three clear and accurate paradigms were gained through selective coding, as shown in Table 1.

Code	Paradigm	Categories
C1	Information Factor	B1 Information Attributes; B2 Information Quality
C2	Personal Factor	B3 Affective Variables; B4 Objective Condition
C3	App Features	B5 Ease of Use; B6 Low Cost; B7 Trust Factor

Table 1. Coding results about the reasons

CONCLUSION

This poster used Amazon MTurk to recruit 86 participants for this study. We used structured diaries and interviews to collect data on app switching behavior during the mobile search in daily scenes. During the mobile search process, we summarized the various behaviors that occurred after switching apps. We also concluded why people switch apps based on information, personal characteristics, and mobile app features.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (No. 72104187), China Postdoctoral Science Foundation (No: 2021M69248), and the Hubei Province Postdoctoral Innovation Research Position.

REFERENCES

- Böhmer, M., Hecht, B., Schöning, J., Krüger, A., & Bauer, G. (2011, August). Falling asleep with Angry Birds, Facebook and Kindle: a large scale study on mobile application usage. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services* (pp. 47-56).
- Carrascal, J. P., & Church, K. (2015, April). An in-situ study of mobile app & mobile search interactions. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 2739-2748).
- Ferreira, D., Goncalves, J., Kostakos, V., Barkhuus, L., & Dey, A. K. (2014, September). Contextual experience sampling of mobile application micro-usage. In *Proceedings of the 16th International Conference on Human Computer Interaction with Mobile Devices and Services* (pp. 91-100).
- Grover, V., Teng, J., Segars, A. H., & Fiedler, K. (1998). The influence of information technology diffusion and business process change on perceived productivity: The IS executive's perspective. *Information & Management*, 34(3), 141-159.
- Huang, J., & Efthimiadis, E. N. (2009, November). Analyzing and evaluating query reformulation strategies in web search logs. In *Proceedings of the 18th ACM conference on Information and knowledge management* (pp. 77-86).
- Jansen, B. J., Booth, D. L., & Spink, A. (2009). Patterns of query reformulation during web searching. *Journal of American Society for Information Science and Technology*, 60(7), 1358-1371.
- Jeong, Y., Jung, H., & Lee, J. (2020). Cyberslacking or smart work: Smartphone usage log-analysis focused on app-switching behavior in work and leisure conditions. *International Journal of Human-Computer Interaction*, 36(1), 15-30.
- Jesdabodi, C., & Maalej, W. (2015, September). Understanding usage states on mobile devices. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 1221-1225).
- Karahanna, E., Straub, D. W., & Chervany, N. L. (1999). Information technology adoption across time: a cross-sectional comparison of pre-adoption and post-adoption beliefs. *MIS quarterly*, 183-213.
- Miranda, J., Mäkitalo, N., Garcia-Alonso, J., Berrocal, J., Mikkonen, T., Canal, C., & Murillo, J. M. (2015). From the Internet of Things to the Internet of People. *IEEE Internet Computing*, 19(2), 40-47.
- Roffarello, A. M., & De Russis, L. (2022). Understanding and Streamlining App Switching Experiences in Mobile Interaction. *International Journal of Human-Computer Studies*, 158, 102735.
- Turner, L. D., Whitaker, R. M., Allen, S. M., Linden, D. E., Tu, K., Li, J., & Towsley, D. (2019). Evidence to support common application switching behaviour on smartphones. *Royal Society Open Science*, 6(3), 190018.

Nudging Learning Behaviour: A Systematic Review

Lim, Kok Khiong

Nanyang Technological University, Singapore | w200004@e.ntu.edu.sg

Lee, Chei Sian

Nanyang Technological University, Singapore | leecs@ntu.edu.sg

ABSTRACT

This review examines nudge intervention in influencing learning behaviours from the context of the information use environment. It was conducted based on PRISMA guidelines. Six major databases were searched, and ten studies were included in this review after screening. Overall, the included studies reported beneficial outcomes in the academic settings, but results were mixed as nudge interventions were context-dependent. To convey course-related information during the nudge intervention, all the included studies used at least one technique from category A-Decision Information and most used a secondary technique to enhance or manipulate the intervention. Among the techniques, A1-Translate Information, A2-Make Information Visible, and C1-Provide Reminders were widely used. However, there was no ideal nudge technique or a combination of nudge techniques that were effective across all situations. Nonetheless, this review underscored the importance of incorporating nudge techniques in an informational learning environment that could shape and motivate learners.

KEYWORDS

Nudge, intervention, information use environment, systematic literature review, learning behaviour

INTRODUCTION

Using the information to encourage individuals to change their behaviour in ways that are beneficial to themselves or society without forbidding their freedom of choice has gained significant attention in recent years (Münscher et al., 2016; Thaler & Sunstein, 2008). This approach, termed ‘nudge’, is based on the basic premise that people’s daily decision-making is inherently autonomous or follows perceived norms (Thaler & Sunstein, 2008). Hence, it is not rational due to their own cognitive biases and heuristics (Thaler & Sunstein, 2008). Nudge exploits these cognitive biases and heuristics to influence people’s decision-making towards an intended outcome (Thaler & Sunstein, 2008). In other words, nudge uses the information to design interventions that could subtly change people’s behaviour. Past reviews have investigated nudge intervention in various domains. For example, Sarpy et al. (2021) reviewed the construction safety domain and highlighted nudge techniques were used to enhance how information was presented, how options and decision formats were arranged, and how people followed through with their intentions. Notably, there were indicators in Sarpy et al.’s (2021) findings that aligned with Taylor’s (1991) notion of information use environment (IUE). Specifically, IUE influences workers’ safety and health decisions, and the environment in which the information is operating further influences workers’ decision-making. IUE, a set comprises “*people*, structure and thrust of *problem* of those set of people, typical *settings*, and what constitutes *resolution* of problems” (p. 221), affects the flow and use of information that determines how the value of information will be judged (Taylor, 1991). However, there is limited review to examine the state of nudge research in the educational domain, particularly in learning behaviour. Guided by the IUE, this review will extend current nudge research in learning behaviour by examining the characteristics of past nudge studies (*people*, *problem*, *setting*), usage and nudge techniques (*resolution*), and their extent of effectiveness with considerations of the antecedent factors.

RESEARCH METHOD

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA; Page et al., 2021) guidelines and guided by the PICO-C (*Population, Intervention, Comparison, Outcomes*, and *Context*) framework to search and select the queried studies for inclusion into this review. The Mixed-Methods Appraisal Tool (MMAT; Hong et al., 2019) was used to assess the quality of the studies and their nudge techniques were categorised according to the taxonomy by Münscher et al. (2016). Six major databases, ERIC, Education Source, PsycINFO, ScienceDirect, Scopus, and Web of Science, were searched for studies published between January 2008 and November 2021 that aligned with the nudge concept developed in 2008. A keyword search strategy (related to nudg*, behavioural economics, intervent*, choice architect*, learn*, educat*, and self-regulat*) was developed to ensure the queries were relevant, optimised, and inclusive of British and American spellings. The queried studies were screened in two stages. First, the studies’ titles, abstracts, and associated metadata, such as types of journals and languages used, were retrieved and reviewed based on the inclusion (published in English language and peer-reviewed journal) and exclusion (grey literature) criteria. In the second stage, the full text was retrieved for further evaluation based on the PICO-C framework before being selected for inclusion in this review. Data were then extracted from the included studies using the ten pre-defined categories for comparison and synthesis. The first author performed the data extraction, and the second author verified the accuracy of the extracted data. Any discrepancies were resolved through discussion and consensus.

RESULTS

A total of 481 records were retrieved from the six databases. Of which, ten studies were selected in this review after the screening. The ten studies covered diverse areas (*setting and problem*) that included evaluating the effects of self-directed scheduling (Baker et al., 2016), improving cooperative learning (Buchs et al., 2016), enhancing learning experiences and outcomes (Dart & Spratt, 2020), reducing procrastination in a competitive learning environment (Li et al., 2021), informing on imminent submission deadlines (Motz et al., 2021), persuading students to improve their study consistency (O'Connell & Lang, 2018), evaluating the impact on procrastination (Onji & Kikuchi, 2011), improving the enrolment rate of tutoring services (Pugatch & Wilson, 2018), evaluating the impact of learning outcome (Sinha et al., 2021), and improving task performance (Smith et al., 2018). These ten included studies were published between 2011 to 2021 across six different countries. The sample size ranged from 187 to 18,043. Participants were recruited from colleges and universities (*people*). The quality of the studies varied. Five studies fulfilled all the MMAT criteria (Dart & Spratt, 2020; Li et al., 2021; O'Connell & Lang, 2018; Pugatch & Wilson, 2018; Smith et al., 2018). Another four were unable to fulfil the randomisation criteria (MMAT 2.1) in the quantitative RCT category as the studies did not specify how the randomisation schedule were generated (Baker et al., 2016; Buchs et al., 2016; Motz et al., 2021; Sinha et al., 2021), and one did not account for the confounders (MMAT 3.4) in the quantitative non-randomised category (Onji & Kikuchi, 2011).

Based on the Münscher et al. (2016) framework (*resolution*), A1-Translate Information ($n = 7$), A2-Make Information Visible ($n = 7$), and C1-Provide Reminders ($n = 7$) were the most popular nudge techniques. Other techniques identified were, A3-Provide Social Reference Point ($n = 3$), B1-Change Choice Defaults ($n = 1$), B2-Change Option-related Effort ($n = 1$), and C2-Facilitate Commitment ($n = 1$). Most studies combined several techniques in their nudge intervention design. Two studies used four types of techniques, five studies utilised three techniques, and one combined two types of nudge techniques in their nudge designs. Only two studies used the single nudge technique. Nine studies reported positive results in improving students' academic performances, learning outcomes and experiences (Buchs et al., 2016; Dart & Spratt, 2020; Li et al., 2021; Motz et al., 2021; O'Connell & Lang, 2018; Onji & Kikuchi, 2011; Pugatch & Wilson, 2018; Sinha et al., 2021; Smith et al., 2018). However, students' characteristics such as gender and contextual variables in the area of prior academic performances and behaviour (Li et al., 2021), the timing of interventions (O'Connell & Lang, 2018; Onji & Kikuchi, 2011), and the presented information exposed to students (Motz et al., 2021), were factors affecting the degree of nudge effectiveness. Meanwhile, the tenth study reported no effect on near-term engagement and a weak significant negative impact on long-term engagement, persistence, and performance (Baker et al., 2016).

DISCUSSION AND CONCLUSION

This review presents preliminary findings on nudge research to ascertain the effective use of information to nudge learning behaviour from the IUE perspective. Overall, the included studies showed positive nudge effectiveness but mixed results. This was mainly due to confounding effects arising from nudge dependency on the intervention's context that affects its outcomes. Hence, future nudge research needs to account for any potential factors during nudge design to mitigate and control the confounding effects. Meanwhile, all the included studies used at least one technique from category A-Decision Information. Most would use at least two techniques within or across different nudge techniques' categories in their intervention design. The secondary technique was either used to reinforce or further manipulate the intervention. These findings were not surprising as course-related information was required to trigger learners' System 2 decision-making process in deliberating and reasoning on their next course of actions that entailed consequences on the learning goals. Key to this process is the role of information use underpinning decision-making and learning behaviour. However, nudge effects in the longer term currently remains a limitation. Future research is suggested to focus on examining the interplay between IUE and nudge to overcome the limitation.

The low numbers of included studies were not unexpected, given that nudge is a relatively new field in education (Weijers et al., 2021). This limitation is well acknowledged by organisations such as Joanna Briggs Institute and Cochrane (Slyer, 2016; Yaffe et al., 2012). Nonetheless, the diversity of countries and study designs provided converging evidence on nudge's applicability worldwide. The four studies that lacked the proper randomisation process might raise concerns on the results' validity due to potential biases and random errors during the experimentation (Kendall, 2003). This is especially crucial in the educational context. Given that there is an added responsibility to ensure that nudge interventions applied should not disrupt educational learning nor have negative spillover effects to other essential areas (Weijers et al., 2021). Hence, attention should be given during the study design to achieve higher quality research. In sum, this review examined exiting nudging research landscape from the IUE perspective. IUE influences learners' learning decisions, and the academic environment would further influence their decision-making. Further, nudge could potentially extend to other information domains, such as information seeking to guide information behaviour. More importantly, the review's findings aligned with prior works (e.g., Ma & Lee, 2020) that advocates developing an informational learning environment that motivates learners. Specifically, this review underscores the importance of incorporating nudge in the informational learning environment.

REFERENCES

- Baker, R., Evans, B., & Dee, T. (2016). A randomized experiment testing the efficacy of a scheduling nudge in a Massive Open Online Course (MOOC). *AERA Open*, 2(4). <https://doi.org/10.1177/2332858416674007>
- Buchs, C., Gilles, I., Antonietti, J. P., & Butera, F. (2016). Why students need to be prepared to cooperate: A cooperative nudge in statistics learning at university. *Educational Psychology*, 36(5), 956-974. <https://doi.org/10.1080/01443410.2015.1075963>
- Dart, S., & Spratt, B. (2020). Personalised emails in first-year mathematics: Exploring a scalable strategy for improving student experiences and outcomes [Article]. *Student Success*, 11(2), 1-12. <https://doi.org/10.5204/ssj.1543>
- Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., & Nicolau, B. (2019). Improving the content validity of the mixed methods appraisal tool: a modified e-Delphi study. *Journal of clinical epidemiology*, 111, 49-59. e41. <https://doi.org/https://doi.org/10.1016/j.jclinepi.2019.03.008>
- Kendall, J. M. (2003). Designing a research project: randomised controlled trials and their principles. *Emergency Medicine Journal*, 20(2), 164. <https://doi.org/10.1136/emj.20.2.164>
- Li, Z. X., Wang, G., & Wang, H. J. (2021). Peer effects in competitive environments: Field experiments on information provision and interventions. *MIS Quarterly*, 45(1), 163-191. <https://doi.org/10.25300/misq/2021/16085>
- Ma, L., & Lee, C. S. (2020, November 30 – December 1). A motivational design approach to integrate MOOCs in traditional classrooms. Proceedings of the 22nd Asia-Pacific Conference on Digital Library ICADL, Kyoto, Japan. https://doi.org/10.1007/978-3-030-64452-9_16
- Motz, B. A., Mallon, M. G., & Quick, J. D. (2021). Automated educative nudges to reduce missed assignments in college [Article]. *IEEE Transactions on Learning Technologies*, 14(2), 189-200, Article 9373909. <https://doi.org/10.1109/TLT.2021.3064613>
- Münscher, R., Vetter, M., & Scheuerle, T. (2016). A review and taxonomy of choice architecture techniques. *Journal of Behavioral Decision Making*, 29(5), 511-524. <https://doi.org/10.1002/bdm.1897>
- O'Connell, S. D., & Lang, G. (2018). Can personalized nudges improve learning in hybrid classes? Experimental evidence from an introductory undergraduate course. *Journal of Research on Technology in Education*, 50(2), 105-119. <https://doi.org/10.1080/15391523.2017.1408438>
- Onji, K., & Kikuchi, R. (2011). Procrastination, prompts, and preferences: Evidence from daily records of self-directed learning activities. *The Journal of Socio-Economics*, 40(6), 929-941. <https://doi.org/10.1016/j.socec.2011.08.019>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews*, 10(1), 89. <https://doi.org/10.1186/s13643-021-01626-4>
- Pugatch, T., & Wilson, N. (2018). Nudging study habits: A field experiment on peer tutoring in higher education. *Economics of Education Review*, 62, 151-161. <https://doi.org/10.1016/j.econedurev.2017.11.003>
- Sarpy, S. A., Betit, E., Barlet, G., & Echt, A. (2021). *A literature review of behavioral economics in the construction industry: Use of choice architecture techniques to accelerate acceptance and adoption of safety and health research findings and solutions*. <https://www.cpwr.com/wp-content/uploads/Behavioral-Economics-Literature-Review.pdf>
- Sinha, T., Kapur, M., West, R., Catasta, M., Hauswirth, M., & Trninic, D. (2021). Differential benefits of explicit failure-driven and success-driven scaffolding in problem-solving prior to instruction. *Journal of Educational Psychology*, 113(3), 530-555. <https://doi.org/10.1037/edu0000483>
- Slyer, J. T. (2016). Unanswered questions: Implications of an empty review. *JBI Evidence Synthesis*, 14(6). <https://doi.org/10.11124/JBISRIR-2016-002934>
- Smith, B. O., White, D. R., Kuzyk, P. C., & Tierney, J. E. (2018). Improved grade outcomes with an e-mailed “grade nudge”. *The Journal of Economic Education*, 49(1), 1-7. <https://doi.org/10.1080/00220485.2017.1397570>
- Taylor, R. S. (1991). Information use environments. In B. Dervin & M. Voigt (Eds.), *Progress in Communication Science* (pp. 173-216). Norwich, NJ: Ablex.
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.
- Weijers, R. J., de Koning, B. B., & Paas, F. (2021). Nudging in education: From theory towards guidelines for successful implementation. *European Journal of Psychology of Education*, 36(3), 883-902. <https://doi.org/10.1007/s10212-020-00495-0>
- Yaffe, J., Montgomery, P., Hopewell, S., & Shepard, L. D. (2012). Empty reviews: A description and consideration of cochrane systematic reviews with no included studies. *PLoS One*. <https://doi.org/10.1371/journal.pone.0036626>

Redeeming by Unlearning: A Critical Discourse Analysis of COVID-19 Vaccine Hesitancy on Facebook

Ma, Jinxuan
Bonnici, Laurie

School of Library and Information Management, Emporia State University, USA | jma4@emporia.edu
College of Communication & Information Sciences, University of Alabama, USA | lbonnici@ua.edu

ABSTRACT

As false and debunked claims about the COVID-19 virus and vaccines were pervasively disseminated across online social networks, their detrimental effects necessitated that social media companies track down and remove the disinformation. To explore the lingering vaccine hesitancy and resistance within a vaccine discussion group on Facebook, this study applies a nascent framework called Information Acts in three communication styles. Employing critical discourse analysis methodology, this study showed that the dominant communication styles of vaccine hesitant participants (VHPs) are locutionary and perlocutionary acts. Those VHPs can reduce and eventually end their vaccine hesitancy and resistance in the process of informative and rational communication and interactive and ad-hoc support for informed decision-making. The results pose challenges and opportunities for public health communication and information provision to cultivate information resilience and interventions targeting VHPs.

KEYWORDS

COVID-19 Pandemic, Disinfodemic, Vaccine Hesitancy, Information Acts, Critical Discourse Analysis

INTRODUCTION

The term disinfodemic, coined by UNESCO (Posetti & Bontcheva, 2020), explicates the dire information crisis that pervaded and subsequently undermined the global efforts to disseminate scientific information and boost public acceptance of COVID-19 restrictions and vaccination. As online social networks afforded immediate transmission of disinformation, social media companies including Facebook initiated fact-checking campaigns in January 2020. These initiatives automatically detected disinformation and subsequently removed over 20 million pieces of content that violated Facebook's rules against spreading disinformation about COVID and vaccination. This study is designed to further explore the manifestations of vaccine hesitancy and resistance among vaccine hesitant participants (VHP) in the new COVID era information environment within a private vaccine discussion group on Facebook. It poses a research question—how VHPs communication style and information-seeking behavior illuminates their actions around vaccination.

THEORETICAL FRAMEWORK

Information Acts are conceived through the lens of Austin's Speech Acts (1962; 1975). In brief, speech utterances embody three forms of intent; locutionary, illocutionary, and perlocutionary. Speakers either make utterances to offer information with no expectation of response (locutionary), to elicit further information to build understanding (illocutionary), and to influence a shift in belief (perlocutionary) (Bonnici & Ma, 2021). Information Acts, which convey information coupled with intent by the sender, serve to influence or make a difference. Receipt of the information packaged in intent may catalyze change in thought, spur action, or result in that "aha moment" where realization brings clarity of thinking and perhaps a resultant change in behavior (Bonnici & Ma, in progress). Three different might declare the importance of receiving the COVID vaccines. But the difference in the message is in how it is delivered: the intent.

RESEARCH METHOD

In analyzing the attributes of communication and information seeking behaviors within this group, this study adopted a critical discourse analysis (CDA) approach to explore how vaccine hesitant people's communication style and information-seeking behavior reveals their actions toward vaccination. As an influential methodology to decode the structural relationships between power and inequity in language, CDA has been efficiently used in uncovering the interplay between linguistic-communicative patterns and concrete actions of a group through their verbal or textual discourses (Blommaert & Bulcaen, 2000; Fairclough & Mauraen, 1997). Framed by the Information Acts framework, this study adopted Fairclough's (1995) systematic analytic method with a three-layer model of critical discourse analysis: description (textual analysis), interpretation (process of discursive practice), and explanation (analysis of social practices) (Abbas, 2022; Jaber, 2021; Ma & Stahl, 2017). After performing a comprehensive search about vaccine hesitancy, a total of 4331 posts (including 26 original posts and 4305 comments) during the periods between July 27, 2020 to May 11, 2022 were manually captured. A codebook based on Information Acts was created in a pilot study of 309 posts (another dataset of this group's postings) and validated through reaching a high intercoder reliability (86%) between two LIS researchers. Applying CDA, a total of 279 collected posts (including 26 original posts and top 10 comments for each post) were analyzed thematically by types of VHP's communication styles and information seeking behavior.

FINDINGS

The results of data analysis reveal that locutionary acts dominated the discourse with 155 (56%) posts compared to perlocutionary acts with 101 (36%) and illocutionary acts with 23 (8%). Different themes in their communication style and information seeking behavior emerged around VHPs' stances:

1. Locutionary: (a) quote mining: one example is a blunt decontextualized claim—"Bill Gates recently made statements concerning omicrons greater effectiveness over vaccines." (b) post hoc, ergo propter hoc, refers to "one of the most powerful fallacies of human logic...Especially when it comes to vaccines." "That's taken out of context." (c) racist infiltrator—"The high-risk genetic signal is carried in sixty per cent of people with south Asian heritage... The gene is present in 15 per cent of those with European ancestry, the study revealed, but is only found in 2 per cent of those with African-Caribbean ancestry."
2. Perlocutionary: (a) ambivalent attitude/feelings to vaccination—"Can vaccines cause ADHD?" (b) unlearning by active communication and information seeking—"I admire people who ask questions and are willing to look at different views and make their own decision, you are attacking a legitimate article that I actually found very helpful with all the links for my own research so I can contrast them to a pro vaccine stance gives me more food for thought instead of just being bullied/manipulated into a certain position."
3. Illocutionary: (a) call for global public health or government data on vaccine side effects—"Has anyone read the first batch of Pfizer data that was revealed a few days ago? ...I haven't read the data and if anyone can link it here would be appreciated. If you've read it what are your thoughts?" (b) call for personal claims or anecdotal experience, or interpretation about vaccine injury—"If you're unvaccinated for Covid why did you make that choice? Is there anything that would change your mind? And has that decision had any impact on your life in regards to work, freedom of travel, relationships with family etc?" (c) seeking affirmation and validation—"Is there any data on giving one COVID vaccine shot for ages 5-11?"

CONCLUSION

Findings indicate that civility, empathy, and genuineness cultivate informative and rational communication within the Facebook group where VHPs are more likely to open up about their concerns, fears, and ignorance (e.g., take second-hand knowledge at face value) when making an informed decision around vaccines. Their indecisiveness is often influenced by their personal beliefs, close-knit relationships, exposure to anti-vaccination rhetoric, emotional distress, fears, and concerns about vaccine safety and efficacy. Among those VHPs, open communication and unlearning could become turning points of breaking down their beliefs, thoughts, feelings, and behaviors towards vaccine hesitancy and resistance. The results pose challenges and opportunities for public health communication and information provision to cultivate information resilience and interventions targeting VHPs.

The implications extend to the development and delivery of public health data/information into different types of conversational messages in layman's terms that could reduce emotional distress, such as embarrassment, vulnerability, uncertainty, distrust, fears, or angst. Furthermore, with renewed urgency to better communicate and inform, it is essential to create and deliver tailored health information based on the differentiation between the misinformed and the uninformed, between their attitudes/emotions and behaviors. VHPs shared that they went through a long journey to get vaccinated by actively communicating, unlearning, or deconstructing disinformation, especially measuring the risks of vaccine side effects alongside the severity of COVID infection.

REFERENCES

- Abbas, A. H. (2022). Politicizing COVID-19 vaccines in the press: A critical discourse analysis. *Int J Semiot Law*, 35(3), 1167–1185. <https://doi.org/10.1007/s11196-021-09857-3>
- Austin, J. L. (1962; 1975), *How to do things with words*. The William James Lectures delivered at Harvard University in 1955. Oxford University Press.
- Blommaert, J., & Bulcaen, C. (2000). Critical discourse analysis. *Annual Review of Anthropology*, 29, 447–466. <http://www.jstor.org/stable/223428>
- Bonnici, L., & Ma, J. (2021). What are they saying? A speech act analysis of a vaccination information debate on Facebook. *Canadian Journal of Information and Library Science*, 44(1), e1–19. <https://doi.org/10.5206/cjilsrscib.v44i1.13342>
- Bonnici, L., & Ma, J. (In progress). Acts & states of being: Expanding characterizations of information. *Journal of Documentation*.
- Fairclough, N. & Mauranen, A. (1997). The conversationalization of political discourse: A Comparative view. *Belgian Journal of Linguistics*, 11 (1), 89–119. <https://doi.org/10.1075/bjl.11.06fai>
- Jaber, B. (2021). A critical discourse analysis of the hashtag #WeAreAllIsraaGhrayeb on Twitter: An exploration of honor killing crimes in Palestine. *Proceedings of the Association for Information Science & Technology*, 58(1), 740-743. <https://doi.org/10.1002/pr2.547>
- Ma, J., & Stahl, L. (2017). A multimodal critical discourse analysis of anti-vaccination information on Facebook. *Library & Information Science Research*, 39(4), 303–310. <https://doi.org/10.1016/j.lisr.2017.11.005>

Posetti, J. & Bontcheva, K. (2020). *Disinfodemic: Deciphering COVID-19 disinformation*. Retrieved from https://en.unesco.org/sites/default/files/disinfodemic_deciphering_covid19_disinformation.pdf

Revisiting Connotations of Digital Humanists: Exploratory Interviews

Ma, Rongqian

Indiana University Bloomington, USA | rm56@iu.edu

ABSTRACT

This ongoing study revisits the connotations of “digital humanists” and explores the reasons why a researcher does or does not self-identify as a digital humanist. Building on semi-structured interview data collected from fourteen researchers and practitioners engaging in digital humanities (DH) projects, this poster illustrates researchers’ various understandings of “digital humanist” as a term and research identity and highlights the complexity of “digital humanists” as a research community. This study contributes to DH scholarship with insights into the collective imaginations of the “digital humanist” as a research community one decade after the early attempts. Findings of this research study also facilitate a more thorough, timely, and dynamic discussion of the major workforce in digital humanities, potentially paving the way for future research on labor and collaboration in the DH research domain.

KEYWORDS

Digital humanist; research community; semi-structured interview

INTRODUCTION AND LITERATURE REVIEW

This poster revisits a classic question in the digital humanities domain: What are the connotations of “digital humanists?” Early research literature has proposed two major definitions. Alvarado (2012) defined a digital humanist as someone who (1) aims to develop the deep domain knowledge of the traditional humanist, (2) learns a wide variety of technologies and programming languages, and (3) critically situates the technologies as cultural artifacts “participating in the production of social and cognitive structures.” This definition of the “digital humanist” requires a scholar to be proficient in both technical skills and humanities knowledge. By contrast, (Ramsay, 2011) argued that a scholar can be called a digital humanist as long as they can build something with digital methods (e.g., applying existing tools or modifying existing codes). This definition embraced a relatively broader reading of the “digital,” emphasizing the gradual transition of a humanities scholar into a digital humanist. However, early definitions still consider a digital humanist as primarily a *humanist* with strong humanities training. Such definitions captured different perceptions of “digital humanists” as a research community in the early stage.

The recent technological advancement has expanded the reach of digital humanist as a research community and raised new questions about the connotations of the term. *How do DH researchers think of the role of technological literacy in their work and research identify? Is a digital humanist still primarily a humanist researcher, or does a digital humanist need formal humanities training?* Jänicke (2016) called for a broader, more inclusive conception of the digital humanists community, arguing that the DH field would benefit from the active participation of researchers from various knowledge domains, especially those from scientific fields. Recent research in the landscape of DH illustrated that “DH is simultaneously a discipline in its own right and a highly interdisciplinary field, with many connecting factors to neighboring disciplines—first and foremost, computational linguistics, and information science” (Luhmann & Burghardt, 2022). In addition to the roles of DH among academic fields, Ma and Li (2022) also demonstrated that humanities and STEM researchers, among others, further shape the DH community with their respective, sometimes competing, research conventions as well as cross-field research collaborations. Empirical works focusing on DH intellectual structures, disciplinary compositions, and collaboration also suggest an increasingly complex picture of DH workforce (Antonijević, 2015; Fiormonte et al., 2015; Griffin & Hayler, 2018; Papadopoulos & Reilly, 2020).

The increasing “complexity” of DH workforce has raised the urgency to revisit the classic question in DH: “Who is in and who is out?” (Ramsay, 2011). Using semi-structured interview method, this study explores the current understandings and conceptions of a “digital humanist,” illustrating why and why not a researcher self-identifies as a digital humanist. Scholarly literature has also shown that active DH researchers do not always think of themselves as “digital humanists” (Burdick, 2012). Rich qualitative data collected from the interviews will help highlight the major perceptions of what a digital humanist means. Ten years into the discussion of the research identities of digital humanists, the findings of this exploratory study will also inform future research on DH workforce, teams, and cross-field collaboration.

DATA AND METHODS

Participant

Fourteen participants were identified through a snowball sampling technique based on the author’s personal network (Biernacki & Waldorf, 1981). Any researcher or practitioner who has been actively engaging in DH projects can be qualified as interview candidates. Using this inclusive screening criterium, this study recruited fourteen participants

85th Annual Meeting of the Association for Information Science & Technology | Oct. 29 – Nov. 1, 2022 | Pittsburgh, PA. Author(s) retain copyright, but ASIS&T receives an exclusive publication license.

representing a diverse pool in terms of disciplines, academic levels, and seniority. The represented fields and disciplines include, for example, history, religious studies, English, art history, sociology, information science, math, and anthropology. The participants also represent various levels of training and research experience, including master's students (n=2), PhD candidates (n=5), postdoc (n=1), assistant professors (n=4), associate professors (n=1), and curator (n=1). Among the fourteen interviewees, eight participants (P2, P4, P6, P7, P9, P10, P11, P14) self-identified as digital humanists while six (P1, P3, P5, P8, P12, P13) did not.

Interview Design and Analysis

Semi-structured interview method was used for this study. Semi-structured interviews have the unique advantage of preserving participants' rich descriptions and detailed responses, especially the original discussions of "digital humanists" and the reasons for their self-identification as digital humanists. Each interview lasted for 45 minutes and consisted of two sections. In the first section, I collected structured demographic information from the participants, such as the participant's current academic position, major field(s) of study, and if they would self-identify as a digital humanist. The second section discussed the participant's self-identification, focusing on the reasons for their choices and how they understand the meaning of "digital humanists." The interviews were audio-recorded and fully transcribed for analysis. Qualitative content analysis was then used to code and analyze the participants' responses, particularly their accounts on the identities of digital humanists.

RESULTS

Preliminary results of this study captured various understandings of "digital humanists" among researchers.

Technological skills define a digital humanist. Results show that some researchers (P2, P6, P9) evaluate their eligibility of being a "digital humanist" based on their technical skills, especially the ability to write codes or run computational models. They claimed that it is the programming ability that made them a "digital humanist" rather than a "humanist." Some participants, however, argued that **actual work experiences define a digital humanist.** P4, for example, said "I self-identify as a digital humanist in the same way as I self-identify as an information designer, a programmer, and a statistician. I do research and work in digital humanities, so I guess *technically* I can say I am a digital humanist." This account treats "digital humanist" as a non-mutually exclusive, fluctuating status that reflects current work state and content. P10 also considered herself as a digital humanist because she had supervised students in DH projects. Compared with the former group who understand digital humanists based on training and skillsets, scholars and practitioners in the latter group believed one can be a digital humanist "by doing it," emphasizing the actual practices over the conceptual, artificial definitions of the term.

The preliminary results also suggest that some researchers did *not* self-identify as digital humanists because they are concerned with the **nature of their research questions and the necessity of using digital methods in research.**

For instance, P1 and P5 did *not* self-identify as digital humanists because they consider their research questions "not ultimately bond with digital methods." A museum art curator trained in art history, museum studies, and sociology, P3 was concerned that new constructs such as "DH" did not generate fundamentally new questions distinct from those in conventional or analog humanities, which made him reluctant to self-identify as a "digital humanist," rather than just a humanist. From a different perspective, P12 and P13 did not self-identify as digital humanists because they would not think of themselves as "humanists," who, according to their accounts, must have humanities training and work on humanities research questions in an in-depth manner. Both trained and working as information scientists, P12 and P13 demonstrated that they have been engaging in DH projects from a technological perspective, either by means of working as a programmer (P12) in the team or by solving problems related to digital methodologies (e.g., how to retrieve and search music information, or create digital simulations for cultural heritage sites, P13).

CONCLUDING REMARKS

The preliminary findings of this study suggest that while technological literacy is still an important factor that impacts researchers' self-identification as a digital humanist, researchers have also developed a more inclusive understanding of the term, which focuses more on the actual DH work engagement rather than the "qualifications" (e.g., if they can code or if they have humanities degrees). These researchers acknowledge that DH work can be of different shapes and each researcher can contribute in a unique and meaningful way – and hence, call themselves as digital humanists. Such a broader view of digital humanists as a research community has potential benefits. It may further facilitate researchers to embrace research methods and paradigms outside of their home fields, as well as to welcome new collaborators, who, for example, speak different disciplinary languages or have distinct ways of thinking. Despite its values, this study bears limitations. Fourteen interviews are not sufficient to conclude all considerations that go into the current definitions of, and researchers' self-identification as, digital humanists. In future work, to address the limitation, more participants will be recruited and the survey method will be applied to expand the sample size. Analysis of a larger dataset with richer details will help develop a more thorough idea of the current conceptions of "digital humanists" among researchers engaging in DH projects.

ACKNOWLEDGMENT

The author would like to thank all the interview participants for their generosity in sharing thoughts, experiences, and ideas. This study would not be possible without their contribution.

REFERENCES

- Alvarado, R. C. (2012). The digital humanities situation. In M. Gold (Ed.), *Debates in the Digital Humanities*. University of Minnesota Press.
- Antonijević, S. (2015). *Amongst digital humanists: An ethnographic study of digital knowledge production*. Palgrave Macmillan.
- Biernacki, P., & Waldorf, D. (1981). Snowball sampling: Problems and techniques of chain referral sampling. *Sociological Methods & Research*, 10(2), 141–163.
- Burdick, A. (Ed.). (2012). *Digital humanities*. MIT Press.
- Fiorimonte, D., Numerico, T., & Tomasi, F. (2015). *The Digital Humanist: A Critical inquiry*. Punctum Books.
<https://doi.org/10.21983/P3.0120.1.00>
- Griffin, G., & Hayler, M. S. (2018). Collaboration in Digital Humanities Research – Persisting Silences. *Digital Humanities Quarterly*, 012(1).
- Jänicke, S. (2016). Valuable Research for Visualization and Digital Humanities: A Balancing Act. *Workshop on Visualization for the Digital Humanities, IEEE VIS 2016, Baltimore, Maryland, USA*.
- Luhmann, J., & Burghardt, M. (2022). Digital humanities—A discipline in its own right? An analysis of the role and position of digital humanities in the academic landscape. *Journal of the Association for Information Science and Technology*, 73(2), 148–171. <https://doi.org/10.1002/asi.24533>
- Ma, R., & Li, K. (2022). Digital humanities as a cross-disciplinary battleground: An examination of inscriptions in journal publications. *Journal of the Association for Information Science and Technology*, 73(2), 172–187.
<https://doi.org/10.1002/asi.24534>
- Papadopoulos, C., & Reilly, P. (2020). The digital humanist: Contested status within contesting futures. *Digital Scholarship in the Humanities*, 35(1), 127–145. <https://doi.org/10.1093/lc/fqy080>
- Ramsay, S. (2011). Who's in and who's out. In M. Terras, J. Nyhan, & E. Vanhoutte (Eds.), *Defining digital humanities: A reader*. Ashgate.

Civic Data Education Series: Developing Civic Data Literacies to Serve Communities

Mattern, Eleanor

University of Pittsburgh, USA | emm225@pitt.edu

Rapchak, Marcia

University of Pittsburgh, USA | mrapchak@pitt.edu

ABSTRACT

This poster introduces the Civic Data Education Series, a collection of openly licensed modules and instructional materials that support the development of learners' civic data literacy. We designed this series for the learner communities of Library and Information Science (LIS) graduate students and library staff, with the goal of equipping library workers to assume civic data intermediary roles. In this poster, we present the meaning of "civic data literacy," our project's aims and instructional design approach, and resulting materials.

KEYWORDS

civic data, instructional design, LIS education, data literacy

INTRODUCTION

Alongside the datafication of society, we have witnessed a rise of open data initiatives, with governments and civic organizations sharing data to support transparency and civic participation. Open civic data is data about our communities that is made available without restrictions. Public, academic, government, and special libraries all have potential roles to play as "civic data intermediaries" organizations that enable the public to find and use data about their communities.

This poster presents an instructional design project that resulted in a series of modules and supporting materials targeted to two groups of learners: LIS students and existing library workers. We designed this series to support the development of civic data literacies among these two learner groups and their adoption of civic data intermediary roles in libraries and other public-facing information organizations. By serving as civic data intermediaries, library and information workers can work with the public to find and use open civic data in ways that benefit their lives and communities.

UNDERSTANDING CIVIC DATA LITERACY

This project understands "civic data" as data produced by governmental and community organizations that describes our communities. This category of data is expansive, including data about local transit systems, environmental conditions, public utilities, and housing. Globally, governments are making this data available as "open data," or data that is absent of restrictions to use. This opening of data, often codified in policy, is driven by expressed commitments to government transparency, accountability, civic participation, and innovation (Attard et al., 2015).

While the open data movement is motivated by the promise of data, this promise is constrained by skill barriers, data anxiety, and lack of awareness of the availability of data. D'Ignazio (2017), for example, observes "a growing gap between those who can work effectively with data and those who cannot" (p. 6). Moreover, data about communities can be used in ways that harm, rather than benefit, communities and, in particular, minoritized individuals. Open civic data, like all data, is not neutral, and library workers must understand the complex information ecosystems that impact their communities and the potential consequences of data use for their patrons. While library engagement with open civic will not eliminate all potential harms, libraries as intermediaries can support broader and meaningful community use, including the use of data to interrogate systems and structures that impact individuals.

For this project, we adopt and expand Ridsdale et al's definition of data literacy: "[Civic d]ata literacy is the ability to collect, manage, evaluate, and apply civic data, in a critical manner" (2015, p. 2). Our project's understanding of "civic data literacy" is informed by work on "critical data literacy." danah boyd & Kate Crawford (2012), Ruha Benjamin (2013) and the Ida B. Wells Just Data Lab, Lauren Klein & Catherine D'Ignazio (2020) and Data for Black Lives (D4BL) are just some of the contributors voicing the importance of moving beyond building technical data skills. Instead, the ability to question, investigate, and evaluate datasets and the systems and methods used to collect the data is foundationally important for social justice and to mitigate data harms.

INSTRUCTIONAL DESIGN METHODOLOGY

The premise of this project is that there are valuable roles that libraries can play as civic data intermediaries. In service of these roles, we designed civic data literacy instructional materials that build capacity for library participation in their civic data ecosystems. Through literature on civic data and data literacy, we identified a set of competencies that underpin civic data work: Civic Data Acquisition and Collection; Analysis; Communication and Visualization; Ethics; Management and Organization; Metadata; Policy and Copyright; Preservation; Publication and Dissemination; and Security and Privacy.

85th Annual Meeting of the Association for Information Science & Technology | Oct. 29 – Nov. 1, 2022 | Pittsburgh, PA. Author(s) retain copyright, but ASIS&T receives an exclusive publication license.

We were guided by the ADDIE instructional design approach for creating instructional materials. The ADDIE approach calls for an iterative process of analysis, design, development, implementation, and evaluation. To inform our analysis stage, we conducted focus groups and administered a survey to learn about barriers to civic data work in libraries and gaps in understanding that affect public and academic library workers' comfort with beginning to engage with civic data. Our focus groups were held with public librarians from our regional libraries. We received 91 responses to our survey, with 57 participants reporting an affiliation with a public library, 27 with an academic library, and the remaining from school, government or unidentified types of libraries. Through a second survey instrument, we also gathered 11 chief data officers' perspectives on skills and understandings that are essential to civic data intermediary roles. We report on data collection methods in this analysis stage in Chaar-Pérez et al., 2021.

In addition to the literacy areas identified above, our data pointed to other competencies to support through our instructional materials, including how to build effective partnerships with local civic data initiatives; the principles and lifecycle of library civic data; and the design of outreach strategies around civic data.

Following the creation of our modules, we invited feedback from LIS students, who reviewed 3-4 modules asynchronously and provided input on the clarity, format and delivery, and connections between their content and other concepts in librarianship. We used this evaluation data as we revised and revisited the materials.

MODULES

We crafted a series of openly licensed online modules and supporting instructional materials (Table 1) and published them through GitHub Pages. These modules reflect the data competencies that are important for civic data intermediary work in libraries. These materials include 10-15 minute recordings, slides, scripts, activities, and resources and can be used for asynchronous learning or integration in classroom teaching.

Segment	Coverage
Segment 1: Introducing Civic Data	This segment provides an introduction to civic data and equitable civic data work in libraries. This segment explains the importance of civic data by discussing its meanings and uses, its lifecycle and principles, and how to analyze it critically.
Segment 2: Civic Data Ecosystems, Partnerships, and Community Needs	This segment provides an overview of how to understand, approach, and work with the people and organizations that are connected to civic data, with a focus on ecosystem mapping, partnership building, and community needs assessment.
Segment 3: Preparing Libraries for Sharing their Data	This segment centers on how libraries can share their data through a process that considers equity, privacy, open data standards, metadata, and documentation.
Segment 4: Community Engagement through Civic Data	This segment focuses on how to engage communities with civic data, using data visualization, narrative strategies, and outreach and programming.
Segment 5: Civic Data Inquiry	This segment applies data skills basics to an example civic dataset created by a library. Learners are supported in cleaning, manipulating, sorting and analyzing a dataset to address an information need.

Table 1. Organization of Modules

CONCLUSION

This poster introduces the ASIST community to modules that can be integrated into coursework and training to support preparation for civic data intermediary roles. In libraries, these intermediary roles may involve supporting patrons' data use through data literacy workshops or publishing library data that provides insight into the community's use of public services. Through a critical approach that undergirds all of the modules and through skills development, this series supports the preparation of civic data intermediaries who work in service of social justice and advance the use of data for public good.

REFERENCES

- Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly*, 32(4), 399-418. <https://doi.org/10.1016/j.giq.2015.07.006>
- Benjamin, R. (2019). *Race after technology: Abolitionist tools for the New Jim Code*. Polity Press.

- boyd, d., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, 15(5), 662–79.
<https://doi.org/10.1080/1369118X.2012.678878>
- Chaar-Pérez, K., Mattern, E., Raphak, M., & Burton, M. (2021). Exploring civic data work in libraries: an opportunity for LIS curriculum and community empowerment. Proceedings of ALISE 2021. <https://hdl.handle.net/2142/110933>
- Mattern, E., Rapchak, M., Chaar-Pérez, K., Monk, L., Thaler, J., Burton, M., & Biehl, J. (2022). Civic Data Education Series. <https://civic-switchboard.gitbook.io/education-series/civic-data-education-series>
- D’Ignazio, C. (2017). Creative data literacy: Bridging the gap between the data-haves and data-have nots. *Information Design Journal*, 23(1): 6-18. <https://doi.org/10.1075/idj.23.1.03dig>
- D’Ignazio, C. and Klein, L.F. (2020). *Data feminism*. MIT Press.
- Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., Kelley, D., Matwin, S., & Wuetherick, B. (2015). *Strategies and best practices for data literacy education: Knowledge synthesis report*. Dalhousie University.
<https://doi.org/10.13140/RG.2.1.1922.5044>

Envisioning Ethical Mass Influence Systems

Mayhew, Alex	University of Western Ontario, Canada amayhew@uwo.ca
Chen, Yimin	University of Western Ontario, Canada ychen582@uwo.ca
Cornwell, Sarah, E.	University of Western Ontario, Canada scornwel@uwo.ca
Delellis, Nicole, S.	University of Western Ontario, Canada ndelelli@uwo.ca
Kelly, Dominique	University of Western Ontario, Canada dkelly48@uwo.ca
Liu, Yifan	University of Western Ontario, Canada yliu4369@uwo.ca
Rubin, Victoria, L.	University of Western Ontario, Canada vrubin@uwo.ca

ABSTRACT

This work envisions the possibility of ethical Mass Influence Systems (MISs). Modern algorithmic MISs, like Facebook and YouTube, have seen a link between the systems design for profit maximization and the increased radicalization of users (Wu, 2017). Using a Goals analysis grounded in philosophy (Falcon, 2022; Lipton, 1990; Bostrom, 2014), we will contrast the goals of existing algorithmic MISs with the goals of a future ethical algorithmic MIS. With the philosophical guidance of the Moral Parliament (Newberry & Ord, 2021) and the Moral Landscape (Janoff-Bulman & Carnes, 2013), we elaborate on a set of goals and mechanisms for promoting human flourishing via ethical MISs.

KEYWORDS

Ethics, Analytic Philosophy, Proximate and Ultimate Causation; Moral Parliament, Moral Landscape, Flourishing

INTRODUCTION

Modern algorithmic MISs, like Facebook and YouTube, are part of a long tradition of systems that influence opinions and behavior, like broadcast mass media. McQuail (1977) distinguishes between the effects and the effectiveness of mass media: The *effects* are any consequences of mass media operation, while the *effectiveness* is determined by mass media's capacity to achieve given objectives or goals. McQuail notes that both effects and effectiveness are important but require different types of analysis.

Philosophy has an even longer tradition of increasingly refined and nuanced understandings of goals. Aristotle distinguished many types of goals, or 'causes' as he called them. One of the most enduring is the *final goal*: the purpose for which a thing exists (Falcon, 2022). In analytic philosophy there is the notion of a *proximate goal*: the adequate proxy for the final goal (Lipton, 1990). Proximate goals can be used due to the difficulty in conceptualizing, measuring, or implementing final goals. The field of Artificial Intelligence ethics has popularized the notion of *instrumental goals*: the actions undertaken to achieve the proximate goal (Bostrom, 2014). In modern algorithmic systems the pursuit of these instrumental goals is often automated (e.g., as illustrated in Rubin (2022) for the purpose of identifying mis- and disinformation in online textual data). Returning to McQuail (1977), the final point of consideration is the actual impact caused by the pursuit of the instrumental goals.

ALGORITHMIC MASS INFLUENCE SYSTEMS GOALS ANALYSIS

An analysis of the various goals of current algorithmic MISs will form a baseline of comparison to a future envisioned more ethical MIS. The final goal of algorithmic MISs is profit generation (Wu, 2017), but simply wanting to make money is not a business plan though, so a proximate goal is needed. The proximate goal of current algorithmic MISs is to maximize user engagement and thus advertisement exposure (Bucher, 2018; Thaler et al., 2014). The instrumental goal pursued to maximize user engagement is the creation of a curated user experience. YouTube (Brown, et al., 2022) and Facebook (Rubin, 2022) create experiences that exploit human emotions and channel all people towards an increasingly narrow range of content. This leads to the actual impact of users becoming radicalized (Tufekci, 2018) and thus more likely to behave predictably when exposed to known types of content (Wu, 2017).

ETHICAL MASS INFLUENCE SYSTEMS GOALS ANALYSIS

When considering the possibility of an ethical algorithmic MISs, we may conclude that perhaps the best MIS is no MIS. Unfortunately, we do not have that luxury as the economic potential of current MISs ensures that they will continue to be used and the lack of boarders online means that the effectiveness of regulation is limited (Wu, 2017). Currently, algorithmic MISs are aimed at profit generation, with all of us as targets (Wu, 2017). To paraphrase a quote attributed to David Horowitz: If you aren't at the table, you are on the menu. As a result, when envisioning a redesign of future MISs, researchers should start at the beginning and ask: What, exactly, are we trying to achieve? Another way to ask this is: If not profits, what final goals should developers pursue with the system design? And, what final goals do users of MISs have in mind? And, how do those ultimate goals can be achieved via proximate or instrumental goals?

Final Goal

A commonly identified final goal for ethics in general is ‘*flourishing*’ or sustained happiness, also called ‘*eudaimonia*’ or ‘a life that goes well for the person leading it’ (Haybron, 2020). How do we navigate the many different conceptions of a ‘*good life*’? Rather than assert a particular conception of flourishing as correct, we can look to the notion of the Moral Landscape which suggests that there can be many different, even incompatible, ways of flourishing (Janoff-Bulman & Carnes, 2013). Peaks on the Moral Landscape are built out of different norms, values, and traditions but each peak is a way of being in the world that results in sustained happiness (Janoff-Bulman & Carnes, 2013). The Moral Landscape framework embraces the diversity of human experience while acknowledging the real harms and miseries that people endure.

Proximate Goal

Even with the notion of the Moral Landscape, flourishing remains difficult to rigorously define. To create an ethical MIS, an adequate proxy for flourishing is needed. Seligman (2012) suggests that flourishing can be thought of as the sustained practice of prosocial behavior which enhances the people’s subjective well-being. *Subjective well-being* (SWB) in online environments, similar to those an MIS would curate, has been the subject of our recent study (Delellis et al., 2022). Psychology research suggests a number of psychological traits that are broadly conducive to subjective well-being and are often considered prosocial, such as gratitude and grit (Seligman, 2012). Many schools of philosophy center virtues, such as honesty (Kaufmann, 2015), courage (Kraut, 2022), and curiosity (Baumgarten, 2001), as valuable for flourishing. Various philosophical virtues, psychological traits, and other factors, may serve as the building blocks of many different peaks on the Moral Landscape, via their connections to individuals’ subjective well-being.

Instrumental Goal

Varieties of human behavior expressed through language are increasingly detectable with natural language processing (NLP) techniques under development. Psychology has established methods for measuring subjective well-being (Layard, 2010). Using a variety of mechanisms, such as the design affordances of the user-interfaces (Kelly et al., 2022) and findings in NLP, it seems in principle possible for an ethical algorithmic MIS to nudge its users in a more prosocial direction (Thaler et al., 2014), and thus towards greater subjective well-being, or greater human flourishing. However, there could be conflicts between different peaks on the Moral Landscape, and there can even be trade-offs between virtues for people occupying a single peak. The Moral Parliament has been suggested as a decision-making mechanism when operating under moral uncertainty (Newberry & Ord, 2021). In a Moral Parliament various ethical theories (Meynell & Paron, 2021) each vote on preferred actions or trade-offs during conflicts. For example, some people may value honesty above all, while others would prioritize devotion and family ties. The outcome of those votes is then used to update the makeup of the parliament for the next vote. The Moral Parliament preserves the diversity of human experience and values instead of holding one theory of ethics above all others as correct (Newberry & Ord, 2021).

Actual Impact

Final, proximate, and instrumental goals are all needed to determine the effectiveness of an ethical MIS, but with reference to its actual effects and impact. Current algorithmic MISs are profit motivated and are known to cause radicalization in their users (Wu, 2017), any MIS attempting to promote flourishing would need to be mindful of such unintended consequences. Making the task more challenging is that the actual impacts of MISs can manifest at “the level of the individual, the group, the institution, the whole society or the culture” (McQuail, 1977).

CONCLUSION

The components laid out above do not represent a full blueprint for the creation of an ethical algorithmic MIS. The goals and impacts outlined here are the starting points for more investigation. The robustness of and translation between each of the goal types and their effects is subject to ongoing investigation from some quarters (Bostrom, 2014), and is ripe for further inquiry. As algorithmic MISs continue to become more powerful and have greater impact on our lives, it is becoming ever more imperative that we find ways to not just mitigate the harm they can cause but also harness their potential to do good, if only to ameliorate the harm already caused. This will require the re-envisioning of algorithmic MISs with ethical considerations in mind.

ACKNOWLEDGEMENTS

This work has been funded by the University of Western Ontario Strategic Support for the Canadian Sciences and Humanities Research Council (SSHRC) Success grant, awarded to Dr. Rubin in fall 2021. The grant was entitled “Rewiring for Happiness: Computing for Kinder, Healthier, and Compassionate Life Experiences with Information Commutation Technologies (ICTs).”

REFERENCES

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

- Brown, M. A., Bisbee, J., Lai, A., Bonneau, R., Nagler, J., & Tucker, J. A. (2022). *Echo Chambers, Rabbit Holes, and Algorithmic Bias: How YouTube Recommends Content to Real Users*. Available at SSRN 4114905.
- Bucher, T. (2018). *If... then: Algorithmic power and politics*. Oxford University Press.
- Baumgarten, E. (2001). Curiosity as a moral virtue. *International Journal of Applied Philosophy*, 15(2), 169-184.
- Delellis, N. S., Kelly, D., Liu, Y., Mayhew, A., Chen, Y., Cornwell, S. E., & Rubin, V. L. (2022).. Applying Positive Psychology's Subjective Well-Being to Online Interactions. *85th Annual Meeting of the Association for Information Science and Technology (ASIS&T2022): "Crisis, Transition, Resilience: Re-Imagining an Information-Resilient Society,"* October 29 - November 1, 2022, Pittsburgh, PA.
- Falcon, A. (2022). Aristotle on Causality, *The Stanford Encyclopedia of Philosophy*.
<https://plato.stanford.edu/archives/spr2022/entries/aristotle-causality>
- Haybron, D. (2020). Happiness, *The Stanford Encyclopedia of Philosophy*.
<https://plato.stanford.edu/archives/sum2020/entries/happiness>
- Janoff-Bulman, R., & Carnes, N. C. (2013). Surveying the Moral Landscape: Moral motives and group-based moralities. *Personality and Social Psychology Review*, 17(3), 219-236.
- Kelly, D., Liu, Y., Mayhew, A., Chen, Y., Cornwell, S. E., Delellis, N. S., & Rubin, V. L. (2022). Supporting Prosocial Behaviour in Online Communities through Social Media Affordances. *85th Annual Meeting of the Association for Information Science and Technology (ASIS&T2022): "Crisis, Transition, Resilience: Re-Imagining an Information-Resilient Society,"* October 29 - November 1, 2022, Pittsburgh, PA.
- Kraut, R. (2022). Aristotle's Ethics, *The Stanford Encyclopedia of Philosophy*.
<https://plato.stanford.edu/archives/sum2022/entries/aristotle-ethics/>
- Kaufmann, W. A. (2015). *The Faith of a Heretic*. Princeton University Press.
- Layard, R. (2010). Measuring subjective well-being. *Science*, 327(5965), 534-535.
- Lipton, Peter (1990). Contrastive Explanation. *Royal Institute of Philosophy Supplement* 27:247–266.
- McQuail, D. (1977). The influence and effects of mass media. *Mass Communication and Society*, 70-94.
- Meynell, L. & Paron, C. (2021). *Applied Ethics Primer*. Pressbooks. <https://caul-cbua.pressbooks.pub/aep/>
- Newberry, T., & Ord, T. (2021). The Parliamentary Approach to Moral Uncertainty. *Future of Humanity Institute*.
- Rubin, V. (2022). Misinformation and Disinformation: Detecting Fakes with the Eye and AI. *Springer*.
- Seligman, M. E. (2012). *Flourish: A visionary new understanding of happiness and well-being*. Simon and Schuster.
- Thaler, R. H., Sunstein, C. R., & Balz, J. P. (2014). Choice architecture. In *The Behavioral Foundations of Public Policy*. Princeton University Press. <http://dx.doi.org/10.2139/ssrn.2536504>
- Tufekci, Z. (2018). YouTube, the Great Radicalizer. *New York Times*.
<https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html>
- Wu, T. (2017). *The attention merchants: The Epic Scramble to Get Inside Our Heads*. Vintage.

A Multi-Methodological Assessment to Further the Role of Alabama's Public Libraries in Addressing the Opioid Crises

Mehra, Bharat

University of Alabama, USA | bmehra@ua.edu

Viehouser, Misha

University of Alabama, USA | miviehouser@crimson.ua.edu

Jaber, Baheya

University of Alabama, USA | bsjaber@crimson.ua.edu

ABSTRACT

This poster highlights the use of a multi-methodological approach to assess the role of Alabama's public libraries in addressing the opioid crises. Select findings/results from mixed research methods are reported to document the experiences of Alabama's library staff about their information offerings, activities, and community engagement and analyze web representation of their relevant services. The following research objectives were achieved over a two-year time-period (2019-2021): 1) We implemented a quantitative online survey to collect feedback from 36 Alabama's librarians about their community engagement experiences and what past and current role public libraries play in community health building processes. 2) We conducted qualitative interviews and focus groups with fifteen state librarians about their role as community hubs to address the opioid crises. From this dataset we developed a programming taxonomy and engagement prototype to identify preliminary activities, partners, challenges, outcomes, potential actions, resources, and best practices. 3) We conducted content analysis on websites of Alabama's 230 public libraries about availability, engagement, and representation of health information support services to overcome the opioid crises. This poster provides a glimpse of select activities conducted in this regard, reports select findings and results from the various methods used and highlights methodological implications and issues.

KEYWORDS

Alabama; multi-methodological approach; opioid crises; public libraries

INTRODUCTION

In 2016, 42,000 Americans died of an opioid overdose, since then the reality of accelerated opioid misuse, everyday addiction, and overdose deaths in the United States has led to a serious national crisis of terrifying proportions (Chase, 2018; King, 2019). Alabama has not been immune to increased age-adjusted drug overdose death rates with the state experiencing limited opioid health prevention, treatment, recovery, and effective relapse-prevention solutions. The National Institute on Drug Abuse (2019) reported that there was a significant increase of 11.1 percent in the age-adjusted drug overdose death rate in Alabama from 2016 (16.2 per 100,000) to 2017 (18.0 per 100,000). Since then, the addiction and misuse of opioids, including prescription pain relievers, heroin, and synthetic opioids (e.g., 3 fentanyl) has become a serious national crisis with more than 130 people deaths reported every day owing to opioid overdoses (Centers for Disease Control and Prevention, 2018). Fifty-five of 67 counties in Alabama are designated rural (82.10%), 37 counties are Appalachian (55.22%), and with the state's Civil Rights struggles even till recent times, the health disparities accumulate (Alabama Rural Health Association, n.d., Appalachian Regional Commission, n.d., Pearl, 2015). For example, the following are a few glimpses of health conditions and health injustices that underly deeper, complex socio-cultural, sociopolitical, socioeconomic, and socio-environmental problematics in the state. Relative to national health, the Commonwealth Fund's (2019) *2019 Scorecard on State Health System Performance* ranked Alabama as 38th in overall health services, 35th in health access and affordability, 30th in health prevention and treatment, 34th in avoidable hospital use and cost, 46th in its residents' healthy lives, and 36th in health disparities. In the *Alabama Health Disparities Status Report 2010* (Alabama Department of Public Health, 2010), the state was comparatively ranked as one of the worst in the country for several years, with higher rates of disease, injury, premature death, and disability, and considerable health care disparities between racial and ethnic minorities compared to Whites (Signorello, Hargreaves, and Blot, 2010). HealthyPeople.gov (2019) identified leading health indicators and health disparities in Alabama to include: cardiovascular conditions and diseases, cancer, diabetes, HIV/AIDS, infant mortality, and mental health illness, related but not limited to the lack of physicians in rural areas, low health literacy, unequal treatment, and exposure to environmental risks. The Alabama Public Health (2017) estimated that approximately 8,600 Alabama adults die each year from their own smoking, 800 adult nonsmokers die each year from exposure to secondhand smoke, and 108,000 Alabama children under 18 years will ultimately die prematurely from smoking.

THEORETICAL UNDERPINNINGS

Public libraries, especially in the southern states in the United States, have played a limited role in addressing the opioid crises at national, state, regional, and local levels (Bishop et al., 2018; Mehra, Bishop, and Partee II, 2018, 2017). As community hubs and information providers in rural and urban counties dispersed across the width and

85th Annual Meeting of the Association for Information Science & Technology | Oct. 29 – Nov. 1, 2022 | Pittsburgh, PA. Author(s) retain copyright, but ASIS&T receives an exclusive publication license.

breadth of the state (and country), public libraries can potentially play a significant role to become active stakeholders in the design and development of effective health information and support services in opioid prevention, treatment, recovery, and relapse prevention solutions (McCook, 2011; Whiteman et al., 2018; Wiegand, 2017). The project goal is to assess existing and future engagement in Alabama's public libraries to advance their roles in the development of health information and support services to overcome the opioid crises. Mixed methods (quantitative and qualitative) in this research included documentation of experiences of Alabama's library staff about their information offerings, activities, and community engagement to address the opioid crises and web content analysis of representation of relevant services. This project extends theoretical underpinnings from past work by: 1) Bridging discourse between human information behavior theories (e.g., information use and information practices) in libraries and library service practices with health impacts (i.e., social justice and community engagement) (Mehra, Sikes, and Singh. 2020) Documenting the experiences and perspectives of public library staff in a state that is painted solely in negative stereotypes recognizes their role in asset management and gives "voice" to an underrepresented population (Gray and Mehra, 2021).

OBJECTIVE 1 & OUTCOMES (PHASE 1)

We implemented a quantitative online survey (with open-ended questions) through Qualtrics and collected feedback from 36 Alabama's librarian staff stakeholders about their community engagement experiences and initiatives to address the opioid crises, and what past and current role public libraries play in community health building processes and community health development (APLS, n.d.). Data analysis provides an initial assessment of the role of Alabama's rural librarians in community development and needed mobilization to address the Opioid crisis.

OBJECTIVE 2 & OUTCOMES (PHASE 2)

We conducted qualitative interviews and focus groups with fifteen state librarian staff stakeholders about their role as community hubs to address the opioid crises. From this dataset we developed a programming taxonomy and engagement prototype to identify preliminary activities, partners, challenges, outcomes, potential actions, resources, and best practices.

OBJECTIVE 3 & OUTCOMES (PHASE 3)

We conducted content analysis on websites of Alabama's 230 public libraries about availability, engagement, and representation of health information and support services to overcome the opioid crises. Contextually relevant strategic information tools developed from data collected (e.g., roadmap, action plan, taxonomic framework of information responses) will strengthen the role of Alabama's 230 public libraries in health partnerships to address the opioid crises.

CONCLUSION

The use of multiple research methods of data collection and data analysis provided a holistic approach to assess the role of Alabama's public libraries to address the opioid crises. Findings from the online survey complemented the information gathered during the interviews and focus groups and content analysis from the websites to help develop tangible action-oriented information products (e.g., drafts of a Roadmap & Strategic Action Plan and programming taxonomic framework of health information and support services) in this regard. The former provided tools to facilitate action (based on current realities) while the latter described and visually articulated those realities. These are providing contextually relevant culturally applicable information strategies to Alabama's public libraries for helping them play a more significant role in addressing the opioid crises. In this manner, they are promoting principles of social justice (i.e., fairness, justice, equity/equality, empowerment, change agency, community development) in a religiously and politically conservative part of the country that has remained marginalized and entrenched in its dated practices in the delivery of health information support services (Mehra, 2022). Finding such novel ways to enhance their roles to operationalize and implement social justice will address gaps between the haves and have-nots (Mehra and Jaber, 2021). This poster provides a glimpse of select research activities in the project. It also reports select findings and results from the various methods used and highlights methodological implications and issues involved in developing a multi-methodological approach to identify complexities and overcome multi-factorial layers of intertwining health injustices (in the plural) experienced in Alabama's communities and their embedded public libraries (Alabama Primary Health Care Association, 2019).

ACKNOWLEDGMENTS

We are grateful to the University of Alabama Office of Research and Economic Development's Small Grant Program for their support of the research reported in this poster via funding of the grant entitled "An Exploratory Assessment of the Role of Alabama's Public Libraries to Address the Opioid Crises" (January 2020 – December 2021).

REFERENCES

- Alabama Department of Public Health. (2010). *Alabama Health Disparities Status Report 2010*. Montgomery, AL: Alabama Department of Public Health, Office of Minority Health. <https://www.astho.org/Programs/Health-Equity/Alabama-Health-Equity-Report/>.
- Alabama Primary Health Care Association. (2019). *Mission. Members. Service. About Us: We are APHCA*. <https://www.alpha.com/about-us/>.
- Alabama Public Health. (2017). *Health Disparities: Eliminating Tobacco-Related Disparities*. Tobacco Prevention and Control, Bureau of Prevention, Promotion, and Support. <https://www.alabamapublichealth.gov/tobacco/health-disparities.html>.
- Alabama Public Library Service. (n.d.) Public Library Listings. <http://fmweb.apls.state.al.us/libinfo2/recordlist.php?-max=500&-action=findall&-skip=0&-link=Find%20All>.
- Alabama Rural Health Association. (n.d.). *Definition of Rural Alabama*. <https://arhaonline.org/definition-of-rural-alabama/>.
- Appalachian Regional Commission. (n.d.). *Counties in Appalachia*. Washington, D. C.: Appalachian Regional Commission. https://www.arc.gov/appalachian_region/CountiesinAppalachia.asp.
- Bishop, M. R., et al. (2018). Establishment of the Alabama Hereditary Cancer Cohort: Strategies for the Inclusion of Underrepresented Populations in Cancer Genetics Research. *Molecular Genetics & Genomic Medicine*, 6(5), 766-778. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6160710/>.
- Centers for Disease Control and Prevention (CDC). (2018). *National Vital Statistics System: Mortality*. Atlanta, GA: US Department of Health and Human Services.
- Chase, D. (2018). *The Opioid Crisis Wake-up Call: Health Care is Stealing the American Dream. Here's How We Take It Back*. Seattle, WA: Health Rosetta Media.
- Gray, L., and Mehra, B. (2021). Going Against the Current of Hegemonic “White-IST” Discourse: A Doctoral Program Journey from Critical Student Z+Z Guide Perspectives. *Journal of Education for Library and Information Science*, 62(2), 182-200.
- HealthPeople.gov. (2019). Healthy People 2020—A Framework for Prevention for the Nation: DATA2020. Office of Disease Prevention and Health Promotion—U.S. Department of Health and Human Services (HHS). <https://www.healthypeople.gov/2020/data-search/>.
- King, T. M. (2019). *Addiction Nation: What the Opioid Crisis Reveals about us*. Harrisonburg, VA: Herald Press.
- McCook, K. de la. P. (2011). *Introduction to Public Librarianship*. Chicago, IL: Neal-Schuman Publishers, Inc.
- Mehra, B. (2022). *Social Justice Design and Implementation in Library and Information Science*. Abingdon, United Kingdom: Routledge.
- Mehra, B., Bishop, B. W., & Partee, R. P. II. (2018). A Case Methodology of Action Research to Promote Economic Development. *Journal of Education for Library and Information Science*, 59(1-2), 48–65.
- Mehra, B., Bishop, B. W., and Partee II, R. P. (2017). How Do Public Libraries Assist Small Businesses in Rural Communities? An Exploratory Qualitative Study in Tennessee. *Libri International Journal of Libraries and Information Studies*, 67(4), 245-260.
- Mehra, B., and Jaber, B. (2021). Opioid Consumer Health Information Literacies (o-CHIL) in Alabama’s Public Libraries: An Exploratory Website Content Analysis. In B. St. Jean, G. Jindal, Y. Liao, and P. Jaeger (eds.), *Roles and Responsibilities of Libraries in Increasing Consumer Health Literacy and Reducing Health Disparities* (Advances in Librarianship Series, Volume 47) (pp. 61-82). Bingley, United Kingdom: Emerald Group Publishing.
- Mehra, B., Sikes, E. S., and Singh, V. (2020). Scenarios of Technology Use to Promote Community Engagement: Overcoming Marginalization and Bridging Digital Divides in the Southern and Central Appalachian Rural Libraries. *Information Processing & Management*, 57(3). Article 102129. <https://doi.org/10.1016/j.ipm.2019.102129>.
- National Institute on Drug Abuse. (2019). *Alabama Opioid Summary (May 2019)*. Washington, DC: National Institute on Drug Abuse.
- Pearl, M. (2015). Fifty Years After Selma, Civil Rights in Alabama Are Still in Rough Shape. VICE: News, March 9. https://www.vice.com/en_us/article/exmj47/50-years-after-selma-civil-rights-in-alabama-are-still-in-rough-shape-992.
- Signorello, L. B., Hargreaves, M. K., and Blot, W. J. (2010). The Southern Community Cohort Study: Investigating Health Disparities. *Journal of Health Care for the Poor and Underserved*, 21(1), 26-37.
- The Commonwealth Fund. (2019). *2019 Scorecard on State Health System Performance: Alabama*. <https://scorecard.commonwealthfund.org/state/alabama/>.
- Whiteman E. D., Dupuis, R., Morgan, A. U., D’Alonzo, B., Epstein, C., Klusaritz, H., et al. (2018). Public Libraries As Partners for Health. *Preventing Chronic Disease: Public Health Research, Practice, and Policy*, 15(170392). Centers for Disease Control and Prevention. DOI: <http://dx.doi.org/10.5888/pcd15.170392> external icon; https://www.cdc.gov/pcd/issues/2018/17_0392.htm.
- Wiegand, W. A. (2017). *Part of Our Lives: A People’s History of the American Public Library*. Oxford: Oxford University Press.

Quantifying Individual Research's Distance from the Trends based on Dynamic Topic Modeling

Meng, Jie University of Chinese Academy of Sciences, People's Republic of China | moonjaymengjie@gmail.com
Lou, Wen East China Normal University, People's Republic of China | wlou@infor.ecnu.edu.cn
He, Jiangen University of Tennessee, Knoxville, USA | jiangen@utk.edu

ABSTRACT

Research trends are the keys for researchers to decide their research agenda. However, only few works have tried to quantify how scholars follow the trends. This paper addresses this problem by proposing a novel measurement for quantifying how a scientific entity (paper or researcher) follows the hot topics in a research field. Specifically, the topic evolution and papers are vectorizing by dynamic topic modeling. Then the degree of hotness tracing is explored from three different perspectives: mainstream, short-term direction, long-term direction. Papers and researchers in the field of Computer Vision from 2006 to 2017 were selected to evaluate our method. Further study will show the results of topic evolution patterns and researchers' clusters.

KEYWORDS

Research trends; Dynamic topic modeling; Hot topics; Research behavior; NLP

INTRODUCTION

The shocking research alerted that the progress of large scientific fields may be slowed canonical (Chu and Evans, 2021), which revealed a phenomenon that many researchers would follow mainstream research over time. Quantifying the mainstream in a research field and identifying researchers with different behaviors can be challenging (Small et al., 2014). We aim to model such research trends and the distances of each scientific entity (e.g. a paper, a researcher) from the trends to explore researchers' behaviors in science communities.

METHOD

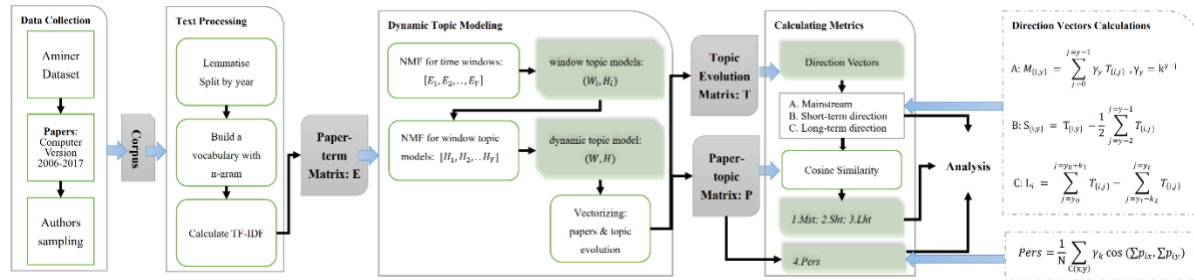


Figure 1. Research design and dynamic topic modeling process

We designed the method as four steps with three direction vectors and four metrics (Figure 1). Firstly, basic text processing techniques are utilized to create a paper-term matrix, E . Secondly, we quantify the trends and papers with a set of dynamic topics $DT = [dt_1, dt_2, \dots, dt_n]$, which is generated by dynamic topic modeling (DTM). DTM is implemented by three layers of Non-negative Matrix Factorization (NMF) (Greene and Cross, 2017), producing the topic model (W, H) . The i -th row of H represents the term distribution of dt_i . The papers P_y (for year y) can be vectorized by solving the linear system $P_y H = E_y$. Thus, a paper is represented by the distribution of DT . Further, we model the topic trends as a topic evolution matrix T . Two steps are included: (1) Topics with n largest weights are assigned to a paper, and other values in P_y are set to zero. (2) Calculating the y -th column of T by aggregating the topic vectors of all papers in one year $T_y = \sum_i P_{iy} \cdot T_{iy}$. T_y represents the distribution of DT in the y -th year of this field.

The trends of hot topics are described from three perspectives. Accordingly, three direction vectors are defined to quantify these trends. The *Mainstream* (ma) of a field is a direction vector representing established research agenda. It's a measure of the accumulated topics over the years, calculated by a weighted summing of T , as in Figure 1. γ represents decay coefficient and $k < 1$. (We set $k = 0.8$). Since papers from far-off years have less of an impact on the mainstream, the decay coefficient decreases with increasing proximity to year y . The *Short-term Direction* (sd) is to measure topics' popularity over a brief period of time (2 or 3 years). The calculation takes into account the difference between the topic vectors for year y and the average of the topic vectors for the two years prior. The

Long-term Direction (ld) depicts a growth tendency over a long period of time, measuring the variation in average topic vectors between the starting and ending years.

Once the pattern of topic trend evolution is established, we can quantify individual research’s similarity with the trends. Given a direction vector $D (ma, sd, ld)$, three metrics for papers are defined: (1) *Mainstream tracing degree (Mst)*, (2) *Short-term hotness tracing degree (Sht)*, (3) *Long-term hotness tracing degree (Lht)*, which are derived by cosine similarity: $\cos \langle P_i, D \rangle$. For authors, these metrics are defined as the arithmetic mean of the indicator of their papers: $(1/N) \sum_{i=1}^N \cos \langle P_i, D \rangle$. For a thorough understanding of the scholar’s research interests, we propose another metric, (4) *Persistence (Pers)*, to quantify the continuation of the researcher’s work. It is derived by weighted summing the similarity of the topic vectors of the scholars’ papers from various years, as illustrated in Figure 1. $\sum p_{ix}$ represents the sum of all paper vectors in year x ; $C(x, y)$ denotes all combinations of (x, y) ; γ_k denotes the decay coefficient, $\gamma_k = k_1 + |x - y| * k_2$, $k_2 < 0$ (Here we set $k_1 = 2, k_2 = -0.25$). The decay coefficient indicates that two closer years are given more weight.

RESULT AND DISCUSSION

The dataset used in this article comes from AMiner (Wan et al., 2019), where the authors are disambiguated. AMiner assigns one or more discipline(s) to each publication. The final dataset, including of 279,875 articles, is composed of computer vision papers published between 2006 and 2017. Authors who are active in the field of computer vision are sampled by the two criteria: 1. They have more than three papers. 2. They have published in the field for more than three years. Finally, 45,203 unique authors are included.

Spearman correlation analysis (Table 1) shows that all metrics for papers are positively correlated with each other, among which *Lht* has more significant correlation with *Mst*, while correlation between *Sht* and *Mst* is relatively weaker. The results reveal that papers following long-term hotspots are more consistent with the mainstream than short-term hot topics. For authors, *Sht* has the most significant positive correlation with *Lht*, revealing that the researcher who prefers long-term hotspots also tends to follow short-term direction. Besides, *Pers* has very little correlation with other indicators, indicating that adherence to a research direction is not related to following the hot topics.

<i>Metrics</i>	<i>Papers</i>			<i>Authors</i>			
	<i>Sht</i>	<i>Lht</i>	<i>Mst</i>	<i>Sht</i>	<i>Lht</i>	<i>Mst</i>	<i>Pers</i>
<i>Sht</i>	1	0.592	0.379	1	0.719	0.413	0.058
<i>Lht</i>	0.592	1	0.711	0.719	1	0.666	0.058
<i>Mst</i>	0.379	0.711	1	0.413	0.666	1	0.049
<i>Pers</i>	\	\	\	0.058	0.058	0.049	1

Table 1. Correlations between indicators for papers and authors

CONCLUSION

The proposed approach explored how individual research in a field follows trends. This study provided a new quantitative analysis method for studying scholars’ research interests and trend analysis. Future work includes optimizing the method to be applicable in a larger-scale dataset in a longer period, identifying and visualizing the exact groups in various behaviors.

REFERENCES

- Chu, J. & Evans, J. (2021). Slowed canonical progress in large fields of science. *PNAS*, 118(41), e2021636118.
- Greene, D., & Cross, J. (2017). Exploring the Political Agenda of the European Parliament Using a Dynamic Topic Modeling Approach. *Political Analysis*, 25(1), 77-94. doi:10.1017/pan.2016.7
- Small, H., Boyack, K. W., Klavans, R. (2014). Identifying emerging topics in science and technology. *Research Policy*, 43, 1450-1467
- Wan, H., Zhang, Y., Zhang, J., & Tang, J. (2019). Aminer: Search and mining of academic social networks. *Data Intelligence*, 1(1), 58-76.

Professional Development: Perception among School Librarians

Munyao, Moses M. University of Oklahoma, United States | Moses.Muendo.Munyao-1@ou.edu

ABSTRACT

The roles of school librarians as well as the needs of library users have been changing. To keep up with the changes, school librarians have been required to engage in professional development which enables them to acquire the necessary knowledge and skills for the effective provision of services and programs to their clients. School districts formulate policies that provide guidelines on the professional development requirements and activities for school librarians but there have been concerns on the effectiveness and approach of professional development as outlined in the policies. Using Ball's policy cycle theoretical framework, the study analyzed professional development policies within schools and interviewed school librarians to understand their perception of professional development. Findings from the study will enable librarians and administrators within the school and districts to understand the current state of professional development and may also suggest effective and innovative strategies in professional development for school librarians.

KEYWORDS

Professional development, School libraries, School librarians, Policy analysis, Policy cycle

INTRODUCTION

The social, cultural, political, and economic aspects of our world continue to change thus affecting the way we carry out our responsibilities, how we learn and grow. Professionals in different sectors strive to keep up with the changes by implementing different strategies, among them, professional development. Professional development (PD) is the process of engaging in activities that enable professionals, regardless of the organization, to improve in their roles by upgrading their competencies, attitudes, knowledge, and skills (Ende, 2016). Even though professionals possess certification acquired through formal education, "learning is a lifelong proposition" and to be the best, one must look beyond to encounter new and challenging ideas (Harlan, 2009, pg. 5). At the heart of professional development are policies that provide guidelines and procedures on professional development. Ball (1994, as cited in Heimans, 2010, p.4) described policy as "both text and action, words and deeds, it is what is enacted as well as what is intended". Apart from outlining the required aspects of PD programs, the policies should be generated by an inclusive team and strive to accommodate the different opinions of all staff within the school (Blandford, 2012). The research reported in this paper sought to provide an understanding of the perception related to the experiences in professional development among school librarians and relating the factors to the PD policies within their schools.

LITERATURE REVIEW

Professional development is evolving from the traditional approaches that include webinars, seasonal meetings, workshops, and summer institutes to include strategies that utilize the latest technology tools such as social media where librarians are creating learning networks that enhance their expertise of the profession (Moreillon, 2016; Cooke, 2012). Zepeda (2019) states that professional development is a continuous process that requires constant support to ensure that what was learned formally or informally is transferred into practice. Some of the traditional approaches of PD characterized by the workshop model fail to provide post-workshop support, hence in certain instances they become less effective at the workplace. According to Lieberman & Wilkins (2006) there is an increasing emphasis on standards, thus creating a shift in the delivery of professional development. School contexts are different and therefore the professional needs for staff will differ. Different professional organizations within the library field have outlined the tools and approaches necessary for an effective PD program. The IFLA (2015) module for education identifies the need for continuous professional development to refine the professional skills that meet the learning needs of the community. The American Library Association (ALA) has a compilation of professional development tools that have been adapted by library organizations and institutions across the country (ALA, n.d.). One of the common beliefs for the American Association of School Librarians (AASL) is that "qualified school librarians lead effective school libraries" (AASL, n.d.). Studies have identified the need for relevancy (Brown, Dotson, & Yontz, 2011; Harada, 2012), collaboration (Kammer, King, Donahay, Koeberl, 2021), ease of access and up to date programs (Hossain, 2017) as some of the important aspects to consider in professional development for school librarians. According to Brown, Dotson, & Yontz (2011), the process for planning, designing, implementing, and evaluation of professional development relies on two main elements "a strong theoretical base" and the knowledge from experts in the profession. Ball, Bowe & Gold (1992) formulated the "policy cycle", a theoretical framework for the analysis of educational policies. Ball et al (1992, as cited in Ball, 2006, p. 51) identified three main contexts in the policy analysis model: 1) the context of influence, 2) the context of policy text production, and 3) the context of practice. The context of influence is the environment through which key

policy concepts are initiated and acquire legitimacy. In this context, stakeholders interact to make decisions on what is to be included or excluded in the policy. Bowe et al (1992, p. 22) argue that “policy texts represent policy” and therefore in the context of text production “policy texts are normally expressed in the language of general public good appealing to popular common sense and written in relation to an ideal set of conditions, that is, idealizations of the real world” (Bowe et al, 1992, p. 21). The context of practice is where policies are subjected to interpretation and produces effect (Vidovich, 2013).

RESEARCH DESIGN

PD programs within schools seem to take a one-size fits all approach focusing more on teachers. Much of the research has also focused on the professional development of schoolteachers but this research study sought to examine the perception of professional development among school librarians as outlined in the policies within their school districts. This paper focuses on the preliminary research findings of the perceptions of school librarians? by analyzing PD policies and conducting interviews with school librarians. Interviews were carried out among school librarians and 10 professional development policies from the southwest region of the United States were analyzed.

Data Analysis

The data, which are largely qualitative, derive from analysis of policy documents and interviews with school librarians. Ball’s (1992) contexts of the policy process provided the primary framework for analyzing the policies. The school policies were sourced online and from requests made to school librarians. The analysis focuses upon identifying inconsistencies between what policy says and policy in practice. Inductive thematic analysis was used to identify themes within the data where NVIVO was used to review and code the data. The themes identified focused on the requirements, strengths, challenges and recommendations regarding professional development.

Findings and Discussion

Some of the important issues in Ball’s context of text production include: whose interests the policy intends to serve, which interest (stakeholder) groups are represented in the production of the policy text, and which are excluded. Analysis of the policies indicate that the format and text of the policies is generalized to school staff, therefore applicable to school librarians too. For example, “The board believes that it has a responsibility to provide opportunities for the continual growth of its professional staff.” (Edmond public schools, 2021). Findings from the school librarians also indicate that even though not all professional needs are met through PD programs, the professional development provided so far is relevant to them as also highlighted by Brown, Dotson, & Yontz, 2011; Kammer, King, Donahay, Koeberl, (2021). One of the strengths noted by librarians is the ease of access of various PD opportunities (Hossain, 2017). The availability of technology and the changes that happened due to the pandemic, provided a myriad of professional development opportunities for free or subsidized prices. As stated by Ball (1992) the context of influence is concerned with the environment through which the policies are initiated therefore technological changes, diverse user groups and the ever-changing user needs are some of the influencing factors during the policy process (Moreillon, 2015; Cooke, 2012). Some of the issues in the context of practice include, who puts the policy into practice, how well the policy is received, the processes used to put the policy into practice and the effect. The policies outline yearly requirements for PD, but the school librarians interviewed were not fully aware of the PD requirements outlined by their school districts since they are in their early years in the profession. Through networking they are made aware of PD opportunities. While Zepeda (2019) recognizes the importance of continuous support in the practice phase of professional development, she also notes the lack of post-workshop support. The policies analyzed do not provide clear guidelines on how to carry out post-workshop support. School librarians identified the lack of follow ups as one of the challenges experienced in their professional development therefore the need for initiating better channels that follow up on their PD recommendations.

CONCLUSION

The continuous and fast changes experienced in society will continue to have an impact on the roles of school librarians. PD delivered to school librarians affects students and therefore it should be aimed at improving the quality of the professionals to enhance student learning and achievement. It is important to analyze the needs of both education professionals and students to identify gaps and determine what areas need more skills and knowledge. The need for relevant, up to date, continuous professional development should therefore be emphasized enough to ensure effective program and service provision to the diverse school library users. Administrators both within the school and the district should also be intentional and inclusive in the formulation of equitable policies that provide guidelines for professional development by taking into consideration the specific professional needs of school librarians and being aware of external influences. School librarians should also be proactive by networking and collaborating with others to help them identify, as well as participate in, relevant PD programs.

REFERENCES

Abilock, D., Fontichiaro, K., & Harada, V. H. (2013). Growing schools: Effective professional development. *Teacher Librarian*, 41(1), 8-13.

- Ball, S. (2006) *Education policy and social class: The selected works of Stephen J. Ball*. Abingdon: Routledge
- Blandford, S. (2012). *Managing professional development in schools*. Routledge.
- Ball, S., Bowe, R. & Gold, A. (1992) *Reforming education and changing schools: Case studies in policy sociology*. New York: Routledge
- Brown, C., Dotson, L., & Yontz, E. (2011). Professional development for school library media professionals: Elements for success. *TechTrends*, 55(4), 56-62.
- Ende, F., (2016). *Professional development that sticks: How do I create meaningful learning experiences for educators*. ASCD.
- Harlan, M. A. (2009). *Personal learning networks: Professional development for the isolated school librarian*. Libraries Unlimited.
- Heimans, S. (2010). Conceptualizing education policy as practice. *In AARE Annual Conference Proceedings*. Melbourne, Australia (pp. 1-16).
- Hossain, Z. (2017). Professional development via Facebook group: Perception of school librarians. *In IASL Annual Conference Proceedings*. Long Beach, CA. United States. <https://doi.org/10.29173/iasl7160>
- Kammer, J., King, M., Donahay, A, Koeberl, H. (2021). Strategies for successful school librarian and teacher collaboration. *School Library Research: Research Journal of the American Association of School Librarians* 24, 1-24.
- Lieberman, J. M., & Wilkins, E. A., (2006). The professional development pathways model: From policy to practice. *Kappa Delta Pi Record*, (Spring), 124-128.
- Moreillon, J. (2016). Building your personal learning network (PLN): 21st-century school librarians seek self-regulated professional development online. *Knowledge Quest*, 44(3), 64-69.
- Vidovich, L. (2013). Policy research in higher education: Theories and methods for globalising times? In J. Huisman & M. Tight (Eds.), *Theory and method in higher education research* (pp. 21–39). Emerald.
- Zepeda, S. J. (2019). *Professional development: What works*. Routledge.

LIST OF SCHOOL PROFESSIONAL DEVELOPMENT POLICIES ANALYZED

1. Edmond public schools
2. Frisco ISD
3. Little Elm ISD
4. McKinney ISD
5. Mustang public schools
6. Norman public schools
7. Perry public schools
8. Plano ISD
9. Tulsa public schools
10. Duncanville ISD

The Effects of Collaborative Country and Discipline on an International Research Collaboration Indicator: The Case of the United States

Nguyen, Ba Xuan

Victoria University of Wellington, New Zealand | ba.nguyen@vuw.ac.nz

ABSTRACT

This study explores the extent that international research collaboration indicators vary according to two factors: countries' global economic levels and type of disciplines. The study examines the case of asymmetric co-authorship relationships between the United States and its collaborative research countries, measured by the affinity index. The study finds that there is no significant interaction between the two factors surveyed. The discipline factor has a main effect on the asymmetric co-authorship relationships. No main effect of the country factor categorised by economic level is found, after allowing for the discipline factor. This study contributes to the social science research on the impact of factors on co-authorship in international research collaboration measurement indicators.

KEYWORDS

international research collaboration; co-authorship indicator; country; economic level; discipline

INTRODUCTION

International research collaboration (IRC) includes the activities engaged in by individuals from different countries in a shared project or research. In this context, IRC measurement (IRCM) is an important means for countries to evaluate the IRC capacity of their partners and themselves (Chen et al., 2019).

IRCs can be specified by many indicators including the numbers of international co-publications, or international co-patents. Generally, co-authorship has been used as a common indicator for research collaborations (Katz & Martin, 1997). It is worth noting that each IRC indicator can suggest a multi-dimensional structure from its own data. Specifically, *country* and *discipline* are two notable dimensions of IRCs. This is because an IRC measurement is often carried out for a country in a specific discipline. These dimensions are possible factors that might affect the outcomes indicated at IRC levels. First, the role of the country factor on IRC indicators is important and deserves to be explored (Glänzel, 2001). For example, the frequencies and benefits of IRC in different groups of countries may not be the same. Countries at the same income levels have been recognised to account for a higher percentage of IRCs in some disciplines (e.g., Ni & An, 2018) although the correlation between income levels and co-authorship is not always significant (Choi, 2012). This recognition can be explained because economic reasons play an important role in IRC activities (Luukkonen et al., 1992), and members in the same international associations tend to develop close research collaborations (Hou et al., 2021). Second, the discipline (i.e., field of study) factor might have an association with indicators' results in IRCM (Frame & Carpenter, 1979). For instance, IRC has different citation impacts on China's overall publications among different disciplines (Zhou & Glänzel, 2010). Previous studies have explored the effects of these factors on the impacts (i.e., citations) of IRC publications (e.g., Thelwall & Maflahi, 2020; Kwiek, 2021), but there is a lack of assessment of these effects on the indicators measuring IRC relations.

The purpose of this study is to determine the effects of the country factor and the discipline factor, in combination, on the asymmetric co-authorship relationships in IRCM. The United States (the USA) is examined in this study. The reason for this choice is because the USA has the highest total number of IRC relationships (Nguyen et al., 2019), and it is also among the top collaborative research partners for many other countries (Kwiek, 2021).

MATERIALS AND METHODS

The international co-authorships of the USA and other countries in the year 2015 were used for this exploratory study. The relevant publications were retrieved from the Dimensions data source using API calls. Regarding the *discipline* factor, Dimensions data source categorises its publications into 22 research disciplines, varying on a basic-to-applied spectrum. Eight disciplines were chosen in this study regarding whether they are predominantly basic or applied disciplines (Coccia & Bozeman, 2016). They include four more basic disciplines (specifically, Mathematical Sciences, Physical Sciences, Chemical Sciences, and Earth Sciences) and four more applied disciplines (specifically, Agricultural and Veterinary Sciences, Engineering, Technology, and Medical and Health Sciences). For the *country* factor, the numbers of international co-authorships between the USA and other countries were categorised by differences in economic levels. Two economic levels were used: OECD countries and non-OECD countries. These two economic levels were used because the GDP per capita of OECD is remarkably higher than that of non-OECD in the period surveyed (Liu & Matsushima, 2019), and the distance in GDP per capita is associated with IRC outcomes (Jiang et al., 2018). The OECD had 35 member countries in 2015, which means that

the USA had 34 OECD partners. Accordingly, the thirty-four most productive non-OECD countries (e.g., non-OECD countries having the most bilateral research publications in 2015) were also selected for this study.

The IRC relations of the USA and its research partners were counted from international co-publications. These values were then normalised by the total numbers of IRC relations credited to the research partners. The received asymmetric IRC relationships, also called the affinity index, reflect the importance (Chinchilla-Rodríguez et al., 2018; Nguyen et al., 2022) of the USA to its collaborative research countries. This study used two-way ANOVA to know how the country factor and the discipline factor, in combination, affect the mentioned research relationships. After an initial check, this study applied the logged values of affinity index in the model because they improved ANOVA assumptions of the errors' normality and homogeneity of variance.

ANALYSIS AND RESULTS

This study started with an interaction test. The null hypothesis is that there is no interaction between the country factor and the discipline factor in their effects on the asymmetric co-authorship relationships. Using R, the ANOVA table was produced (Figure 1). As the p-value for the interaction is higher than 0.05, the null hypothesis cannot be rejected at the 5% significance level. The interaction graph also shows that there is only a weak interaction between the two factors (Figure 2), so the study proceeded to the main effects tests. The study confirmed that the discipline factor has a main effect on the logged affinity index values at the 5% significance level. No significant main effect of the country factor is found, after allowing for the discipline factor.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
discipline	1	14.41	14.408	54.477	1.57e-11	***
country	1	0.10	0.101	0.381	0.538	
discipline:country	1	0.49	0.489	1.850	0.176	
Residuals	132	34.91	0.264			

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figure 1. The ANOVA table

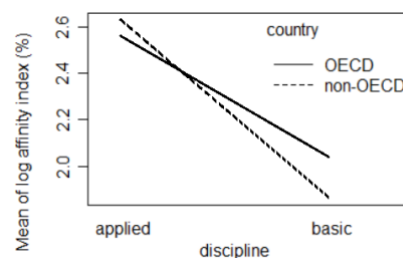


Figure 2. The interaction graph

Although there are notable differences in economic levels between the OECD countries and non-OECD countries, it is unexpected that this factor does not significantly affect these countries' affinity index values. A possible explanation is that the effects of GDP per capita on IRC relations between the USA and other countries are relatively corresponding to the effects on these countries' total IRC relations, and therefore these effects are eliminated in the affinity index. In contrast, the type of disciplines (i.e., basic or applied), has a constant effect across all logged values of the collaborative research countries' affinity index. The main effect of the discipline factor on the co-authorship indicator has been recognised in previous studies (e.g., Frame & Carpenter, 1979; Luukkonen et al., 1992; Coccia & Bozeman, 2016). In this study, Figure 2 shows that the affinity index values are higher for the applied disciplines than for the basic disciplines. These collaborative countries, therefore, realise that the importance of having collaborators in the USA changes according to the discipline type.

CONCLUSION

This study examined two sources (*country*, categorised by economic level, and *discipline*) of the asymmetric co-authorship relationships' variations in IRCM. A two-way ANOVA test was used to determine their effects. The study found that there is no significant interaction between the two factors surveyed. Furthermore, the discipline factor was the only factor that has a main effect on the asymmetric co-authorship relationships of the USA.

The findings in this study reveal the roles of underlying aspects to the asymmetric co-authorship relationships in IRCM studies. The surveyed asymmetric co-authorship relationships are significantly affected by the type of disciplines surveyed and not by the economic level of the authors' countries. This study may have limitations because the numbers of co-publications of each country might be affected by other determinants (e.g., social, historical, political, and geographical aspects), and other variables than GDP per capita might better capture an "economic level". Other limitations are that the contribution of the *time* factor, and the effects on symmetric co-authorship indicators' variations should also be examined. These limitations suggest further research in the future.

Despite the above limitations, this study contributes to research in IRCM studies. This study verifies the impact of the discipline factor on the asymmetric co-authorship relationships and shows that this factor is more significant than the economic level of countries in determining international co-authorship relationships.

ACKNOWLEDGMENTS

The author thanks Dr Ryan Admiraal at Victoria University of Wellington for helpful comments about the methods used in this study.

REFERENCES

- Chen, K., Zhang, Y., & Fu, X. (2019). International research collaboration: An emerging domain of innovation studies?. *Research Policy*, *48*(1), 149-168.
- Chinchilla-Rodríguez, Z., Bu, Y., Robinson-García, N., Costas, R., & Sugimoto, C. R. (2018). Travel bans and scientific mobility: utility of asymmetry and affinity indexes to inform science policy. *Scientometrics*, *116*(1), 569-590.
- Choi, S. (2012). Core-periphery, new clusters, or rising stars?: international scientific collaboration among 'advanced' countries in the era of globalization. *Scientometrics*, *90*(1), 25-41.
- Coccia, M., & Bozeman, B. (2016). Allometric models to measure and analyze the evolution of international research collaboration. *Scientometrics*, *108*(3), 1065-1084.
- Frame, J.D., & Carpenter, M. P. (1979). International research collaboration. *Social studies of science*, *9*(4), 481-497.
- Glänzel, W. (2001). National characteristics in international scientific co-authorship relations. *Scientometrics*, *51*(1), 69-115.
- Hou, L., Pan, Y., & Zhu, J. J. (2021). Impact of scientific, economic, geopolitical, and cultural factors on international research collaboration. *Journal of Informetrics*, *15*(3), 101194.
- Jiang, L. A., Zhu, N., Yang, Z., Xu, S., & Jun, M. (2018). The relationships between distance factors and international collaborative research outcomes: A bibliometric examination. *Journal of Informetrics*, *12*(3), 618-630.
- Katz, J. S., & Martin, B. R. (1997). What is research collaboration?. *Research policy*, *26*(1), 1-18.
- Kwiek, M. (2021). What large-scale publication and citation data tell us about international research collaboration in Europe: Changing national patterns in global contexts. *Studies in Higher Education*, *46*(12), 2629-2649.
- Liu, B., & Matsushima, J. (2019). Annual changes in energy quality and quality of life: A cross-national study of 29 OECD and 37 non-OECD countries. *Energy Reports*, *5*, 1354-1364.
- Luukkonen, T., Persson, O., & Sivertsen, G. (1992). Understanding patterns of international scientific collaboration. *Science, Technology, & Human Values*, *17*(1), 101-126.
- Nguyen, B. X., Dinneen, J. D., & Luczak-Roesch, M. (2022). Closed Shop or Collaborative Hub? An Analysis of the Partners' Importance in CANZUK Countries' Research Collaborations. *Frontiers in Research Metrics and Analytics*, *7*.
- Nguyen, B. X., Luczak-Roesch, M., & Dinneen, J. D. (2019). Exploring the effects of data set choice on measuring international research collaboration: An example using the ACM digital library and microsoft academic graph. *arXiv preprint arXiv:1905.12834*.
- Ni, P., & An, X. (2018). Relationship between international collaboration papers and their citations from an economic perspective. *Scientometrics*, *116*(2), 863-877.
- Thelwall, M., & Maflahi, N. (2020). Academic collaboration rates and citation associations vary substantially between countries and fields. *Journal of the Association for Information Science and Technology*, *71*(8), 968-978.
- Zhou, P., & Glänzel, W. (2010). In-depth analysis on China's international cooperation in science. *Scientometrics*, *82*(3), 597-612.

Is This a Good Idea? A Case Study on Civil Perception of Smart City Projects

Nguyen, Bich Chau
Zimmer, Franziska

Heinrich Heine University, Germany | bich.nguyen@uni-duesseldorf.de
Heinrich Heine University, Germany | franziska.zimmer@uni-duesseldorf.de

ABSTRACT

The purpose of this research is to pursue a better understanding of citizens' needs in a fast-changing medium-sized town transitioning to a smart city. The investigation focuses on the citizens' perceptions of smart city projects, the use intensity of said projects, and further examines what other factors play a part in shaping a citizen's opinion on smart city initiatives concerning their hometown. It was discovered that the majority of citizens perceive smart city projects positively, even though the projects are rarely used. Their behaviour and opinion were not related to their age, gender, and education in contrast to claims from previous research. Citizens had positive perception of the city's offer to involve them in urban planning, but otherwise refrain from actively getting involved. This was mostly due to the fact that they either did not feel the need to do so or were not confident enough in their contribution.

KEYWORDS

smart city; citizen participation; urban management; perceived usefulness; use intensity; transition

INTRODUCTION

The focus of smart city research has so far been on large cities or pilot projects such as Songdo or PlantIT Valley (Griffinger et al., 2007). But a city planning to transition to a smart city faces different hurdles and, in some cases, unique challenges (Heo et al., 2014) compared to "cities built from scratch" (Shelton et al., 2015, p. 14). The goal, therefore, is to shift the focus away from artificially built cities and metropolitan areas, as most city dwellers do not live in metropolitan areas or artificially created cities but in small and medium-sized ones (Giffinger et al., 2007). City construction should also be driven by and be tailored to the citizens' needs, investing in the overall improvement of the citizens' quality of life (Caragliu & Del Bo, 2019; Neirotti et al., 2014). Nevertheless, social aspects and the welfare of people and communities seem to "have a secondary role in smart city strategies" (Angelidou, 2017, p. 12). As such this study will focus on citizen participation in smart city projects. The city chosen is a small medium-sized town in Germany called Monheim am Rhein (around 43,000 residents). As of 2016, the city administration has been consistently implementing strategies intending to develop Monheim am Rhein into a smart city (Monheim am Rhein, 2020).

Based on the research model (Figure 1), the research questions (RQs) are:

- (RQ1) Do citizens perceive smart city projects as useful?
- (RQ2) Do citizens use the implemented smart city projects? If so, does increased use have an impact on perceived usefulness?
- (RQ3) Does citizens' attitude towards smart city projects differ depending on their age, gender, and education?
- (RQ4) Does city attachment influence a citizens' attitude towards city projects?
- (RQ5) Does citizens' attitude towards the city administration influence their perception towards smart city projects?

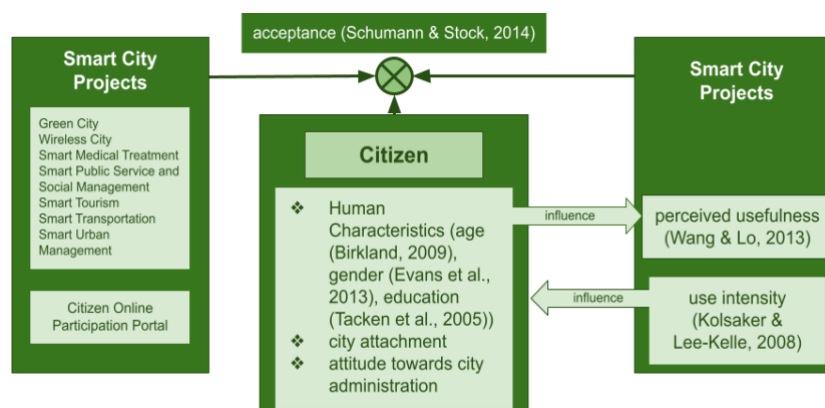


Figure 1. Research Model Based on the ISE Model (Schumann & Stock, 2014)

METHODS

To obtain the data on perceived usefulness and use intensity of smart city projects, an online survey was conducted. The survey was tailored to smart city projects, which are currently being worked on in Monheim am Rhein and which have already been completed during the time the survey was online. The data set was collected between May 2021 and July 2021. The survey was created with the Unipark/Questback Online Umfrage Tool for the citizens of Monheim am Rhein and, thus, was in German. It was mainly distributed on social media sites such as Facebook and Instagram. A total of 110 subjects completed the questionnaires of which twelve were submitted by people not currently living in Monheim. As such these questionnaires could not be used in this research, reducing the sample size to 98. Perceived usefulness and other sentiments were measured with a five-point Likert scale ranging from “strongly agree” to “strongly disagree” with an option for abstinence. Use intensity of projects was measured with a six-point Likert scale ranging from “daily” to “never”. The survey also gathered data on demographic variables such as age, gender, and education. For the data analysis the program SPSS ver. 26.0 was used to conduct a dependence analysis that examines the presence or intensity of a one-way associative relationship between one or more dependent variables and one or more independent variables.

RESULTS

RQ1. The majority of participants agree that smart city projects are useful. Projects concerning environmental issues were especially well received, while projects that focused more on the town’s appearance and marketability for tourism ranked the lowest.

RQ2. It turns out that projects that can be actively used by citizens are rarely or never used by most participants with the exception of the free Wifi provided by the town, which the majority claimed to be using several times a week. There is an overall correlation between perceived usefulness and use intensity of said projects.

RQ3. Neither age, gender, nor education had any significant impact on a participants’ attitude towards smart city projects.

RQ4. The attitude towards several projects differs depending on the participants’ satisfaction with living in Monheim am Rhein. Participants who do not like living there, perceived projects as less useful than their peers.

RQ5. Attitude towards projects differs depending on the participants’ feeling of involvement in said projects and their satisfaction with the city administration. If participants felt involved by the administration, projects were rated as useful. If participants liked the city administration they also perceived the projects as being useful.

CONCLUSION

The results show that smart city projects are seen as being useful overall. The tendency for older people to use technology less, as previous literature often claims (Fietkiewicz et al., 2016; Tacken et al., 2005), could not be proven by the study in Monheim am Rhein. In principle, the generations were in agreement with most of the projects and did not show significant differences. Instead, it turned out that participants (25 and below years old) were the largest group that did not know about the OPP, while all participants (70 and above years old) at least heard about it and some even used it. Concerning demographic change, it is particularly important to get more data on the age group above 70 years old in future research, as transportation can be an important issue. From what the input responses indicate, the lack of active participation is more likely due to citizens thinking that they have nothing to share. After all, sharing an opinion requires having an opinion in the first place (Broekens et al., 2010). It is, therefore, not always because of the system’s high degree of complexity that citizens do not participate, as some studies suggest (Cardullo & Kitchin, 2019; Townsend, 2013). Participants who do not like living in Monheim am Rhein perceived projects as less useful than their peers with an opposite opinion. Reasons for this are either a participant’s personal negative emotions towards the city which influence their judgement on the usefulness of projects or the projects have been affecting the quality of life of those participants in a negative way. Some implications of this study are that a transition towards a smart city is relevant for all ages. Therefore, it is important to design smart city projects according to all age groups for a more resilient society. In general, it is important to consider more aspects when planning and constructing a smart city, be it its people, their culture or its location. A next step could be to conduct a similar study like Ilhan, Möhlmann, & Stock (2015) did in Songdo, where they not only surveyed citizens but also included expert opinions in their analysis.

REFERENCES

- Angelidou, M. (2017). The role of smart city characteristics in the plans of fifteen cities. *Journal of Urban Technology*, 24(4), 3-28. <https://doi.org/10.1080/10630732.2017.1348880>
- Broekens, J., Pommeranz, A., Wiggers, P., & Jonker, C. M. (2010). Factors influencing user motivation for giving online preference feedback. *5th Multidisciplinary Workshop on Advances in Preference Handling (MPREF'10)*. http://mmi.tudelft.nl/~catholijn/publications/sites/default/files/Broekens_Factors%20Influencing%202010.pdf
- Caragliu, A., & Del Bo, C. F. (2019). Smart innovative cities: The impact of Smart City policies on urban innovation. *Technological Forecasting and Social Change*, 142, 373-383. <https://doi.org/10.1016/j.techfore.2018.07.022>

- Cardullo, P., & Kitchin, R. (2019). Smart urbanism and smart citizenship: The neoliberal logic of 'citizen-focused' smart cities in Europe. *Environment and Planning C: Politics and Space*, 37(5), 813-830. <https://doi.org/10.1177/0263774X18806508>
- Fietkiewicz, K. J., Lins, E., Baran, K. S., & Stock, W. G. (2016). Inter-generational comparison of social media use: Investigating the online behavior of different generational cohorts. In *2016 49th Hawaii International Conference on System Sciences (HICSS)* (pp. 3829-3838). IEEE. <https://doi.org/10.1109/HICSS.2016.477>
- Giffinger, R., Fertner, C., Kramar, H., & Meijers, E. (2007). City-ranking of European medium-sized cities. *Cent. Reg. Sci. Vienna UT*, 1-12. http://www.smartcity-ranking.eu/download/city_ranking_final.pdf
- Heo, T., Kim, K., Kim, H., Lee, C., Ryu, J. H., Leem, Y. T., Jun, J. A., Pyo, C., Yoo, S., Ko, J. (2014). Escaping from ancient Rome: Applications and challenges for designing smart cities. *Transactions on Emerging Telecommunications Technologies*, 25(1), 109–119. <https://doi.org/10.1002/ett.2787>
- Ilhan, A., Möhlmann, R., & Stock, W. G. (2015). Customer value research and ServQual surveys as methods for information need analysis - The ubiquitous city Songdo as a case study. In F. Pehar, C. Schlögl, & C. Wolff (Eds.), *Re:inventing Information Science in the Networked Society. Proceedings of the 14th International Symposium on Information Science (ISI 2015)* (pp. 457–468). Verlag Werner Hülsbusch.
- Monheim am Rhein (2020). Aktuelle Projekte. *Monheim am Rhein*. Retrieved from <https://www.monheim.de/stadtleben-aktuelles/mitmach-portal/aktuelle-projekte/>
- Neirotti, P., De Marco, A., Cagliano, A. C., Mangano, G., & Scorrano, F. (2014). Current trends in Smart City initiatives: Some stylised facts. *Cities*, 38, 25-36. <https://doi.org/10.1016/j.cities.2013.12.010>
- Schumann, L., & Stock, W. G. (2014). The Information Service Evaluation (ISE) Model. *Webology*, 11(1). <http://www.webology.org/2014/v11n1/a115.pdf>
- Shelton, T., Zook, M., & Wiig, A. (2015). The 'actually existing smart city'. *Cambridge journal of regions, economy and society*, 8(1), 13-25. <https://doi.org/10.1093/cjres/rsu026>
- Tacken, M., Marcellini, F., Mollenkopf, H., Ruoppila, I., & Szeman, Z. (2005). Use and acceptance of new technology by older people. Findings of the international MOBILATE survey: 'Enhancing mobility in later life'. *Gerontechnology*, 3(3), 126-137. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.474.3979&rep=rep1&type=pdf>
- Townsend, A. M. (2013). *Smart cities: Big data, civic hackers, and the quest for a new utopia*. WW Norton & Company Inc.

Cross-Analysis of Researchers' Different Shared File Management Activities in Cloud Storage

Oh, Kyong Eun

Simmons University, USA | kyongeun.oh@simmons.edu

ABSTRACT

Using cloud storage for collaborative projects has become common practice among researchers. However, despite its popularity, we lack an understanding of researchers' shared file management practices in cloud storage. More importantly, although researchers engage in various shared file management activities (i.e., create, update, organize, find/re-find, keep, archive, and delete) in cloud storage, no studies have comparatively investigated these activities. Based on 475 responses from researchers collected through an online survey, this study examined researchers' different shared file management activities by focusing on the differences in the frequency, perceived ease of use, and satisfaction for each activity, and identified well-supported activities as well as activities that require increased support. This study deepens our understanding of researchers' shared file management practice in cloud storage by cross-examining an array of shared file management activities, and contributes to the development of tools and applications that better support researchers' shared file management in cloud storage.

KEYWORDS

Shared information management; cloud storage; information behavior; personal information management

INTRODUCTION

Researchers' use of cloud storage such as Google Drive, Dropbox, or OneDrive for their collaborative projects has been increasing, and even more so in response to the pandemic. However, despite the widespread use of cloud storage for collaborative research projects, little is known about researchers' shared file management practices in cloud storage. Previous studies on file management in cloud storage have examined individuals' personal file management rather than shared file management, and focused on specific aspects such as users' intention of using cloud storage (Ghaffari & Lagzian, 2018; Li et al., 2020) or security issues (Alsmadi & Prybutok, 2018; Widjaja et al., 2019) rather than everyday life file management practices. Some studies investigated how people manage shared files in cloud storage. However, their focus was on a specific activity or aspect such as re-finding shared files (Bergman et al., 2014, 2015, 2019, 2020a, 2020b) or file synchronization (Capra et al., 2014; Marshall & Tang, 2012; Marshall et al., 2012) rather than on an array of activities researchers engage in including creating, updating, organizing, finding/re-finding, keeping, archiving, and deleting shared files.

In order to fill this gap, this study explored the following research questions: (1) To what extent do various shared file management activities differ in terms of their frequency? (2) To what extent do researchers' perceived ease of use for various shared file management activities in cloud storage differ? (3) To what extent do researchers' satisfaction with various shared file management activities in cloud storage differ? and (4) Which activities are well-supported and which activities need increased support?

METHODS

Data Collection and Analysis

Data were collected via an online survey from 475 researchers in the US who all had ongoing collaborative projects that use cloud storage. The survey asks how researchers manage shared files in cloud storage. It was distributed to 220 universities by contacting administrative staff at each university and asking them to share the link to the survey with the researchers at their institutions. Universities were randomly selected from the list of doctoral universities in the United States (Carnegie Classification of Institutions, 2018). Collected data were analyzed by conducting statistical analyses including descriptive and inferential statistics.

Participants

In terms of gender, 60.1% were female ($n=282$), 36.7% were male ($n=172$), and 2.8% were non-binary or preferred not to answer ($n=13$). In terms of age, 42.4% were 20s ($n=199$), 30.1% were 30s ($n=141$), 14.7% were 40s ($n=69$), 8.1% were 50s ($n=38$), and 4.3% were 60 or older ($n=20$). In terms of ethnicity, 69.1% were White ($n=324$), 16.4% were Asian ($n=77$), 6.4% were Latino/Hispanic ($n=30$), 3.0% were Black ($n=14$), 2.3% were multiracial ($n=11$), and 1.2% were American Indian/Alaska Native/Pacific Islander and other ($n=6$). In terms of discipline, 43.5% were from Social Sciences ($n=204$), 39.2% were from Sciences ($n=184$), 16.6% were from Arts & Humanities ($n=78$), and 0.4% were from other disciplines ($n=2$). Among participants, 58.2% were graduate students ($n=273$), 30.0% were faculty members ($n=141$), and 11.3% were postdoctoral researchers, lecturers, or research specialists ($n=53$).

DATA ANALYSIS AND PRELIMINARY RESULTS

Frequency of shared file management activities in cloud storage

Participants were asked to indicate how often they update, find/re-find, create, organize, archive, and delete files of their current project in the shared cloud storage. Among different shared file management activities, updating files was the most frequent activity ($M=4.44$, $SD=1.44$), followed by finding/re-finding files ($M=4.38$, $SD=1.43$), and creating files ($M=4.25$, $SD=1.43$). The least frequent activity was deleting files ($M=2.07$, $SD=1.08$) followed by archiving files ($M=2.76$, $SD=1.65$). See Table 1 for more information. A one-way ANOVA analysis revealed statistically significant differences in frequency across file management activities, $F(5, 2792)=218.49$, $p < .01$, $\eta^2 = .28$.

	<i>M</i>	<i>SD</i>	<i>N</i>
Updating files	4.44	1.44	472
Finding/Re-finding files	4.38	1.43	463
Creating files	4.25	1.43	469
Organizing files	3.31	1.54	468
Archiving files	2.76	1.65	457
Deleting files	2.07	1.08	469

Table 1. Frequency of Shared File Management Activities (Likert scale 1-7, 1=never, 7=always)

Perceived ease of use and satisfaction with shared file management activities in cloud storage

Participants were asked to rate how easy it is to manage files in the shared cloud storage for their current collaborative project for each shared file management activity. Participants perceived creating files ($M=6.29$, $SD=1.20$) as most easy, followed by keeping ($M=6.26$, $SD=1.10$) and updating files ($M=6.14$, $SD=1.27$). The least easy activity was finding/re-finding files ($M=5.33$, $SD=1.54$), followed by archiving ($M=5.39$, $SD=1.46$) and organizing files ($M=5.49$, $SD=1.55$). A one-way ANOVA analysis indicated statistically significant differences in researchers' perceived ease across file management activities, $F(6, 3132)=46.55$, $p < .01$, $\eta^2=.08$.

Participants were also asked to rate how satisfied they were with their file management practices in the shared cloud storage for their current collaborative project for each shared file management activity. Similar to the perceived ease of use, participants found creating files ($M=6.09$, $SD=1.25$) most satisfactory, followed by keeping ($M=5.93$, $SD=1.33$) and updating files ($M=5.80$, $SD=1.40$). Finding/re-finding ($M=5.29$, $SD=1.63$) and archiving files ($M=5.29$, $SD=1.60$) were least satisfactory, followed by organizing files ($M=5.36$, $SD=1.65$). A one-way ANOVA analysis indicated statistically significant differences in satisfaction across file management activities, $F(6, 3081)=22.06$, $p < .01$, $\eta^2=.04$. Table 2 presents more detailed information.

<i>Perceived Ease</i>				<i>Satisfaction</i>			
	<i>M</i>	<i>SD</i>	<i>N</i>		<i>M</i>	<i>SD</i>	<i>N</i>
Creating files	6.29	1.20	468	Creating files	6.09	1.25	460
Keeping files	6.26	1.10	467	Keeping files	5.93	1.33	461
Updating files	6.14	1.27	467	Updating files	5.80	1.40	465
Deleting files	6.12	1.22	426	Deleting files	5.73	1.39	410
Organizing files	5.49	1.55	468	Organizing files	5.36	1.65	464
Archiving files	5.39	1.46	374	Archiving files	5.29	1.60	364
Finding/Re-finding files	5.33	1.54	469	Finding/Re-finding files	5.29	1.63	464

Table 2. Perceived Ease of Use and Satisfaction with Shared File Management Activities (Likert scale 1-7)

DISCUSSION AND CONCLUSION

The results showed that creating and updating shared files were frequent activities and were perceived as easy and satisfactory. However, finding/re-finding shared files was the least easy *and* least satisfactory activity, although it was one of the top three most frequent activities, indicating it's a crucial activity that requires greater support. Archiving and organizing shared files were other activities that were neither easy nor satisfactory, also requiring improvement. The findings suggest that not only finding/re-finding behavior but also archiving and organizing behavior need further attention and improved support. By comparatively analyzing different shared management activities, this study advances our knowledge of researchers' shared file management behaviors in cloud environments, which is a timely and critical area to explore. The findings also have practical implications for developing tools and applications that better support researchers' shared file management in cloud storage.

ACKNOWLEDGMENTS

The author gratefully acknowledges the grant from the Alfred P. Sloan Foundation (#G-2021-16833).

REFERENCES

- Alsmadi, D., & Prybutok, V. (2018). Sharing and storage behaviors via cloud computing: Security and privacy in research and practice. *Computers in Human Behavior*, *85*, 218-226.
- Bergman, O., Elyada, O., Dvir, N., Vaitzman, Y., & Ben Ami, A. (2015). Spotting the latest version of a file with Old'nGray. *Interacting with Computers*, *27*(6), 630-639.
- Bergman, O., Israeli, T., & Whittaker, S. (2020a). Factors hindering shared files retrieval. *Aslib Journal of Information Management*, *72*(1), 130-147.
- Bergman, O., Israeli, T., & Whittaker, S. (2020b). The scalability of different file-sharing methods. *Journal of the Association for Information Science and Technology*, *71*(12), 1424-1438.
- Bergman, O., Whittaker, S., & Falk, N. (2014). Shared files: The retrieval perspective. *Journal of the Association for Information Science and Technology*, *65*(10), 1949-1963.
- Bergman, O., Whittaker, S., & Frishman, Y. (2019). Let's get personal: The little nudge that improves documents retrieval in the Cloud. *Journal of Documentation*, *75*(2), 379-396.
- Capra, R., Vardell, E., & Brennan, K. (2014). File synchronization and sharing: User practices and challenges. In *Proceedings of American Society for Information Science and Technology*, October 2014, Seattle, WA, pp. 1-10.
- Carnegie Classification of Institutions. (2018). *Standard listings*.
https://carnegieclassifications.iu.edu/lookup/standard.php#standard_basic2005_list
- Ghaffari, K., & Lagzian, M. (2018). Exploring users' experiences of using personal cloud storage services: A phenomenological study. *Behaviour & Information Technology*, *37*(3), 295-309.
- Li, Y., Chang, K., & Wang, J. (2020). Self-determination and perceived information control in cloud storage service. *Journal of Computer Information Systems*, *60*(2), 113-123. DOI: 10.1080/08874417.2017.1405294
- Marshall, C., & Tang, J. (2012). That syncing feeling: Early user experiences with the cloud. In *Proceedings of the Designing Interactive Systems Conference*, June 2012, Newcastle, UK, pp. 544-553.
- Marshall, C., Wobber, T., Ramusubramanian, V., & Terry, D. (2012). Supporting research collaboration through bi-level file synchronization. In *Proceedings of the 17th ACM International Conference on Supporting Group Work*, October 2012, Sanibel Island, FL, pp. 165-174.
- Widjaja, A., Chen, J., Sukoco, B., & Ha, Q. (2019). Understanding users' willingness to put their personal information on the personal cloud-based storage applications: An empirical study. *Computers in Human Behavior*, *91*, 167-185.

Impacts of Information Source and Perceived Health on Information Behavior

Park, Min Sook

University of Wisconsin-Milwaukee, U.S.A. | minsook@uwm.edu

Lee, Hyun Seung

University of Wisconsin-Milwaukee, U.S.A. | lee649@uwm.edu

ABSTRACT

This preliminary study revisits a fundamental information problem of information behavior, focusing on needs, overload, and information source use, in the context of the COVID-19 pandemic. The associations between the impact of information source use on the extent of information needs, being exposed to information, and feeling of overload was examined. Furthermore, to understand the impact of context on information behavior, the differences in the degree of information resource use, needs, exposure, and overload between the two groups with different levels of health were investigated.

KEYWORDS

Health information behavior, information needs, information overload, COVID-19

INTRODUCTION

Provision of the right amount of information that meets user needs is a fundamental information problem. The importance of this problem has been reemphasized since the public health crisis that arose from severe acute respiratory syndrome coronavirus 2 disease (hereafter, COVID-19) has significantly impacted the globe. To decide how to respond to the pandemic (e.g., compliance with preventive measures), individuals need information about situations and the virus to make decisions on their health issues such as compliance with preventive measures.

The simple provision of information, however, is not always sufficient. It may or may not meet individuals' needs, be overly excessive, or vary by context. It is a known fact that both scarcity and overload of information are "a hindrance rather than a help (Bawden Holtham, & Courtney, 1999, p. 250)." The availability of a wider range of information sources (e.g., online resources, TV, etc.) than ever before further complicates this fundamental problem throughout society (Bawden & Robinson, 2020). Paradoxically by having more information sources, information users may easily get exposed to excessive information than they need. Furthermore, information needs would vary by group in our society. For example, to navigate a health crisis like the pandemic, in particular, variances in individuals' overall wellness can lead to different information needs and user behavior.

As an initial effort of revisiting this fundamental information problem, this preliminary study investigates if the amount of information that has been provided during the COVID-19 pandemic was associated with the extent of information resource use and the overall health of individuals.

METHOD

Data

In this study, a total of 7,541 responses collected in the United States were used after removing incomplete responses. This is a subset of the COVID-19 Preventive Behavior Survey (Collis et al., 2020) data collected by the Massachusetts Institute of Technology and Facebook's Data for Good from 45 countries between October 2020 and March 2021.

Measures and Data Analysis

Five survey questions were used to measure the extent of information resource use, needs, exposure, and overload of COVID-19 information. *Information needs* was measured using the question: "Which of the following aspects of COVID-19 do you have the most questions about?" We operationalized a continuous integer variable, ranging from 0 to 17, based on the sum of the selection of *yes=1* or *no=0*. *Information exposure* was measured using the question: "In the past week, how much, if anything, have you heard or read about coronavirus (COVID-19)?" with a 3-point scale (*nothing/a little=1, a moderate amount=2, a lot=3*). The degree of *information overload* was measured by the question: "In the past week, did you see more or less than you wanted to see about coronavirus?" with a 3-point scale (*much less/less=1, about the right amount=2, more/much more=3*). The *information source use* was measured by two multiple-choice questions "In the past week, from which of the following, if any, have you received news and information about COVID-19?" and "In the past week, from which of the following, if any, have you received news and information about COVID-19?" We operationalized a continuous integer variable, ranging from 0 to 13, based on the number of selections (e.g., *television, online, local health worker, etc.*). We also considered demographic and socioeconomic characteristics: Gender (*female, male*), age (*younger than 20, 20–30, 31–40, 41–50, 51–60, 61–70, and 71 or older*), education level (*less than secondary school =1, college or university=2, and graduate school=3*) and self-reported overall health (*poor/fair vs. good/very good/excellent*).

85th Annual Meeting of the Association for Information Science & Technology | Oct. 29 – Nov. 1, 2022 | Pittsburgh, PA. Author(s) retain copyright, but ASIS&T receives an exclusive publication license.

Data were analyzed using the Pearson correlation coefficient (r) (Benesty, Chen, Huang, & Cohen, 2009) to understand the relations among the variables. Welch's t-test (Welch, 1947) was also performed to understand if there are differences between two groups who self-rated their overall health as poor/fair or good/better in their information resources use, needs, exposure, and overload.

RESULTS

About 63.9% ($n=4822$) of the sample was female. Approximately half (49.7%, $n=3,747$) were between 40 years old or younger. 46.7% ($n=3,523$) had a college or graduate degree. Weak or moderate correlations were found among the variables. The extent of diverse information source use was found to be positively correlated with information needs ($r=.209, p < 0.01$), information exposure ($r=.291, p < 0.01$) and feeling of overload ($r=.102, p < 0.01$). Information exposure was also positively correlated with overload ($r=.211, p < 0.01$) (see Table 1).

Variables	n(%)	M(SD)	Sources	Needs	Exposure	Overload	Health
Sources	-	5.84(2.45)	1				
Needs	-	5.03(3.40)	0.209***	1			
Exposure							
<i>Nothing/ A little</i>	911(12.08)		0.291***	0.072***	1		
<i>A moderate amount</i>	2,178(28.88)						
<i>A lot</i>	4,452(59.04)						
Overload							
<i>Much less/ Less</i>	892(11.83)		0.102***	-0.0491***	0.211***	1	
<i>About the right amount</i>	4,080(54.10)						
<i>More/Much more</i>	2,569(34.07)						
Self-reported overall health							
<i>Poor or Fair</i>	1,161(15.40)		0.030***	-0.096***	0.008	0.036***	1
<i>Good or Better</i>	6,380(84.60)						

Table 1. Descriptive Statistics of Sample Data and Correlations Between Variables (N=7,541. P** < 0.05, P*** < 0.01)

Welch's t-test results demonstrated that there was a significant effect of health on information use behavior. To illustrate, those who rated their overall health as good or better demonstrated they use significantly more diverse information sources ($M=5.87, SD=6.05$) than those who rated their health as fair or poor ($M=5.66, SD=5.87$), $t(1626)=-2.68, p < .01$. Those who have fair or worse health showed a significantly higher degree of information needs ($M=5.79, SD=13.85$) but experience overload less ($M=2.46, SD=.50$) than those who rated their health as good or better, $t(1514)=7.77, p < .00$; $t(1584)=-.73, p < .00$ (see Table 2).

	Self-reported overall health				<i>df</i>	<i>T</i>	<i>Sig.</i>
	Poor or Fair ($n=1,161$)		Good or Better ($n=6,380$)				
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Sources	5.66	5.87	5.87	6.05	1626	-2.68	.00
Needs	5.79	13.85	4.89	10.99	1514	7.77	.00
Exposure	2.46	0.50	2.47	0.49	1597	-0.73	.00
Overload	2.17	0.44	2.23	0.40	1571	-3.01	.00

Table 2. Welch's T-Test Results Comparing the Extent of Information Source Use, Information Exposure, Needs, and Feeling of Overload (N=7,541)

CONCLUSION

The results demonstrate that the extent of information source use is associated with the users' information needs, and the extent of getting exposed to information. It was also found health can yield significant differences in the amount of information needed, the number of information sources used, and experiences of having overly excessive information (i.e., overload). The result of this preliminary study restresses the importance of context and the right amount in providing information to successfully navigate crises such as a pandemic. The results also imply that information professionals need to actively engage in providing tailored information to meet different needs among groups to effectively respond to crises in our society.

ACKNOWLEDGMENTS

The study was approved under MIT's Committee on the Use of Humans as Experimental Subjects (IRB) protocols E-2578 and E- 2859.

REFERENCES

Bawden, D., Holtham, C., & Courtney, N. (1999). *Perspectives on information overload*. 51, 249–255.

- Bawden, D., & Robinson, L. (2020). Information overload: An overview. In *Oxford encyclopedia of political decision making* (pp. 1–60). Oxford University Press.
- Benesty, J., Chen, J., Huang, Y., & Cohen, I. (2009). Pearson correlation coefficient. In *Noise Reduction in Speech Processing* (pp. 37–40). Springer.
- Collis, A., Garimella, K., Moehring, A., Rahimian, M., Babalola, S., Gobat, N., Shattuck, D., Stolow, J., Eckles, D., & Aral, S. (2020). *Global survey on COVID-19 beliefs, behaviors and norms*. Technical report, MIT Sloan School of Management.
- Welch, B. L. (1947). The generalization of 'student's' problem when several different population variances are involved. *Biometrika*, 34(1-2), 28–35.

Seeing Resilience in Students: Information Literacy as a Social Practice for Academic Instruction Librarians

Rath, Logan

SUNY Brockport | lrath@brockport.edu

ABSTRACT

This poster will describe a dissertation project that consisted of a questionnaire and interview phase followed by a focus group and semi structured diary phase. This project investigated the theory-to-practice gap in seeing information literacy as a socially constructed practice among academic instruction librarians. Phase one consisted of 17 interviews that were coded and analyzed using grounded theory approaches. The second phase of the research introduced the resulting grounded theory to new participants. The participants then took part in two focus groups, one at the start of the semester and one at the end of the fall 2021 semester. Between the two focus groups participants completed semi structured diary entries related to their instruction sessions. Grounded theory approaches were used to modify the original theory into the current iteration.

KEYWORDS

Information literacy, information landscapes, practice theory, grounded theory, threshold concepts

INTRODUCTION

The field of information literacy (IL) suffers from a gap between theory and practice, meaning that librarians, as practitioners, do not adequately use and apply the theory of information science in their work (Julien et al., 2013; Nguyen & Hider, 2018). Information literacy is defined by the Association of College and Research Libraries (2016) as “the set of integrated abilities encompassing the reflective discovery of information, the understanding of how information is produced and valued, and the use of information in creating new knowledge and participating ethically in communities of learning” (p. 8). Concurrently, Lloyd (2017) defines IL as “a way of knowing” (p. 94). Both definitions acknowledge that IL is theorized as a social practice. This has been the case in scholarly thinking for almost 30 years (Hjørland & Albrechtsen, 1995). However, practitioners appear to focus on skills and not practices (Graves et al., 2021). Without demonstrating an understanding of, and integrating, theory, practitioners run the risk of explaining IL concepts in outdated ways or using outdated or ineffective methods when providing IL instruction. While many studies have examined the nature of IL, none have examined the theory-to-practice gap in understanding information literacy as a social practice among academic instruction librarians.

This study took place in two phases. The first phase asked the following research question: What are academic instruction librarians’ understandings of information literacy as a set of social practices? Grounded theory analysis resulted in the discovery that seeing information literacy as a social practice is a threshold concept for academic instruction librarians. A threshold concept transforms the learner so that they must find a new way forward as a result of shifting their understanding of a given concept (Meyer et al., 2010). The second phase of this study attempted to refine that finding with new participants in a new context, asking: “To what extent, if any, does the proposed threshold concept describe librarians’ beliefs about information literacy as a social practice?”

CONCEPTUAL FRAMEWORK

The conceptual framework for this study consisted of information literacy landscapes (Lloyd, 2010), as well as Schatzki’s (2001) practice theory. Information landscapes emphasizes how students build their knowledge of different features of the information environment. Practice theory examines how human activity plays a role in the creation, change, and maintenance practices. Together these focus on the valued ways of knowing that information literacy practices are performed. The frameworks were used to develop the original grounded theory in phase one. Understanding IL as a social practice requires a grounding in practice theory and by extension aligning one’s view with Lloyd’s definition.

METHODS

In the first phase of this study, 17 academic librarians who have responsibility for instruction from community colleges, comprehensive colleges and doctoral granting institutions took part in semi structured interviews over Zoom lasting approximately one hour in length. The recordings were transcribed and then open coded (Bryant & Charmaz, 2019). After open coding, codes were collapsed into axial codes which lead to the discovery of categories and the relationship between those categories (Miles et al., 2020). Finally, selective coding (Saldaña, 2016) was completed to identify key quotes and other data necessary to confirm the chosen categories. Member-checking was performed by presenting each participant with an analysis of their interview as it related to the overall grounded theory. Participants then had a chance to provide corrections and confirm the researcher’s analysis. All participants agreed with the analysis and threshold concept.

In the second phase of the study, nine new participants who were all also academic librarians, agreed to take part in focus groups and complete semi structured diary entries during the fall 2021 semester. Seven participants finished all study activities. Focus groups were used for their ability to bring about greater findings than individual interviews alone (Krueger & Casey, 2015). Semi structured solicited diaries were used for their ability to approximate participation observation while allowing for the participant to answer additional reflective questions that would not have been possible during participant observation (Bartlett & Milligan, 2015). The first focus group introduced a simplified version of the grounded theory. Each participant then completed a weekly semi structured diary entry responding to specific prompts related to the ways in which the threshold concept fit or was in need of revision. Member checking in this phase of the study took place during the second focus group. Consensus was found in each focus group and between focus groups with regard to the revisions to the grounded theory.

FINDINGS

The first phase of this study discovered that seeing information literacy as a social practice is a threshold concept for academic instruction librarians. It also found that there were external tensions that forced librarians to teach from a skills-based lens even though they did not agree with it. In the second phase of the research, the roles of skills, practices, the threshold, and the force of tensions were supported. There was one revision to the threshold concept showing that a counteracting force of supports that allowed participants to resist the tensions that exist.

DISCUSSION

The use of a threshold concept worked to define the shift in librarian's thinking because it allowed participants to see IL in a different way and required that librarians change their pedagogical practices. Librarians who had crossed the threshold also shunned a deficit perspective of students. Instead of seeing students as lacking, participants saw students as bringing several strengths to their interactions, as well as attempting to learn needed practices based on a new context. This shift in conceptualization of IL resulted in richer conversations with students and a sense of professional fulfillment. Students were seen as more resilient under this approach. Participants in the second phase of the study enjoyed the in-depth time to reflect on their own teaching practices and appreciated their participation as a form of professional development.

IMPLICATIONS

This study has implications for LIS faculty, library administrators, and practitioners. LIS faculty are encouraged to include discussion of this research in their IL courses as a way to prepare pre-service librarians to recognize differences in conceptualization of information literacy among competing discourses. Library administrators can use this study to help their staffs move forward in a sustainable manner related to their library's IL program. Lastly practitioners can benefit from the outcomes a practice-based approach to IL.

CONCLUSION

This study adequately used multiple methods to create and refine a grounded theory stating that seeing information literacy as a social practice is a threshold concept for academic instruction librarians. Limitations of the study include a relatively small number of participants in the second phase, the impacts of COVID-19 on participant availability and desire to participate in a study, and a lack of greater participant diversity. Future research can be undertaken with additional groups of librarians, especially those who have completed their graduate degrees more recently.

REFERENCES

- Association of College and Research Libraries. (2016). *Framework for information literacy for higher education*. <http://www.ala.org/acrl/standards/ilframework>
- Bartlett, R., & Milligan, C. (2015). *What is diary method?* Bloomsbury Publishing.
- Bryant, A., & Charmaz, K. (Eds.). (2019). *The SAGE handbook of grounded theory* (2nd ed.). SAGE.
- Graves, S. J., LeMire, S., & Anders, K. C. (2021). Uncovering the information literacy skills of first-generation and provisionally admitted students. *The Journal of Academic Librarianship*, 47(1), 102260. <https://doi.org/10.1016/j.acalib.2020.102260>
- Hjørland, B., & Albrechtsen, H. (1995). Toward a new horizon in information science: Domain-analysis. *Journal of the American Society for Information Science*, 46(6), 400–425. [https://doi.org/10.1002/\(SICI\)1097-4571\(199507\)46:6<400::AID-AS12>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1097-4571(199507)46:6<400::AID-AS12>3.0.CO;2-Y)
- Julien, H., Tan, M., & Merillat, S. (2013). Instruction for information literacy in canadian academic libraries: A longitudinal analysis of aims, methods, and success. *Canadian Journal of Information and Library Science*, 37(2), 81–102. <https://doi.org/10.1353/ils.2013.0007>
- Krueger, R. A., & Casey, M. A. (2015). *Focus groups: A practical guide for applied research* (5th ed.). Sage.
- Lloyd, A. (2010). *Information literacy landscapes: Information literacy in education, workplace and everyday contexts*. Chandos.
- Lloyd, A. (2017). Information literacy and literacies of information: A mid-range theory and model. *Journal of Information Literacy*, 11(1), 91–105. <https://doi.org/10.11645/11.1.2185>
- Meyer, J. H. F., Land, R., & Baillie, C. (Eds.). (2010). *Threshold concepts and transformational learning*. Sense.

- Miles, M. B., Huberman, A. M., & Saldaña, J. (2020). *Qualitative data analysis: A methods sourcebook* (4th ed.). SAGE.
- Nguyen, L. C., & Hider, P. (2018). Narrowing the gap between LIS research and practice in Australia. *Journal of the Australian Library and Information Association*, 67(1), 3–19.
- Saldaña, J. (2016). *The coding manual for qualitative researchers* (3rd ed.). SAGE.
- Schatzki, T. R. (2001). Introduction: Practice theory. In T. R. Schatzki, K. K. Cetina, & E. von Savigny (Eds.), *The practice turn in contemporary theory*. Taylor & Francis Group. <http://ebookcentral.proquest.com/lib/buffalo/detail.action?docID=235322>

Give or Take?: Conceptualizations of Data and Its Creation in Information Science

Ridenour, Laura

University of Missouri, USA | lerbhn@missouri.edu

Moulaison-Sandy, Heather

University of Missouri, USA | moulaisonhe@missouri.edu

ABSTRACT

In information science, what data itself *is*, how it is defined, and the distinctions between data and information are unclear, yet *data* is a foundational construct at the heart of the research enterprise. This poster probes the question of how data is conceptualized in information science through an analysis of works by Jonathan Furner, Birger Hjørland, and Michael Buckland, who have contributed substantially to this critical issue. It finds that notions of the nature of data vary widely, and are often contradictory. Complexities are identified in the approach to data as both documents and as records. In the process of this analysis, it identifies the essential question of whether [research] data is *given* or *taken*: Is data found/identified by the researcher, or is it created? Arguments for both approaches are fleshed out, providing a basis for further work in information science in this essential space.

KEYWORDS

Data; nature of data; data creation; science of science; research data

INTRODUCTION

Data, documents, and records exist in multitudes of types, formats, and spaces, and are used extensively in the creation of new knowledge. Calls in the field of information science to clearly define data exist (Borgman, 2010; Furner, 2016; Hjørland, 2019), yet a fundamental agreement on what data itself *is* lacks, and the distinctions between data and information is far from resolved. Data is an encoded representation of the nature of the phenomenon under examination, and as such, is an abstraction removed from an instance or series of instances. In this work-in-progress poster, we explore tensions in descriptions of data, datasets, documents, and records from the perspectives of Furner (2016), Hjørland (2019), and Buckland (2018). How are data conceptualized and created? Humanistic perspectives on the nature of data as a phenomenon as discussed by Drucker (2011) and Thorpe (2021) leads to the ultimate question this project considers: In light of what we know about data, is data *given* or is it *taken*?

DATA AND DATASETS AS DOCUMENTS AND RECORDS

Historical interpretations of the term “data” and their respective relationships to “documents” vary in proximity and the precision of the relationship between the two terms (Furner, 2016). In this work, Furner examines nine historical kinds of data, presenting a logic-based argument concluding that a computational model of data and documents is incorrect, incoherent, and useless; instead, Furner recommends an informational interpretation of data and documents as it affords the ability to name frequently used concepts. To interest in our discussion, Furner’s claim is that “document” is primary and comes before “data.” Furner emphasizes that a document is a type of dataset, as it is a carefully arranged aggregation of data.

Hjørland (2019) contrasts definitions of “data” and “documents,” claiming that the two concepts are interrelated, yet discrete, referencing several sources that (according to Hjørland) do not sufficiently distinguish between the two (citing the Consultative Committee for Space Data Systems, 2002; Borgman, 2007; Bugaje & Chowdhury, 2015). According to Hjørland, records yield data, but are not themselves data, as records are a form of document.

Buckland (2018) identifies documents as being related to “data, facts, texts, works, information, knowledge, signs, and other documents,” (p. 1) and having tangible (“physical”), cognitive, and social aspects. Documents contain rich potential for the long-term preservation of records, but ultimately, what is called a document is a pragmatic decision. Buckland argues that the contrast between data and documents is a matter of form, as texts are often considered to be documents, and numerical representations are called data.

From these views, the relationships between data, datasets, documents, and records are not only unclear, they simultaneously share perspectives while clashing. Table 1 illustrates the complexity of the approaches at a high level, in the final poster, the table will include more detailed considerations for each author’s views of documents, data, and records in a graphic.

Source	Document	Data	Data vs. document	Records
Furner, 2016	Document is primary, datasets are a type of document	Data is in datasets as “numeric attribute-values”	Document is primary, preceding data	Each record is a document
Hjørland, 2019	Evidence in support of a fact (from Briet)	Data are produced for a purpose p.699	Interrelated but discrete concepts	Records yield data, but are documents
Buckland, 2018	Physical, often textual, affords preservation, asserts a perspective	Often numeric, fragments from within one or more documents	Distinction between data and document are an issue of format	Individual entities (text or object) that are represented within documents

Table 1. Approaches to Data as Documents and as Records

THE GIVE AND THE TAKE OF DATA CREATION

Drucker (2011) introduces the concept of data as “capta,” where instead of being “given,” as implied by the Latin root of the word, it is actively taken from observations. Taking data imbues it with the context of the individual or group’s purpose, rendering it subject to interpretation through the lens through which it was constructed. Only the pieces of interest in relation to a given observed phenomenon are recorded. This notion of taking what is of interest is reflected by Thorpe (2021), who states that “...data is not found, it is constructed” (p. 50). Similarly, documents are constructed to reflect the nature of curated records. The construction of data, in turn, impacts the world, “that in making data we change the systems from whence it came” (Thorpe, 2021, p. 50). Applying the notion of actively taking and constructing data, applying it to the perspectives on documents, records, and data as stored in their information ecosystems. Table 2 presents the competing notions of giving and taking data in the document/records approach, which will be expanded to include identified give/take dichotomy for each subtype of document and data discussed in the three selected pieces for analysis. Furner’s include a historical overview of data as 1) gifts, 2) metadata, 3) gifts of God, 4) geometric premises, 5) mathematic premises, 6) evidence, 7) attribute-values, 8) bits, 9) differences; and logical permutations of documents, document-sets, data, datasets, and documents of documents. Hjørland’s include his discussions of the definitions for data, documents, and records supplied by various authors; and contexts of big data, metadata, standards, and database semantics. Buckland’s include contrasts between documents and data, facts, texts, works, information, knowledge, signs, and inter-document relationships; functions of documents; evidence and experience; and an overview of the approaches of Otlet, Briet, Lund, and Frohmann. This will be presented in clearly outlined tables with a brief discussion of identified implications; graphical mappings illustrating interrelated views; and diagrams of relationships between documents, data, records, and other related concepts as discussed in each of the three identified works.

	Document	Data
Furner	<i>Taken</i> in that they are curated.	<i>Given</i> , then taken in the form of documents. Historical examples discussed vary between given and taken.
Hjørland	Evidence, is <i>taken</i> to illustrate a case or make an argument.	<i>Taken</i> , in some cases, <i>given</i> , in others.
Buckland	<i>Given</i> , in that documents are tools for learning, which is <i>taken</i> from documents.	<i>Given</i> , in that they are represented in documents.

Table 2. Giving and Taking in Documents and Data

CONCLUSION

This poster presents a preliminary analysis of contrasting conceptualizations of data, focusing on the literature in information science and the approaches that consider data to be evidentiary presented here. Implications of this line of inquiry are also explored; they include developing understanding regarding the nature of data, irrespective of disciplinary perspective, as reflected in the discourse. Theoretical implications include bettering our understanding of the rich, underexplored dimension of the undercurrents of science (under the penumbra of science of science) using the simple dichotomy of give and take. Practically, the give/take framework has potential to illuminate decisions made during data collection and analysis through the examination of a dataset as its own entity, giving context to datasets irrespective of the provenance provided by authors. Ultimately, this work is a first step in a larger project to investigate dichotomous approaches to considering data, and it aligns with information science’s quest to better understand the fundamental nature of meaning, especially in data-intensive environments lacking inherent context.

ACKNOWLEDGMENTS

The authors would like to thank Devon Whetstone for initial data collection and analysis.

REFERENCES

- Borgman, C. L. (2010). Research data: Who will share what, with whom, when, and why? *Proceedings of the Fifth China-North American Library Conference 2010*, September 2010, Beijing, China, pp. 1-21.
- Buckland, M. (2018). Document theory. *Knowledge Organization*, 45(5), 425-436.
- Drucker, J. (2011). Humanities Approaches to Graphical Display. *Digital Humanities Quarterly*, 5(1), 1-21.
- Furner, J. (2016). "Data": The data. In M. Kelly, & J. Bielby (Eds.), *Information cultures in the digital age: A festschrift in honor of Rafael Capurro* (pp. 287-306). Springer.
- Hjørland, B. (2019). Data (with big data and database semantics). *Knowledge Organization*, 45(8), 685-708.
- Thorpe, J. (2021). *Living in data: A citizen's guide to a better information future*. MCD, Farrar, Strauss and Giroux.

A User Study in a Pandemic: Some Lessons Learned

Rogers, Kristine M. University of Maryland, College Park, USA | krogers@umd.edu
Trubey, Emily Florida, USA | emilynhr@gmail.com
Oard, Douglas W. University of Maryland, College Park, USA | oard@umd.edu

ABSTRACT

During the COVID-19 pandemic, some research that otherwise would have been conducted in person pivoted to online platforms. This poster paper describes lessons learned from an online study of information behavior by individuals with long-term information needs, focusing on what was learned about how to conduct such a study online. Broadly, three themes are evident: (1) Trust mechanisms were weaker than would be expected for an in-person study, resulting in greater coordination difficulties; (2) What seemed to be a fair reimbursement rate appears to have provided an outsized incentive for fraud; and (3) Zoom proved to be sufficiently capable as a platform for remote use of software for a study that had not originally been designed with remote use in mind.

KEYWORDS

User study; virtual study; qualitative research; research issues

INTRODUCTION

While online studies have been conducted for many years (Gould and Lewis, 1985; Moreno et al., 2013), the COVID-19 pandemic led researchers to pivot to remote studies in cases that may otherwise have been performed in-person (Falter et al., 2022; Larsen et al., 2021; Torrentira, 2020). Prior to the pandemic we had planned an in-person study for formative evaluation of techniques supporting long-term information seeking behaviors; this shifted to an online format. We had developed our in-person instrumented search software using a laptop, and chose to run the software on that laptop for the remote study. We shared our screen and provided keyboard and mouse control to the participant via Zoom. Each session (90-140 minutes) was with one participant, the researcher (first author of this paper), and an assistant. Sessions were recorded for transcription and coding (Glaser & Strauss, 1967; Saunders et al., 2018); the assistant took notes, which were also subsequently coded for triangulation. After completing the study, participants received an Amazon gift card. Of more than 200 initial responses to the Qualtrics selection questionnaire, we scheduled sessions with 27 participants, 15 of which yielded useful data. Our study revealed a number of challenges; here we focus on those stemming from moving the study online.

RECRUITING PARTICIPANTS

Enrolling a sufficient number of participants is a difficult challenge for any study. Because of the topics selected for the study, we were seeking adults located in the U.S. We faced a number of challenges in getting enough responses.

Advertising: We emailed our announcement to professors and researchers across the U.S. in the fields of journalism, information and library sciences, and related disciplines. We posted about the study to research-related groups and hashtags on Facebook and Twitter. Although the posts went to a large number of people, they likely were viewed by many outside the targeted study population; this turned out to be the first link in a chain that led to fraudulent participation. We had chosen to reward participants (adults in the U.S.) at what we thought to be a fair rate (\$25 for approximately 90 minutes); that rate seems to have been high enough to provide an incentive for fraud, at least in some regions. We learned later that our invitation email messages sometimes went to a spam folder—possibly because we sent standardized, formally-worded messages to multiple recipients. For both reasons, we might consider a more focused recruiting effort in future studies, relying more heavily, for example, on snowball sampling.

Fraudulent Qualification Responses: The first step was a qualification questionnaire, implemented as a Qualtrics survey. We identified some fraudulent responses by observing multiple submissions with similar details that were all sent 2-5 minutes apart, at fairly regular intervals—some of which originated from the same IP address. When IP addresses differed in such sequences, we suspect the use of a Virtual Private Network (VPN) to mask the true IP address. Some such respondents attempted to appear legitimate by creating multiple email addresses and usernames. Responses with these characteristics originated from locations that appeared to be in the U.S., China, India, or Kenya, based on Qualtrics metadata. Survey fraud has been noted by others as well, particularly during the pandemic (Brazhkin, 2020; Singh & Sagar, 2021; Storozuk et al., 2020; Zhang et al., 2022). Qualtrics metadata helped us identify some problematic responses, but much of our review for fraud was manual. We recommend using different survey links for each advertising approach so that if fraud is detected in one response, others advertised in the same way can be more closely examined.

CHALLENGES WHILE RUNNING THE STUDY

We initially selected participants to maximize demographic diversity for each of the pre-defined long-term interest types in our study (e.g., sports or health), but exogenous factors later unbalanced those samples. Participants were

85th Annual Meeting of the Association for Information Science & Technology | Oct. 29 – Nov. 1, 2022 | Pittsburgh, PA. Author(s) retain copyright, but ASIS&T receives an exclusive publication license.

lost to both expected factors (e.g., nonresponsive people or incomplete sessions) and from a surprising number of participation attempts from people outside the Institutional Review Board (IRB)-approved demographic (adults in the U.S.). This demographic was selected because information seeking behaviors may vary in other countries.

Caught in a Lie: Some fraudulent participation was detected only after a session began. One participant connected using a mobile phone with a U.S. IP address, but when we asked them to move to a laptop (which the qualification questionnaire had stated was needed for the study), their laptop IP was geolocated to Africa. In another case, a participant insisted that they were connected from a computer, but according to metadata, they were using a mobile phone. This precluded their ability to control the mouse, and thus prevented their participation in the study. We also saw some implausible inconsistencies between the demographic data collected with the qualification questionnaire and from responses during the actual session. We interpreted this as indicating that the participant may have provided false data on the qualification questionnaire, and then a few weeks later did not remember what they had originally submitted. IRB guidance specifically discouraged us from repeating questions during the study that we had asked during qualification. While minimizing repeating questions is excellent advice in a face-to-face setting (to minimize burden on participants), in an online study we recommend the practice as a fraud detection measure.

Repeated Participation: In one case, the same individual attempted to participate three times, having signed up under multiple names and email addresses, having provided different enough data on interests and demographic details that we selected them three times! One way of dissuading or detecting such cases would be to require that participants share video using their camera. We did not require this because we did not want to unnecessarily exclude some participants for technical reasons—such as bandwidth or lack of a camera. Our pattern-based fraud detection had failed in this case (and in the two cases described above), so we looked more closely to see if there were other cues that we had missed during qualification. We noted no instances of fraud for people using organizational email addresses (e.g., university email accounts), although those accounted only a minority of the respondents that we selected. Some email accounts using services such as Gmail or Yahoo Mail displayed additional information that suggested actual personal use (e.g., a profile photo or a chat link). Limiting participants to those whose email accounts have such markers of veracity would likely have excluded many qualified respondents as well; however, the absence of such markers might suggest the need for additional vetting.

Zoom Limitations: Fraudulent participation caused a good deal of wasted effort, but the majority of our participants were indeed real people who had seriously intended to participate in the study. Some of these authentic participants had issues connecting to the Zoom call or lacked experience using Zoom features. One factor that we had not anticipated was that the small Zoom video window obscured portions of our interface as seen on the participant's screen. We offered brief instruction on moving that window. In the future, we would design our screen layout to include an unused place where that window can be parked. Our choice to run our instrumented prototype locally on our laptop, sharing keyboard and mouse control with participants so that they could manipulate the system themselves, generally worked well. Latency was noticeable at times, but tolerable. Some participants did comment on this in the semi-structured interview at the end of their session. If we were to do this again we would want to test the system under a broad range of latency conditions (e.g., by using distant VPN servers) prior to the main study.

Research Team: We had only one system malfunction. Near the end of a session, the laptop where we were running the prototype froze, requiring a reboot. We had planned for such an eventuality, and had designated the assistant as a Zoom co-host. The assistant kept the participant informed as we worked through this brief delay. We originally included the assistant simply as a note-taker, but there are a number of ways in which the assistant was helpful. In another session, we forgot to start the recording, resulting in the inability to use the results from that session. Had we thought of it in advance, the assistant could have ensured that recordings were started for all sessions.

Other Issues: We also encountered the usual problems of scheduling difficulties and no-shows, though in somewhat greater numbers than would have been expected for an in-person study. After selection, we scheduled sessions using YouCanBook.Me; some selected participants did not respond to our scheduling email. Some invited participants may have been concerned that the unfamiliar domain was a phishing attempt. We might avoid some of these problems by using personal messages to schedule sessions, which is practical for a study of this size. Despite the scheduling site sending out an invitation plus an email reminder before the session, and our personal email reminder 12-24 hours in advance, some people who had signed up did not show up for their session. Our study was conducted in Dec. 2021 and Jan. 2022, a period when holiday schedules may have resulted in changing availability.

CONCLUSION

The COVID-19 pandemic led to innovations in study design that may have continued utility for similar studies. Online studies with remote participants will likely continue as a preferred option in some cases (e.g., to reach populations not available locally). We found that Zoom provided a useful platform for such a study. Our goal in this paper has been to share our experience, and to identify options to reduce issues similar to those we encountered.

REFERENCES

- Brazhkin, V. (2020). "I have just returned from the moon:" Online survey fraud. *Supply Chain Management: An International Journal*, 25(4), 489–503. <https://doi.org/10.1108/SCM-12-2019-0466>
- Falter, M., Arenas, A. A., Maples, G. W., Smith, C. T., Lamb, L. J., Anderson, M. G., Uzzell, E. M., Jacobs, L. E., Cason, X. L., Griffis, T. A. N., Polzin, M., & Wafa, N. Z. (2022). Making Room for Zoom in Focus Group Methods: Opportunities and Challenges for Novice Researchers (During and Beyond COVID-19). *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 23(1), Article 1. <https://doi.org/10.17169/fqs-23.1.3768>
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Aldine Publishing. <http://lib.myilibrary.com/browse/open.asp?id=620109&entityid=https://idp.brunel.ac.uk/entity>
- Gould, J. D., & Lewis, C. (1985). Designing for usability: Key principles and what designers think. *Communications of the ACM*, 28(3), 300–311. <https://doi.org/10.1145/3166.3170>
- Larsen, L. B., Øvad, T., Nielsen, L., & Larusdottir, M. (2021). Remote User Testing: Experiences and Trends. *Human-Computer Interaction – INTERACT 2021*, V, 579–583. https://doi.org/10.1007/978-3-030-85607-6_81
- Moreno, A. M., Seffah, A., Capilla, R., & Sánchez-Segura, M.-I. (2013). HCI Practices for Building Usable Software. *Computer*, 46(4), 100–102. <https://doi.org/10.1109/MC.2013.133>
- Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H., & Jinks, C. (2018). Saturation in qualitative research: Exploring its conceptualization and operationalization. *Quality & Quantity*, 52(4), 1893–1907. <https://doi.org/10.1007/s11135-017-0574-8>
- Singh, S., & Sagar, R. (2021). A critical look at online survey or questionnaire-based research studies during COVID-19. *Asian Journal of Psychiatry*, 65, 102850. <https://doi.org/10.1016/j.ajp.2021.102850>
- Storozuk, A., Ashley, M., Delage, V., & Maloney, E. (2020). Got Bots? Practical Recommendations to Protect Online Survey Data from Bot Attacks. *The Quantitative Methods for Psychology*, 16, 472–481. <https://doi.org/10.20982/tqmp.16.5.p472>
- Zhang, Z., Zhu, S., Mink, J., Xiong, A., Song, L., & Wang, G. (2022). Beyond Bot Detection: Combating Fraudulent Online Survey Takers. *Proceedings of the ACM Web Conference 2022*, 699–709. <https://doi.org/10.1145/3485447.3512230>

Self-disclosure of Mentally Ill individuals in Private Facebook Groups: The Means-End Chain Model

Rothschild, Nava
Aharony, Noa

Bar-Ilan University, Israel | rotnava@gmail.com
Bar-Ilan University, Israel | noa.aharony@biu.ac.il

ABSTRACT

This study focuses on mentally ill individuals' self-disclosure in Facebook private groups. The research corpus consists of 30 semi-structured personal interviews. The study was conducted according to a qualitative paradigm and the means-end approach. The aim of the study was to examine which features of Facebook groups, and which personal values encourage the mentally ill to reveal their personal story in private groups. Results indicate that a sense of personal security in the group, self-fulfillment and, most importantly a sense of belonging to the group are the three main values that contribute to mentally ill individuals' choice to engage in self-disclosure on Facebook private groups. This research can assist developers to design better tools that will suit the unique needs of mentally ill people and provide them an appropriate platform where they can express their voice.

KEYWORDS

Social Media; Facebook groups; Self-disclosure; Mental illness; Means-End Chain

INTRODUCTION

Mentally ill people use social networks to communicate with others who share similar struggles (Lewis & Arbuthnott, 2014) in order to talk about their illness, ask for advice, learn from others' experiences, and receive and give support (Manikonda & De Choudhury, 2017). So far, only a few studies have examined the extent and nature of mentally ill individuals' disclosure on social media and especially in Facebook private groups, mainly because of ethical issues. The current study examines the extent to which mentally ill people's self-disclosure in Facebook private groups is a result of the attributes of the platform, the social and community advantages resulting from using these features, and the fulfillment of personal values that can be achieved through self-disclosure in the group.

Researchers used the means-end chain approach (MECA) that Gutman (1982) developed, based on expectancy-value theory, to understand individuals' decision-making processes. The underlying aim of Gutman's (1982) theory is to provide a cognitive structure that connects individuals' values to their behavior (Reynolds & Gutman, 1988). MECA focuses on a product's or service's meanings at three levels of abstraction: attributions, consequences, and values. Attributes ("means") refer to a product's or service's physical or observable properties, consequences are the benefits attained by the attributes, and values ("ends"), imply highly abstract motivation that guides usage behavior (Klenosky, 2002). An attribute-consequence-value chain is usually expressed by a hierarchical map that consists of nodes (i.e., attributes, consequences, and values) and the relationships among them.

A means-end chain analysis typically depends on a laddering interview technique, which was often used in research that tried to understand consumers' preferences towards products or services (e.g., Klenosky, 2002; Reynolds & Rochon, 2001). A laddering procedure typically includes three questions: the attribute question (What attribute makes the product (or service) attractive to you?), the consequence question (Why is the attribute important or desirable to you?), and the value question (Why is your response important?). All responses are coded, and then a hierarchical map is produced:

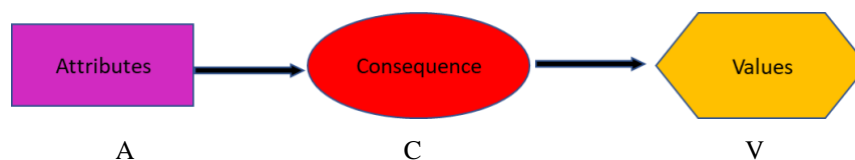


Figure 1. Laddering Procedure

For example, 'size' as an attribute can lead to 'convenient' as a consequence, which can then lead to 'satisfaction' as a value (Al-Abdullatifl & Aladsani, 2021). In the context of the present study, the aim of MECA is to understand how individuals with mental illness perceive their use of Facebook private groups for the purpose of self-disclosure. The product attributes are the group's app on Facebook that encourage the interviewees to disclose themselves.

RESEARCH CORPUS AND RESEARCH PROCESS

Data Collection and Analysis

Researchers conducted 30 semi-structured, in-depth, one-to-one laddering interviews with participants in Facebook private groups for those with mental illness. The interviews were performed according to the guidelines of the

University Ethics Committee and strict adherence to the anonymity of the interviewees. This phase (the interviews) created meaningful A-C-V sequences.

The analysis consisted of three stages: coding, constructing a summary implication matrix (SIM), and generating a hierarchical values map (HVM). The coding procedure resulted in 323 ladders containing 40 detailed codes that are presented in the data, and composed of 10 attribute codes, 21 consequences, and 9 values.

The next step was to take the information from the coding ladders and insert it into a SIM that presents the connections among the coding units. In the third step, the SIM was converted into an HVM, that presented all important chains in order to facilitate interpretation (Leão & Mello, 2007)

RESULTS AND DISCUSSION

The current poster focuses on two chains resulting from the HVM.

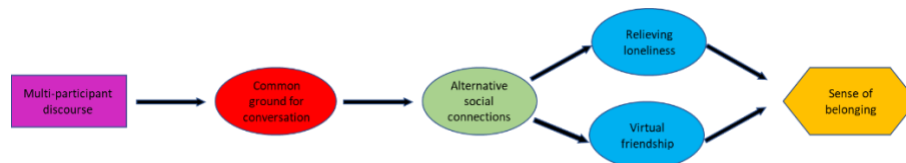


Figure 2. The Means-End Chain – Value Sense of belonging

The chain above highlights the importance of group belonging value and its contribution to self-disclosure (Luo & Hancock, 2020). Participation in private Facebook groups for those with mental illness creates a safe zone that encourages dialogue with others who share similar problems, at a high level of openness. This was already reported by Naslund et al. (2019), who focused on the importance of social media for mentally ill individuals. The discourse in the group reflects the understanding that there are others who face mental difficulties, and it lowers individuals' level of loneliness, as they feel part of a large group. Therefore, in this way, the mentally ill succeed in creating social relationships with others.

Generally, this relationship exists only in the virtual world but occasionally they become friends in the physical world as well (Perry et al., 2018). Exploring the discourse reveals that members consider the group as a family, a place where they have no secrets, and can share their life experiences and receive support.

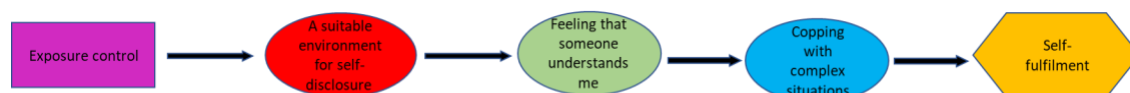


Figure 3. The Means-End Chain – Value self-fulfillment

The second chain focuses on self-fulfillment as a leading value and its contribution to encouraging self-disclosure in the groups. The sense of self-fulfillment is created as a result of receiving and giving help. The sense of control over the disclosure consists of two significant attributes of the group. The first is created because the group is private, and people can participate only after being approved by the group manager. The second is the separation between the time of writing the post and its publication time in the group (Meredith, 2019). These two factors reduce the risk of disclosure and encourage self-expression. As Smith et al. (2021) mentioned, social media technology can be both a deterrent and a boost to the sense of social well-being, depending on the nature of its use.

Our research revealed that individuals who are exposed to the messages have a similar life story and therefore their responses will usually contain empathetic messages and solidarity. Hence, the discourse may improve “support giver” and “receiver” self-esteem. This follows Burrow and Rainone’s (2017) who claimed that feedback such as ‘I like’ in Facebook may enhance advertiser’s sense of self-worth.

LIMITATIONS

Researchers were able to interview only people who participated in the discourse but didn’t interview lurkers. Therefore, the results of this study cannot be generalized to those who chose to participate only as lurkers.

Further, participants in the study were people who, according to their statement, suffer from a mental illness. However, it was impossible to verify that they have a valid medical diagnosis.

REFERENCES

Al-Abdullatif, A. M., & Aladsani, H. K. (2021). Understanding Instructors’ Cognitive Structure Toward the Academic Use of Social Network Sites: The Means–End Chain Theory. *SAGE Open*, *11*(3), 1-17.

- Burrow, A. L., & Rainone, N. (2017). How many likes did I get? Purpose moderate's links between positive social media feedback and self-esteem. *Journal of Experimental Social Psychology, 69*, 232–236.
- Gutman, J. (1982). A means-end chain model based on consumer categorization processes. *Journal of marketing, 46*(2), 60-72.
- Klenosky, D. B. (2002). The “pull” of tourism destinations: A means-end investigation. *Journal of travel research, 40*(4), 396-403.
- Leão, A. L. M., & Mello, S. C. (2007). The means-end approach to understanding customer values of a on-line newspaper. *BAR-Brazilian Administration Review, 4*(1), 1-20.
- Lewis, S. P., & Arbuthnott, A. E. (2014). Non-suicidal self-injury, eating disorders, and the Internet. In L. Claes & J. J. Muehlenkamp (Eds.), *Non-suicidal self-injury in eating disorders* (pp. 273-293). Springer-Verlag.
- Luo, M., & Hancock, J. T. (2020). Self-disclosure and social media: motivations, mechanisms and psychological well-being. *Current Opinion in Psychology, 31*, 110-115.
- Manikonda, L., & De Choudhury, M. (2017). Modeling and understanding visual attributes of mental health disclosures in social media. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 170-181.
- Meredith, J. (2019). Conversation analysis and online interaction. *Research on Language and Social Interaction, 52*(3), 241-256.
- Naslund, J. A., Aschbrenner, K. A., McHugo, G. J., Unützer, J., Marsch, L. A., & Bartels, S. J. (2019). Exploring opportunities to support mental health care using social media: A survey of social media users with mental illness. *Early intervention in psychiatry, 13*(3), 405-413.
- Perry, R., Drachen, A., Kearney, A., Kriglstein, S., Nacke, L. E., Sifa, R., & Johnson, D. (2018). Online-only friends, real-life friends or strangers? Differential associations with passion and social capital in video game play. *Computers in Human Behavior, 79*, 202-210.
- Reynolds, T. J., & Gutman, J. (1988). Laddering theory, method, analysis, and interpretation. *Journal of advertising research, 28*(1), 11-31.
- Reynolds, T. J., & Rochon, J. P. (2001). Consumer segmentation based on cognitive orientations: the ChemLawn case. In T. J. Reynolds & J. C. Olson (Eds.), *Understanding Consumer Decision Making: The Means-End Approach to Marketing and Advertising Strategy* (pp. 283-298). Lawrence Erlbaum Associates.
- Smith, D., Leonis, T., & Anandavalli, S. (2021). Belonging and loneliness in cyberspace: impacts of social media on adolescents' well-being. *Australian Journal of Psychology, 73*(1), 12-23. <https://doi.org/10.1080/00049530.2021.1898914>

Data Care as Conversation: A Resource to Support Dialogue, Decision Making, and Design Activities that Respect Newcomer Communities and their Data

Shankar, Saguna

University of British Columbia, Canada | saguna.shankar@ubc.ca

Nathan, Lisa P.

University of British Columbia, Canada | lisa.nathan@ubc.ca

ABSTRACT

Arts and design-based approaches hold potential for supporting deliberation among groups with different views and values who collaborate through their information practices. This paper describes the design of such an approach, entitled *Sangam: Prompts for Discussions, Decision Making, and Design Activities that Respect Newcomer Communities and their Data*. Grounded in empirical research, *Sangam* is a set of physical cards that prompt reflection on and discussion of ethical complexities that arise when coordinating immigration and settlement services in Canada. This report describes the process of designing the *Sangam* cards, staying true to the project's research findings while also leaving space for interpretive flexibility.

KEYWORDS

Design, Ethics of Care, Immigration, Information Practice, Values

INTRODUCTION AND CONTEXT

Recent scholarship draws our attention to practices and systems through which the use of communities' data can be exploitative, further entrench deficit framings and power imbalances, and miss opportunities to work with communities as partners (e.g., Benjamin, 2016; Costanza-Chock, 2020; D'Ignazio & Klein, 2020; Eubanks, 2018). This paper is motivated by these calls to broaden participation in design and decision making. To join efforts for more deliberative, participatory processes of developing and adapting information practices, a resource in the form of a card set is introduced. The card set is entitled *Sangam: Prompts for Discussions, Decision Making, and Design Activities that Respect Newcomer Communities and their Data*. "Sangam" is a Sanskrit word for a confluence, or meeting point. This resource is informed by early findings from an ongoing research project examining *practices of caring for newcomers and their data*. This write-up offers a brief overview of the research findings, followed by a description of *Sangam*'s design and intended use.

Empirical Research: The initial research project, conducted 2020–2022, identified points of coordination and tension articulated by different groups creating, collecting, analyzing, and sharing the data of newcomers to Canada. The term "newcomer" is used to be inclusive of individuals who hold various immigration statuses. Newcomers' data was understood to comprise their personal information and accounts of experiences, as well as details such as histories of interacting with government agencies, organizations, community groups, and digital immigration systems and technologies. 14 individuals participated in the project. Each person belonged to one of five key groups supporting or studying immigration in Canada: 1) settlement service providers; 2) migrant justice activists; 3) systems designers; 4) academics specializing in immigration; and 5) government staff responsible for managing immigration and citizenship. The theoretical framework of the research project was influenced by *ethics of care* (e.g., Murphy, 2015; Puig de la Bellacasa, 2017; Tronto, 1998), and *values as hypotheses* scholarship. Previous work on *values as hypotheses* observes that values are employed in cycles of reflection and action, often negotiated through story (JafariNaimi et al., 2015). These ideas supported the project's conceptualization of *data care*, integrating attention to groups' interdependence, labor, materiality, and power with information practice theory.

Interviews with participants suggest they perceive their *data care* work as interconnected. They speak of influencing one another's activities and capabilities to make decisions about organizing their labor. Certain groups, especially academics and government staff, were seen to be claiming power over other groups by making demands for data as well as direct access to newcomer communities. These groups were also wielding power by enforcing their own expectations for the value and validity of data collected or created by other groups. This included making strong claims about epistemic authority, such as professional background and training, institutional affiliations, desirability of the content and size of data sets, or the use of certain methods for analyzing data. The array of participants' activities was striking: While some groups were reliant on volunteer support to create databases for documenting their services or advancing grassroots activism, others were experimenting with designing systems to predict global patterns of migration or developing chatbots to provide advice to newcomers on immigration processes. Activities of *data care* were marked by signs of competition, confusion, and conflict, and at times, coordination between groups. The ethics of collecting, analyzing, and sharing communities' data are therefore understood as *complex and changeable*. Implications of the project suggest further studies are needed of groups navigating collective issues

across a range of cultural norms, professional roles and responsibilities, and ethical and political positions, particularly as our information and technology assemblages become more intricate and intertwined.

DESIGNING THE RESOURCE

Motivations: Why Cards?

Physical cards were chosen as a format for enabling deliberation for several reasons. The five key groups that support or study immigration through their use of data work in very different contexts and have varying interests. These groups are understood as potential “users” of the cards. The flexibility of a physical set of cards from which prompts to discuss *data care* can be selected supports groups’ choice and adaptation. The cards acknowledge a plurality of perspectives that one or more groups may hold about data, ethics, and immigration. One knowledge system, setting, or technological system is not assumed. The card set, however, does emphasize a shared concern for newcomers’ wellbeing and a commitment to the need to support respectful deliberation.

Methods and Intended Outcomes

The *Sangam* card set communicates preliminary findings of the research project (Shankar, 2021) through scenarios, questions, and four themes by which the cards are organized. The content of the cards is drawn from actual events, conflicts, dilemmas, and priorities raised by participants in the interviews. The four themes of the cards are *Care*, *Culture*, *Relationships*, and *Stories*, which represent issues brought up by participants as they talked about their information practices of studying or supporting immigration. The cards were created during the research project’s interpretation and writing stages, with the guidance of Franks’ *dialogical narrative analysis* (2010) and *reflexive thematic analysis* from Braun and Clarke (2019). Narrative and thematic analysis of interviews from the research project were used to generate textual content for the cards, with the intention to integrate findings back into practice. This led to the final set of 24 cards in addition to several blank cards to support prompt development by those engaging with the set. The printed card set will be distributed to each participant in the research project. A digital version of the cards will also be published under Creative Commons licensing (CC BY-NC-SA), and made freely accessible to others who wish to adapt it to suit their own decision-making processes.

Artistry, Existing Work, and Future Inquiries

The visuals for the cards were created by Olúwáṣọlá Kèhìndé Olówó-Aké, an artist who offered *Aroko* as a framework from her Yorùbá heritage. *Aroko* has traditionally been a means of carrying messages between groups through significant symbols (Ojo, 2013). We have borrowed from *Aroko* to visually represent themes of the cards and carry findings of the project to groups in the immigration sector. Together we designed the cards, influenced by the *Envisioning Cards* which were developed to elicit stakeholders’ values as they relate to long-term effects of technologies (Friedman et al., 2008; Yoo et al., 2022). The card set is also inspired by other arts and design-based interactive resources, such as Gomez et al.’s *Mind the Five Card Game* (2020) and the Humans Understanding Humans *Conversation Cards*. These are only some examples of a growing genre of interactive resources within and outside the design domain — e.g., *The Tarots of Tech Cards*; Wong et al.’s *Timelines* (2021). Future collaborative inquiries will investigate whether and how this project’s card set and other techniques like it might advance dialogues on information practice and policy in community spaces, classrooms, and design labs.

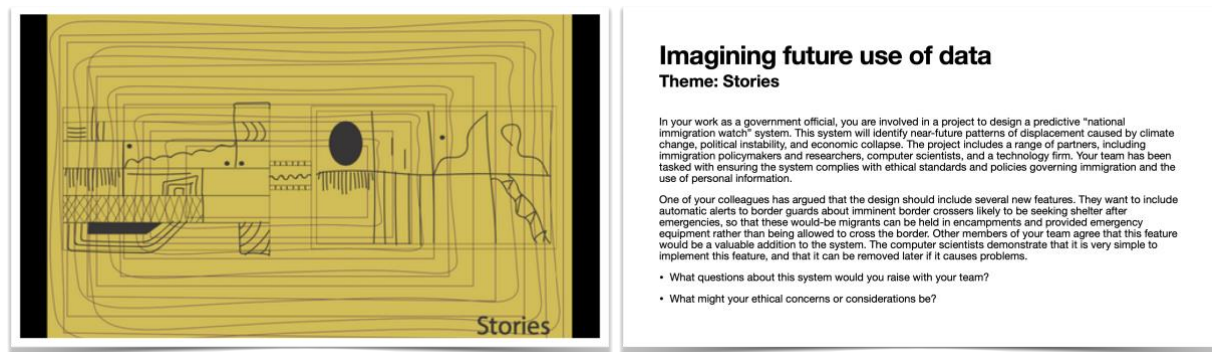


Figure 1. Example of a card in the “Stories” theme entitled, “Imagining future use of data”

CONCLUSION: A PLACE TO START CONVERSATIONS

The cards are informed by the potential for deliberation, exchange, and mutual learning to shape changes in information practices. *Sangam* is a resource which can be used to draw out understandings of different expectations and norms about how newcomer communities’ data should be cared for by various groups. By encouraging conversations, *Sangam* and other resources like it generate possibilities for negotiating how information practices may support the coordination of collective processes with considerations for ethical complexity.

ACKNOWLEDGMENTS

Many thanks for the bold creativity and generosity shared by Olúwásolá Kẹ̀hìndé Olówó-Aké, Ruby Pang, Arina Sin, Katherine Gillieson, Rodrigo dos Santos, and Michelle Kaczmarek.

REFERENCES

- Benjamin, R. (2016). Informed refusal: Toward a justice-based bioethics. *Science, Technology, & Human Values*, 41(6), 967–990.
- Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4), 589–597.
- Costanza-Chock, S. (2020). Design justice: Community-led practices to build the worlds we need. The MIT Press.
- D'Ignazio, C., & Klein, L. F. (2020). *Data feminism*. The MIT Press.
- Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.
- Frank, A. W. (2010). *Letting stories breathe: A socio-narratology*. University of Chicago Press.
- Friedman, B., Nathan, L., Kane, S., & Lin, J. (2008). *Envisioning Cards*.
- Gomez, R., Newell, B. C., & Vannini, S. (2020). Mind the Five Card Game. *The International Journal of Information, Diversity, & Inclusion (IJIDI)*, 4(2), 116–122.
- JafariNaimi, N., Nathan, L., & Hargraves, I. (2015). Values as hypotheses: Design, inquiry, and the service of Values. *Design Issues*, 31(4), 91–104.
- Mol, A., Moser, I., & Pols, J. (Eds.). (2010). *Care in practice: On tinkering in clinics, homes and farms*. transcript Verlag.
- Murphy, M. (2015). Unsettling care: Troubling transnational itineraries of care in feminist health practices. *Social Studies of Science*, 45(5), 717–737.
- Ojo, M. O. D. (2013). Symbols of warning, conflict, punishment and war and their meanings among the pre-colonial Yorùbá natives: A case of Àrokò. *Antropologija*, 13(1), 39–60.
- Puig de La Bellacasa, M. (2017). *Matters of care: Speculative ethics in more than human worlds*. University of Minnesota Press.
- Shankar, S. (2021). Coordinating migration: Caring for communities & their data. *Proceedings of the Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing*, 295–298. <https://doi.org/10.1145/3462204.3481798>
- Tronto, J. C. (1998). An ethic of care. *Generations*, 22(3), 15–20.
- Wong, R. Y., & Nguyen, T. (2021). Timelines: A world-building activity for values advocacy. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–15. <https://doi.org/10.1145/3411764.3445447>
- Yoo, D., Logler, N., Ballard, S., & Friedman, B. (2022). Multi-lifespan Envisioning Cards: Journeying from design theory to tools for action. *Proceedings of the 2022 ACM Designing Interactive Systems Conference*, 557–570. <https://doi.org/10.1145/3532106.3533495>

Factors Influencing Criminologists' Open Research Data Practices: Trust, Contract, and Value

Shih, Yi-Ru Academy for the Judiciary, Ministry of Justice, Taiwan | yiru840321@gmail.com
Yang, Yu-Ju Carnegie Mellon University, USA | yujuy@andrew.cmu.edu
Jeng, Wei National Taiwan University, Taiwan | wjeng@ntu.edu.tw

ABSTRACT

This study investigated how researchers in criminology dealt with their research data and explored the potential factors that would influence their willingness to share and reuse others' research data. Our findings unveiled three factors, trust, contract, and value, deeply rooted in researchers' attitudes and their data sharing and reuse patterns and behaviors. Through comprehending their concern behind sharing and reusing research data, we seek crucial insights that could be beneficial in developing a criminological research data repository, further consolidating the research data infrastructure in social science. Sharing and reusing data can advance scientific research and provide innovative possibilities for practice in criminology (e.g., new correctional methods, new crime prevention policies), providing sustained and thriving energy for the field.

KEYWORDS

Research data infrastructure, Data in Criminology, Research data management, Open research data

INTRODUCTION

Since the concept of data-intensive science, the fourth paradigm of science, has been proposed for more than a decade (Hey, 2009), it accurately predicts a research team's current state in most disciplines: the role of research data has become a more critical asset than ever. Today, researchers derive insights through a large amount of gathered data through computational and analytical techniques. Governments, disciplinary communities, and research institutions are building research data infrastructures to better support researchers' research activities around data. This trend has undoubtedly blown over to the social sciences as well. Researchers in social science often handle very complex data (Jeng & He, 2022). Taking criminology as an example, a researcher may involve primary data, secondary data, and data from vulnerable populations (e.g., prisoners) (Thanos, 2011). Understanding how these researchers coping with complex nature of data, and what special considerations are addressed, becomes useful materials for strengthening current social science data infrastructures.

In this study, we report a complete qualitative study with 13 criminologists in the form of an extended abstract (as a poster paper), in which two focus group sessions and in-depth interviews were conducted. The study's goal was to capture the perspectives and considerations of researchers in criminology on how they use and share data for their research within a data-intensive paradigm. Specifically, we investigated their willingness, experiences, and considerations for sharing and reusing research data and anticipates for data curation professionals to establish an ideal research data repository based on their needs to better build the research data infrastructure and foster the research data sharing and reuse in social science (Jeng & He, 2022).

METHODOLOGY

This study used a qualitative approach with two small focus group sessions (one consisting of five individuals from a criminology research institution and one consisting of three university faculty members who work in criminology departments), as well as five semi-structured, in-depth interviews. Interviews and focus groups were held in Taiwan and in spring 2021. The choice of conducting small focus groups for capturing scholars' responses on open science topics has been seen in previous work (Lyon et al., 2020). Before each interview, we sent out an online questionnaire built on Qualtrics, an online survey software, to gather our participants' training backgrounds and research interests. This particular process helped us better tune the protocol, improve its accuracy, and help avoid technical problems, e.g., grouping interviewees with advisor-advisee relationships into the same focus group.

As for the data analysis, we first performed a round of open coding for the transcribed audio data and gradually reached several axial coding labels. After collecting these axial coding labels, we categorized them into three overarching central themes for criminologists' data practices—trust, contract, and value as our findings.

RESULTS

In this extended abstract, we present the most key notions that influence criminologists' data-sharing and reuse practices in: *trust*, *contracts*, and *values*. Participants' demographical information, including their research interests, is available at https://osf.io/a8fh5/?view_only=726379edffd044b7a042c58b5aa3cb2e for review.

Trust

Lower levels of *trust* between data reusers and data sharers can inhibit open research data's use in both ways: data reusers do not use data because lacking of trust from data sharers, and data sharers do not release the data because they do not trust potential data users (Zuiderwijk et al., 2020). Based on our results, we further identify two sub-categories of concerns amid the notion of trust: Data quality and interpersonal trust.

Data quality. Being aligned with the previous study (Yoon & Lee, 2019), we confirmed that criminologists usually check the target dataset's transparency, accuracy and examine whether their collecting processing is reliable and solid before reusing others' data, no matter what types of data, e.g., qualitative data, quantitative data, data from other individual researchers, government statistics, or documents from public sectors. If the data lacks accuracy and completeness, it may lead to severe problems such as misuse and misinterpretation by scholars who access the data (G06, G08). In terms of data processing, participants reported that they usually check how data are collected, cleaned, and analyzed, which is basically aligned with the research data lifecycle. In addition, data's quality also related to researchers' perceived transparency: if a data sharer provides sufficient description of the released data, researchers were less likely to misinterpret or misuse the data when using it (G06).

Interpersonal trust. Interpersonal trust also influences researchers' willingness to collect, share, and reuse data. Scholars rely on interpersonal relationships to access unpublished data (the 'gray' data) and reach hard-to-reach interviewees such as gang members. Our interviews summarized three foundations of interpersonal trust: honesty and integrity, ability to use, and past partnerships. Not surprisingly, data providers (e.g., government agencies) are only willing to release relevant data under-promise. Trust can only be strongly built on the basis that both parties to the data sharing/use share a common belief in each other's capability to use the data accurately and productive. Finally, the well-established partnerships in the past by all means create trust in the future.

Contract

We found that criminologists consider various forms of contracts during the course of using or sharing research data with others. Contracts usually involve norms such as the ownership and the right to use research data, which influences scholars' willingness and attitudes toward sharing research data. Signing a research proposal commissioned by the government or a memorandum of understanding (MOU) is a part of having the contract, which establishes the obligations between the government and scholars for providing and using data through concrete terms. In addition, the right to use research data (e.g., whether the data can be adapted and published after the research is completed) is also declared (G01, G04, G05). If the ownership and rights to use the data are not declared in the contract, scholars will hesitate to share or disclose data (P01).

With its sensitive nature, data used by criminologists strictly follow laws and regulations. For example, to protect the youth's rights and interests, the government usually cancels the youth's criminal record, requiring additional work to fill in the gap to carry out the intended study related to juvenile delinquency. When scholars share or disclose research data, they are often constrained by their general ethical decision-making and research ethics. Finally, the *silent agreement* shared among scholars refers to an implied contract between scholars. They are willing to share research data with whom and only on the premise that lives a partnership (P01, P02, G06, G07, G08). On the other hand, criminologists are often synchronized in their minds as they understand all the efforts it takes to collect and process the data, making them feel hesitant to request others' data.

Value

Data's value is interwoven with the efforts that researchers pay during the research data lifecycle. Our participants reveal that collecting data from scratch, data cleansing, and developing code books all speak the same story about the unique value individual researchers add to their research data to make them unique. Data's publication value also greatly influences criminologists' attitudes and decision-making. In general, when sharing and disclosing research data, scholars consider the subsequent publication value of the data. They tend to share or disclose data only when the data has completed its mission, e.g., completed the academic publication process (G01, G02, G03, G04, G05).

NEXT STEPS

The current study unveiled three components, namely, trust, contract, and value, which are deeply rooted in researchers' attitudes and their data sharing and reuse patterns and behaviors. While observing how these components interweave, we hoped that a clear picture of open research data practices can be provided to the field of criminology in Taiwan.

Finally, it is worth mentioning that the current repository, the National Archive of Criminal Justice Data (NACJD), primarily sponsored by the U.S. Department of Justice, facilitates criminological data-sharing. However, research on the needs of NACJD users is relatively lacking. Our next step is to extract participant narratives relevant to the use of data repositories and try to relate them to the current data infrastructure design and ultimately, develop a data infrastructure that would inspire future research data infrastructure in social science.

ACKNOWLEDGEMENTS

This work was financially supported by the Ministry of Science and Technology (MOST) in Taiwan, under MOST 111-2636-H-002-004-. Please address all correspondence to Wei Jeng (wjeng@ntu.edu.tw).

SUPPLEMENTAL MATERIALS

Participants' demographical information is available at <https://osf.io/a8fh5/>

REFERENCES

- Hey, T., Tansley, S., & Tolle, K.M. (2009). *The fourth paradigm: data intensive scientific discovery*. Redmond, WA: Microsoftresearch.
- Jeng, W., & He, D. (2022). Surveying research data-sharing practices in US social sciences: a knowledge infrastructure-inspired conceptual framework. *Online Information Review*. <https://doi.org/10.1108/OIR-03-2020-0079>
- Lyon, L., Jeng, W., & Mattern, E. (2020). Developing the tasks-toward-transparency (T3) model for research transparency in open science using the lifecycle as a grounding framework. *Library & Information Science Research*, 42(1), 100999.
- Murillo, A. P., Curty, R. G., Jeng, W., & He, D. (2020). Confronting the Challenges of Computational and Social Perspectives of the Data Continuum. *Data and Information Management*, 4(2), 119-126.
- Thanos, C. (2011). Global Research Data Infrastructures: The GRDI2020 Vision. Report of the GRDI2020 project funded under the 7th Framework Programme, Capacities-GEANT & eInfrastructures.
- Yoon, A., & Lee, Y. Y. (2019). Factors of trust in data reuse. *Online Information Review*, 43(7), 1245-1262.
- Zuiderwijk, A., Shinde, R., & Jeng, W. (2020). What drives and inhibits researchers to share and use open research data? A systematic literature review to analyze factors influencing open research data adoption. *PLoS one*, 15(9), e0239283.

Organizing Common Bean (*Phaseolus vulgaris*) Research: A Model for Information Resources Integration

Shobhane, Mahbub
Ram, Shri

East West University, Bangladesh | mahbubshobhane@gmail.com
Sikkim University, India | shriram.tiet@gmail.com

ABSTRACT

Common Bean (*Phaseolus vulgaris* L.) an essential nutrition source for protein and is quite common in the Indian subcontinent, China, Brazil, and the USA. Common Bean contains nutritional component which will contribute to the reduction of food scarcity worldwide in the coming years (Nadeem et al. 2021). In recent years, the research trends applying informatics techniques have given opportunities to harness genomic level information for better sequencing, generating a large volume of data. The data is generally being utilized for better crop production and improved varieties. The generated data are stored and developed as databases, web servers, publications, and patents. However, managing data in a helpful way is a challenge. This paper proposes a model for developing an integrated platform that comprehends various Common bean information. The proposed model shall provide curated information resources about Common Bean, facilitating the single platform access. The intended benefits can be visualized as the researcher can find all information in one place to showcase the progress of research taking place in this field.

KEYWORDS

Common Bean, *Phaseolus vulgaris*, Collection Development, Information Resource Integration, Informatics

INTRODUCTION

The Common Bean is taxonomically known as *Phaseolus vulgaris* L. (POWO, 2017). The scientific name has a list of synonyms listed by the International Plant Names Index (IPNI) (POWO, 2017). India, Myanmar, Brazil, China, Mexico, and the USA have been the highest common bean-producing countries in the last five years (FAO, 2020). The application of informatics in biological research has broadened the scope of publicly available web-based resources developed to conduct computational modeling (Benson et al., 2012). Various models have been proposed to organize resources, and it is argued that the model needs to be flexible in terms of data sources, programming language, and informatics tools to remain endorsed and user-oriented (Bruneau et al., 2019). Figure 1 depicts the publication's growth in the last seventy-five years indexed in the SCOPUS database.

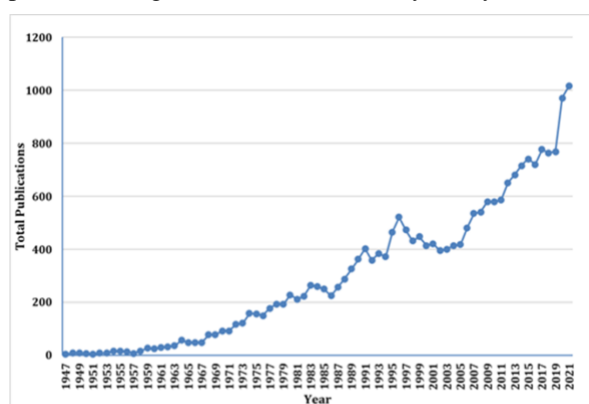


Figure 1. growth of literature in *Phaseolus vulgaris* (Source: SCOPUS)

Several web portals have been developed focusing on particular specimens of crops; one of the most comprehensive and developing portals for legume research is the Legume information system (Dash et al., (2016). Despite being a major grain crop, *Phaseolus vulgaris* has been studied very few times compared to crops with similar importance and sparsely distributed information makes the searching process extra laborious (Kafer et al., 2021). Egan & Vatanparast (2019) proposed an even more advanced portal for legume biology research using Next-generation sequencing (NGS) methods. Ram & Rao (2012) suggested a single platform-based solution to organize heterogeneous bioinformatics resources in one place.

This study aims to design and develop an individual species or clades-based information gateways (Bruneau et al, 2019) integrating heterogeneous information resources on *Phaseolus Vulgaris* available over the internet. The process includes the identification of various forms and formats of information resources available on *Phaseolus vulgaris*, secondly, the design and development of an integrated platform for information resource integration

METHODOLOGY

According to Bruneau et al. (2019), the construction of the model relies on the following four main steps for designing and developing an integrated web-based platform. (i) Definition of the type of information to aggregate; (ii) Selection of online data sources; (iii) Development of scripts to centralize the information; and (iv) Production of an online graphical user interface (GUI) to retrieve and visualize information. This paper proposes a multi-method approach consisting of Information identification and aggregation; computational programming and scripting; and user testing.

TYPES OF INFORMATION ON *PHASEOLUS VULGARIS*

Related peer-reviewed publications: Peer-reviewed publications are resources with a heightened level of empirically. Many harvesting sites make automatic harvesting of published literature. Web of Science, PubMed & SCOPUS can be used for custom-curated content. **Webservers:** PvTFDB, a web-accessible transcription factors (TFs) database for *P. Vulgaris* containing 2370 putative TF gene models in 49 TF families. (Bhawna et al., 2016) The University of California Davis developed PhaseolusGenes to identify and explore markers, quantitative trait loci (QTLs), and SSRs region information for common beans. Phaseolus vulgaris Gene Expression Atlas (PvGEA) allows researchers to search for genes expressed and query gene expression profiles (O'Rourke et al., 2014). **Databases:** IPNI is the biggest plant bibliographic information database linked with the biodiversity heritage library (BHL). Global Biodiversity Information Facility (GBIF) offer harvesting data from different national or regional aggregators, and they are comprehensive. Crop-based images and photo databases are the most growing online image data portals that have been developed in the last few years; this portal includes a rich source of photographs and user observations. Image recognition tools are improving rapidly because of the application of machine learning and artificial intelligence. TreeBASE, DAYAD GitHub, Figshare, etc. are the most growing source of phylogenetic information. International Nucleotide Sequence Database Collaboration (INSDC), NCBI's GenBank, the European Nucleotide Archive, and the DNA Databank of Japan are some of the major data sources of molecular data. The International Center for Tropical Agriculture, also known as CIAT, contains the world's most extensive germplasm collection of common Bean 41500 accessions from the donor countries. Currently, it is coordinating 29 countries in Sub-Saharan Africa, and the activity called the Pan-Africa Bean Research Alliance (PABRA) is supported by the Canadian International Development Agency (CIDA) (Islam, 2006). Public institutions such as EMBRAPA, IAPAR, and IAC are the highest contributors to research on common beans (Kafer et al, 2021). **Software tools:** Software tools are essential for informatics; depending on the data types, software analysis can vary. Duck et al. (2016) highlighted the rapid growth of software used in bioinformatics practice and emphasized the reuse, choice, and sharing of software. To incorporate all the types of information mentioned above, the model will require some specific features. An architectural design of the integrated model is presented in the next section.

PROPOSED MODEL

The proposed information resource integration system generally involves three levels. Such models have been proposed to integrate resources for other subjects (Ram & Rao, 2012; Li et al. 2010). *The Application Level:* The user interface (UI) used for primary communication with the system. Searching, fetching, and visualization are the main features of this level. The user interface (UI) is essential, primarily for proper user communication; the front must be interactive. Now a day, there are specialized scripted frameworks for unique interface design. *Information Resource Level:* This level is concerned with the categorization of various heterogeneous information on the subject area. Since the data sources are very heterogeneous, there will need additional standardization protocol, especially for the phylogenetic tree, geospatial, image gallery, and sequential data. *Data Source Level:* This level is related to the storage of the data sources as per the standard metadata formats. Depending on the data types, the features of the databases will vary, like Gene transfer format (GTF) is a file format that holds information about gene structure; modern phylogenetic trees are based on molecular data, geospatial data, image data, etc. Each data type will require a specialized file format to get the data stored in the database using the proper harvesting protocol.

DISCUSSION

This paper proposes a layered architecture model to organize heterogeneous common beans information resources. The proposed web-based model is based on curated data sources available as digital documents (Articles, conference papers, books, chapters), databases, web servers, and patents, which are distributed at various sources, local databases, and institutional archives. Gathering such information is a challenge for novice researchers. This proposed web-based meta-server helps to identify the resources in one place. The growing dominance of technology in biological research has broadened the scope of web-based resource development in bioinformatics (Egan & Vatanparast, 2019). New server specialized in biological specimen information is continuously added with the internet. Integration of all these resources can benefit the researchers in identifying resources essential for research, teaching, and learning. Moreover, it will be helpful to the scientists who are engaged in crop development. Resource integration is limited to all the digital resources available in the form of published literature, databases, web servers, patents, other genome informatics and related resources.

CONCLUSION

The development of technology and implementation of next-generation sequencing (NGS) has broadened the scope of research in genetics and genomes (Egan & Vatanparast, 2019). Common Bean is essential for globally food security and one of the major grain crops with a long domestication history. The crop has emerged with a scattered genetic diversity. Availability of this diverse species is essential for crop development, especially as an individual species or clades-based information gateways. This study is a novice information researcher approach to facilities the bioinformatics community with an aggregate and latest resource on common Bean.

REFERENCES

- Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2012). GenBank. *Nucleic acids research*, 41(D1), D36-D42.
- Bhawna, B., Bonthala, V. S., & Gajula, M. N. V. (2016). PvTFDB: a Phaseolus vulgaris transcription factors database for expediting functional genomics in legumes. *Database*, 2016.
- Bruneau, A., Borges, L. M., Allkin, R., Egan, A. N., De La Estrella, M., Javadi, F., ... & Zhang, R. (2019). Towards a new online species-information system for legumes. *Australian Systematic Botany*, 32(6), 495-518.
- Duck, G., Nenadic, G., Filannino, M., Brass, A., Robertson, D. L., & Stevens, R. (2016). A survey of bioinformatics database and software usage through mining the literature. *PloS one*, 11(6), e0157989.
- FAO (2020). FAOSTAT Retrieved from-<https://www.fao.org/faostat/en/#data/QCL/visualize>
- Kafer, J. M., Gobatto, D. R., Woyann, L. G., Carneiro, E., da Silva, G. R., & Finatto, T. (2021). The impact of molecular markers in common Bean through a scientometric approach. *Euphytica*, 217(7), 1-12.
- Li, G., Li, W., & Ma, Q. (2010, December). Research of information resources integration model based on grid service. In *The 2nd International Conference on Information Science and Engineering* (pp. 1421-1424). IEEE.
- O'Rourke, J. A., Iniguez, L. P., Fu, F., Bucciarelli, B., Miller, S. S., Jackson, S. A., ... & Vance, C. P. (2014). An RNA-Seq-based gene expression atlas of the common bean. *BMC genomics*, 15(1), 1-17.
- Nadeem, M. A., Yeken, M. Z., Shahid, M. Q., Habyarimana, E., Yilmaz, H., Alsaleh, A., ... & Baloch, F. S. (2021). Common bean as a potential crop for future food security: an overview of past, current and future contributions in genomics, transcriptomics, transgenics and proteomics. *Biotechnology & Biotechnological Equipment*, 35(1), 759-787.
- POWO (2017). Phaseolus vulgaris L. Retrieved from-<https://powo.science.kew.org/taxon/514191-1#bibliography>
- Ram, S., & Rao, N. L. (2012). iBIRA-integrated bioinformatics information resource access: Organizing the bioinformatics resourceome. Reference services review.

Visual Semantics of Memes: (Re)Interpreting Memetic Content and Form for Information Studies

Smith, Alexander O. School of Information Studies, Syracuse University, USA | aosmith@syr.edu
Tacheva, Jasmina School of Information Studies, Syracuse University, USA | ztacheva@syr.edu
Hemsley, Jeff School of Information Studies, Syracuse University, USA | jjhemsle@syr.edu

ABSTRACT

Here we attempt to contextualize memetics for Information Studies by reinterpreting Internet Memetics. In doing so, we propose a way to empirically study semantics of informational content. This is done by collecting images from meme entries on KnowYourMeme, a digital meme repository, and clustering features found in those images. The features are found using Google Cloud Vision, Google's computer vision project. We suggest how informational meaning (semantics) is shaped between similarity (operating as the idea of the meme) and difference (features in the image). Early results show that memetic ideas may be referenced as the shape of information across image collections.

KEYWORDS

Memetics; Visual Methods; Computer Vision; Network Clustering; Semantic Information

INTRODUCTION

In Internet Memetics (IM), the study of memes online, Shifman outlines three dimensions in defining a meme: content, form, and stance. She reduces information to either material behavior or ideas (2013). Instead, we reposition information *between* what Shifman calls "content" (i.e. the idea) and "form" (i.e. the physical aspects) of memes. The goal of this work is to suggest where information is located in contemporary memetic theory. We were motivated to show that information can contribute in explaining memetic expressions with precision to Information scholars who wish to understand memetics or visual culture more broadly.

We make a contribution using network clustering to recontextualize IM for Information research. Following from previous memetic clustering publications on Internet memes (Dang et al., 2015; Leskovec et al., 2009; Theisen et al., 2020; Zannettou et al., 2018) we provide an interpretation of commonly agreed upon "memes" by gathering meme entries from KnowYourMeme. By bridging IM with clustering approaches, this work moves towards offering a more Informationally interpretable meme. In particular, we attempt to answer one theory question:

Do visual features (i.e. "physical aspects") of an internet meme proxy for the shape of a cultural "idea"?

This question suggests a new interpretation in which Information Studies can analyze memes. Traditionally, "ideas" are mental objects which cannot be directly detected. However, clustering the object tags in a collection of memetic image documents offers a representation operating as a "cultural idea." Strongly clustered aspects of the network operate as the "form" of a "cultural idea." This provides a "shape" of a meme that might be considered the information the expression of culture more so than the total information contained in the image collection. Our data provides an informational process of meaning making. A practical implication of this research is in locating structural properties of visual information which so far has been understudied (Cho et al., 2021) we suggest a exploratory methodological way to contribute to this space.

METHOD

We extend our informational interpretation of memes to a computer vision approach: an algorithmic method of extracting features from images. Then, we cluster images based on these features. We use a Bayesian Planted Partition (BPP) algorithm clustering method on the features returned from Google Cloud Vision's pretrained computer vision algorithms.

Our data was collected from the image galleries of KnowYourMeme entries, a popular archive of memes often used in research (Pettis, 2021). We chose four popular meme entries: "I Took an Arrow to the Knee", "Me Gusta", "Rick-Roll", and "Slenderman". These were selected because they offered differences in the way images were changed or collaged within each meme.

For computer vision, Google Vision API has been used in other meme studies (Beskow et al., 2020), and returned image tags in JSON files. Within each meme, we performed feature clustering on the tags using Gephi. We built a bipartite network of tags to each image. Then, we used multimode network projection which creates an image network where the edges between the image nodes are weighted by their shared tags.

Using a clustering method, we intend to interpret potential “ideas” in a meme entry. We initially used Louvain method of community detection (Blondel et al., 2008). With this method, we found smaller numbers of clusters, however they were less stable across all the meme entries. In order to find more stable clusters, we instead opted to use the BPP algorithm (Zhang & Peixoto, 2020) which sacrificed smaller, more interpretable numbers of clusters. However using Zhang and Peixoto’s clustering algorithm enables us to test the following hypotheses with more clarity. Our hypotheses are as follows:

H1: Memes with similar visual content (i.e. “Me Gusta” and “Rick-Roll”) will have fewer and denser clusters.

H2: Memes with visually dissimilar content will have more distributed clusters (i.e. “I Took an Arrow in the Knee” and “Slenderman”).

These hypotheses are motivated by the properties related to the use of the meme. Me Gusta is a popular template meme in that commonly uses a similar background image: i.e. an image macro. Rick-Roll is commonly used as a trick. Being “Rick-Rolled” is to click on a link or see an image referencing Rick Astley’s song “Never Gonna Give You Up” despite expecting something else. For the trick to be successful, the reference to needs to be obvious, hence a minimal clustering. Slenderman, on the contrary, is more often referencing narrative-building and thus a set of less connected stories. Because of lack of narrative cohesion, Slenderman’s features are expected to be more distributed. Similarly, “I Took an Arrow in the Knee” is often more recognized by text references to a quote from the videogame narrative The Elder Scrolls V: Skyrim.

RESULTS

The output suggests that there are GCV tags which cluster more regularly than others in each meme. Also importantly, in support of our hypotheses, the number of clusters does generally follow the pattern we expected. Rick-Roll and Me Gusta do have fewer clusters than I Took an Arrow and Slenderman.

Meme_Entry	# of Images	# of Tags	# of Clusters	Med Cluster Size	Mean Cluster Size	Max Cluster Size	Min Cluster Size
Rick-Roll	197	337	41	8.0	8.2	33	1
Me Gusta	488	579	58	6.0	7.9	48	1
I Took An Arrow	348	469	62	6.5	7.6	32	1
Slenderman	686	459	84	5.5	6.9	25	1
All Combined	1719	872	133	6.0	7.7	38	1

Table 1. Bayesian Planted Partition Algorithm Clustering Results

It is also notable that the combined image clusters are not the sum of the clusters for each meme entry separately. This implies there are a significant number of tags that are in multiple meme entries. While there are not enough meme entries here to say definitely, it does not appear that the number of images are responsible for the number of clusters or each cluster’s size, however further exploration of this might generate more robustness once more meme entries are gathered with a more clear sampling strategy.

CONCLUSION

Early results suggest visual properties expected from a meme entry’s use do impact how many clusters exist across the images, suggesting more or less ideas. This means a clustering approach shows where information joins with semantic properties through clustering visual data in the images. The tags proxy for how a human might make meaning of those visuals. To address limitations, more meme entries could be collected to address significance and robustness.

We have shown theoretical value in empirically discovering common clusters of “ideas” across these images. Noticing the distinctions in cluster structure across each meme entry supports our initial question. Informational distinctions in structure may proxy for the semantic structure of memes. Further structural analysis of clusters could continue to develop this argument with more nuance.

REFERENCES

- Beskow, D. M., Kumar, S., & Carley, K. M. (2020). The evolution of political memes: Detecting and characterizing internet memes with multi-modal deep learning. *Information Processing & Management*, 57(2), 102170. <https://doi.org/10/ggfp6g>
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008. <https://doi.org/2008101003130400>
- Cho, H., Pham, M. T., Leonard, K. N., & Urban, A. C. (2021). A systematic literature review on image information needs and behaviors. *Journal of Documentation*. <https://doi.org/10/gk5b38>

- Dang, A., Moh'd, A., Gruzd, A., Milios, E., & Minghim, R. (2015). A visual framework for clustering memes in social media. *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 713–720. <https://doi.org/10/gg7jdz>
- Leskovec, J., Backstrom, L., & Kleinberg, J. (2009). Meme-tracking and the Dynamics of the News Cycle. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 497–506. <https://doi.org/10/dp59g8>
- Pettis, B. T. (2021). Know Your Meme and the homogenization of web history. *Internet Histories*, 0(0), 1–17. <https://doi.org/10/gmkh47>
- Shifman, L. (2013). *Memes in Digital Culture*. The MIT Press.
- Theisen, W., Brogan, J., Thomas, P. B., Moreira, D., Phoa, P., Weninger, T., & Scheirer, W. (2020). Automatic Discovery of Political Meme Genres with Diverse Appearances. *ArXiv:2001.06122 [Cs]*. <http://arxiv.org/abs/2001.06122>
- Zannettou, S., Caulfield, T., Blackburn, J., De Cristofaro, E., Sirivianos, M., Stringhini, G., & Suarez-Tangil, G. (2018). On the Origins of Memes by Means of Fringe Web Communities. *ArXiv:1805.12512 [Cs]*. <http://arxiv.org/abs/1805.12512>
- Zhang, L., & Peixoto, T. P. (2020). Statistical inference of assortative community structures. *Physical Review Research*, 2(4), 043271. <https://doi.org/10.1103/PhysRevResearch.2.043271>

Influence of Interdisciplinarity on Scientific Impact: Case of Climate Change Field

Song, Yijia

School of Public Affairs, Zhejiang University, China | jnsyj818@outlook.com

Fang, Zhichao

Renmin University, China; CWTS, Leiden University, The Netherlands | z.fang@cwts.leidenuniv.nl

Fu, Hui-zhen

School of Public Affairs, Zhejiang University, China | fuhuizhen@zju.edu.cn

ABSTRACT

As interdisciplinary research attracts more attention, interdisciplinarity has become a significant factor to be explored for scientific impact. This study explored the influence of interdisciplinarity on scientific impact in the field of climate change, based on Web of Science data. Taking newly-developed indicator DIV* and its components as interdisciplinarity measures, we used OLS regression and topic analysis to analyze their relationships with short-term and long-term scientific impact, respectively. We found that DIV* had an inverted U-shaped relationship with scientific impact in overall positive correlation, and both long- and short-term scientific impact increased with the growth of interdisciplinarity with a declining growth rate in most cases. For the three components of DIV*, balance and disparity are negatively and positively correlated with scientific impact, respectively, while variety has an inverted U-shaped relationship. These findings provided theoretical and decision-making support for evaluating the impact of interdisciplinary research in the field of climate change.

KEYWORDS

Interdisciplinarity; Scientific Impact; Climate Change

INTRODUCTION

Interdisciplinary research, as an important scientific research innovation method, provides humans to solve complex problems in depth (National Academy of Sciences, 2005). To analyze, synthesize and harmonize the connections between disciplines into a coherent whole, interdisciplinarity emerged, referring to a synthesis of two or more disciplines (Choi & Pak, 2006). As interdisciplinary research draws more attention from academia and government departments, it is interesting to explore the role of interdisciplinarity in scientific impact. Scholars have found that more than 90% of highly cited research usually had a higher degree of interdisciplinarity (Chen et al., 2015).

Although there have been studies that focused on the relationship between interdisciplinarity and scientific impact (Larivière & Gingras, 2010; Leahey et al., 2017; Yegros-Yegros et al., 2015; Zhang et al., 2021), diverse research options have resulted in a lack of consensus among related studies. Scientific impact, usually represented as citations, behaved differently in the long and short term (Baumgartner & Leydesdorff, 2014; Garfield, 1998). However, most studies only considered a certain citation time window and ignored the difference between short-term and long-term citations. The citation life cycle, which has rarely been used to describe the aging process of interdisciplinary research (Cano & Lind, 1991), could help to determine the appropriate citation window.

Climate change has been listed as one of the major global issues today by the United Nations. Promoting interdisciplinary research is a necessary strategy to deal with the threat of global climate change (Hellsten & Leydesdorff, 2016; McCright et al., 2013; Weart, 2013). This study firstly revealed interdisciplinarity by newly-developed indicator DIV* (Leydesdorff et al., 2019) and its components variety, balance, and disparity, and then identified citation life cycle of climate change research. After understanding interdisciplinarity characteristics and scientific impact respectively, we explored the role of interdisciplinarity in the scientific impact in the whole field of climate change and several sub-topics.

METHOD

We downloaded publications related to climate change in Web of Science Core Collection between 2001 and 2020. The document type was limited to articles and reviews. The annual citation data of each paper were also extracted from the Web of Science Core Collection. Since the subsequent analysis was performed by using the citation life cycle of 3 years and 10 years, two different data sets of 2001-2017 (N=109,282) and 2001-2010 (N=31,991) were screened out as long-term and short-term respectively for subsequent plotting and regression analysis.

For the measurement of interdisciplinarity in our study, we took the integrated measure DIV* as the main comprehensive indicator for interdisciplinarity. The three sub-indicators representing three classic aspects of interdisciplinarity: variety, balance (1-Gini), and disparity, were also employed to give detailed information to describe interdisciplinarity. For a specific paper, its interdisciplinarity measures were all calculated based on the distribution of the subject categories of its references. In terms of scientific impact, we chose citation count as its measure, and used the citation life cycle to observe. Publications were divided into eight different groups according to the level of interdisciplinarity, and the citation life cycles of these eight groups could be compared.

85th Annual Meeting of the Association for Information Science & Technology | Oct. 29 – Nov. 1, 2022 | Pittsburgh, PA. Author(s) retain copyright, but ASIS&T receives an exclusive publication license.

To explore the relationship between article-level interdisciplinarity and scientific impact, we adopted ordinary least-squares linear (OLS) regression with robust standard errors. We used $\lg_citations$ as the new dependent variable and defined it as $\lg(citations+1)$ (Thelwall & Wilson, 2014). In order to eliminate the potential interference of other variables on scientific impact, several control variables were used in the regression model.

RESULTS

Statistics found that the DIV^* ranged from 0 to 273.92 in the sample literature, of which 90.4% had DIV^* less than 40. Figure 1 presents citation life curves of the eight groups of publications classified by DIV^* . It is notable that the higher DIV^* of one group is, the longer citation life cycles and the greater cited half-life.

To further explore the relationship between integrated interdisciplinarity indicator and citations, the quadratic prediction of citations using DIV^* were described in Figure 2. Whether it is short-term or long-term, an inverted U-shaped relationship was found between DIV^* and scientific impact.

While the inflection point of long-term citation impact is around DIV^* equal to 80, lower than that of the short-term citation impact. The OLS regression report also achieved significant results for both linear and quadratic terms. For most papers, the initial growth of interdisciplinarity has brought about a larger increase in scientific impact, but with the growth of interdisciplinarity, the growth rate of scientific impact has reduced for only a small number of papers.

As for the relationship between the three sub-indicators and scientific impact, variety, disparity, and balance shown different characteristics. Figure 3 shows the short-term quadratic prediction plots, which are similar to the long-term ones. To be specific, variety and scientific impact showed an inverted U-shaped relationship, similar to DIV^* . Balance had an overall negative correlation with citations, while disparity and citations are positively correlated. In general, with respect to interdisciplinary research, the number of disciplines (high diversity) and the distance between disciplines (high disparity) have a strong and positive impact on the scientific impact, but this impact may be offset by excessive disciplines or too even distribution of disciplines (high balance).

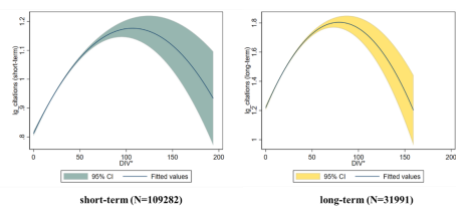


Figure 2. Relation between log-transformed scientific impact and DIV^* in quadratic prediction plots with confidence interval using the long-term and short-term logarithmic citations as the y-axis.

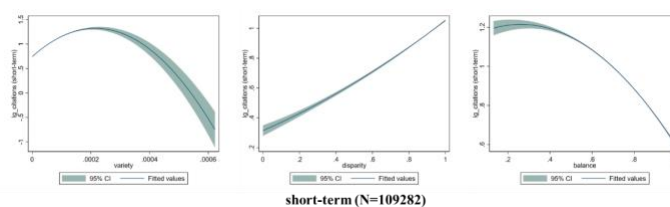


Figure 3. Relation between log-transformed scientific impact and interdisciplinary sub-indicators, including variety, disparity, and balance.

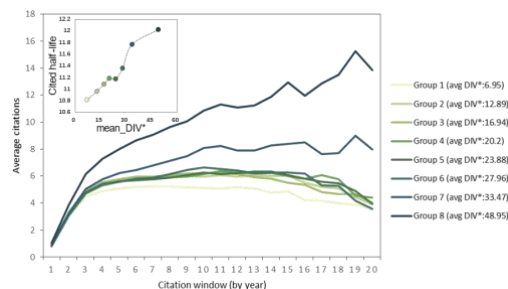


Figure 1. Annual evolution of citations and scatter plot of mean DIV^* versus cited half-life for eight groups of DIV^*

CONCLUSION

This study explored the relationship between interdisciplinarity and scientific impact in the field of climate change, using the new interdisciplinarity measure DIV^* containing three dimensions of variety, balance, and disparity. An inverted U-shaped relationship has been found between the interdisciplinary indicator DIV^* and the scientific impact, and the results of interdisciplinary sub-indicator variety have similar laws. Sub-indicator disparity had a positive relationship with scientific impact, while balance had a negative correlation. Relationships behaved similarly in the long- and short-term scientific impact.

The exploration of the mechanism of interdisciplinarity influence on climate change knowledge production not only helps enhance the output and influence of interdisciplinary research, but also provides clues for academic evaluation. The long citation life cycle of literature in the field of climate change suggested a longer evaluation period for interdisciplinary research in comparison to common research. The objective and fair academic evaluation could influence the research interests of scholars and the healthy development of academia. We will further explore the improvement of interdisciplinarity indicators and methods, and apply the methods to larger data sets for more in-depth analysis. Other impacts besides scientific impact, such as policy impact and altmetrics were planned to be discussed.

REFERENCES

- Baumgartner, S. E., & Leydesdorff, L. (2014). Group-based trajectory modeling (GBTM) of citations in scholarly literature: Dynamic qualities of “transient” and “sticky knowledge claims”. *Journal of the Association for Information Science and Technology*, 65(4), 797-811.
- Cano, V., & Lind, N. C. (1991). Citation life cycles of ten citation classics. *Scientometrics*, 22(2), 297-312.
- Chen, S. J., Arseneault, C., & Larivière, V. (2015). Are top-cited papers more interdisciplinary? *Journal of Informetrics*, 9(4), 1034-1046.
- Choi, B. C., & Pak, A. W. (2006). Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 1. Definitions, objectives, and evidence of effectiveness. *Clinical and Investigative Medicine*, 29(6), 351.
- Garfield, E. (1998). The multiple meanings of impact factors. *Journal of the American Society for Information Science*, 49(8), 768-768.
- Hellsten, I., & Leydesdorff, L. (2016). The construction of interdisciplinarity: The development of the knowledge base and programmatic focus of the journal *Climatic Change*, 1977–2013. *Journal of the Association for Information Science and Technology*, 67(9), 2181-2193.
- Larivière, V., & Gingras, Y. (2010). On the relationship between interdisciplinarity and scientific impact. *Journal of the American Society for Information Science and Technology*, 61(1), 126-131.
- Leahey, E., Beckman, C., & Stanko, T. (2017). Prominent but Less Productive: The Impact of Interdisciplinarity on Scientists' Research. *Administrative Science Quarterly*, 62.
- Leydesdorff, L., Wagner, C. S., & Bornmann, L. (2019). Diversity measurement: Steps towards the measurement of interdisciplinarity? *Journal of Informetrics*, 13(3), 904-905.
- McCright, A. M., O'Shea, B. W., Sweeder, R. D., Urquhart, G. R., & Zeleke, A. (2013). Promoting interdisciplinarity through climate change education. *Nature Climate Change*, 3(8), 713-716.
- National Academy of Sciences, N. A. o. E., & Institute of Medicine. (2005). *Facilitating Interdisciplinary Research*.
- Thelwall, M., & Wilson, P. (2014). Regression for citation data: An evaluation of different methods. *Journal of Informetrics*, 8(4), 963-971.
- Weart, S. (2013). Rise of interdisciplinary research on climate. *Proceedings of the National Academy of Sciences*, 110(Supplement 1), 3657-3664.
- Yegros-Yegros, A., Rafols, I., & D'Este, P. (2015). Does Interdisciplinary Research Lead to Higher Citation Impact? The Different Effect of Proximal and Distal Interdisciplinarity. *PLoS One*, 10(8), e0135095.
- Zhang, L., Sun, B., Jiang, L., & Huang, Y. (2021). On the relationship between interdisciplinarity and impact: Distinct effects on academic and broader impact. *Research Evaluation*.

On the Reliability of Funding Acknowledgements as Research Data: Evidence from Astronomy

Stahlman, Gretchen

Rutgers University, USA | gretchen.stahlman@rutgers.edu

ABSTRACT

Online bibliographic databases have enabled new research through which bibliographic records are analyzed as data about science. Within these records, the acknowledgements sections of papers are often used to draw conclusions about funding support for published research. While acknowledgements and funding statements can be informative for research and policy development, this poster adds to a body of literature that highlights limitations of funding data for scientometric and policy research, using evidence gathered from a questionnaire of authors of astronomy journal articles. The study shows that only 71.4% of a sample of authors of papers tied to NSF grants through acknowledgements reported in the survey that NSF funded the research presented in the respective papers. A brief analysis of the questionnaire followed by recommendations and considerations for further research are presented. The discrepancy in reporting appears to indicate that funding streams can be fluid and not always apparent to authors, overall raising the question of what sorts of research should be addressed with funding statements, where conceptually tying a paper directly to a grant is not straightforward.

KEYWORDS

research funding, scholarly communication, acknowledgements, scientometrics, astronomy

INTRODUCTION

The evolution of online bibliographic databases has enabled scientometric and policy-oriented research, where bibliographic records can be analyzed as data. Within these records, the acknowledgements sections of papers are often used to draw conclusions about research support and conflicts of interest, indirectly influencing the evolution of research policy and funding programs across disciplines. Authors typically mention funding agencies and grant numbers within acknowledgements, leading to efforts to connect specific grants and amounts of funding directly to papers to evaluate research outcomes and return on investment. A few examples of this type of research include Kurczynski & Milojevic (2020), Larivière & Sugimoto (2018), and Tatsioni, Vavva & Ioannidis (2010). While acknowledgements and funding statements can be informative, this poster explores potential limitations of using funding data for research and policy through evidence gathered from a survey of authors of astronomy journal articles.

BACKGROUND

Scholars have identified limitations of using funding acknowledgements to draw conclusions about research practices. For example, Rigby (2011) shows that there is a tendency to exaggerate productivity of certain grants. Paul-Hus, Desrochers and Costas (2016) note that “funding acknowledgement data remain self-declared information and are thus subject to unethical or inconsistent behaviours, either when authors fail to acknowledge funding sources or when, on the contrary, they acknowledge support they did not actually receive” (p. 3). As machine learning and natural language processing techniques are increasingly used to process funding information corresponding to papers, it is necessary to consider the types of research questions and methods that can benefit from funding acknowledgements. Through a dissertation study documented in Stahlman (2020) and Stahlman & Heidorn (2020), a series of incidental findings is presented here, adding insight to conversations about research with funding statements.

METHODS AND RESULTS

Using author email addresses obtained from Web of Science bibliographic records, a survey was sent to corresponding authors of astronomy papers acknowledging a sample of NSF Division of Astronomical Sciences (AST) grant numbers. The sampled grants originated in 2016, and the survey was conducted between May and June 2019. The survey was sent to 477 authors of papers associated with sampled grants, and 107 responses were received corresponding to papers published between 2016 and 2019 in 13 astronomy journals. The original purpose of the survey was to obtain information about the locations and characteristics of astronomical data that correspond to the papers, and the same survey was also sent to a second population not discussed here. As part of this larger overarching study (Stahlman, 2020), participants associated with the known sample of NSF grants through acknowledgements in the literature were nevertheless asked the multiple-choice question: *Which agency or agencies funded the research presented in this paper? Select all that apply.* Options for responses were: 1) NASA, 2) NSF, 3) DOE, 4) DOD, 5) Institutional or university support, 6) Private foundation(s), 7) International (Non-U.S.) agency, 8) Other (specify), and 9) Not applicable. Surprisingly, only 71.4% of the authors of papers linked to NSF grants through Web of Science indicated NSF as a funder of the research in the respective question (n=107).

Review of the funding statements eliminated the possibility of “false positives” originating with the Web of Science search, as all acknowledgements referred to NSF grants in some way. However, a variety of acknowledgement styles appeared across both the group that indicated NSF funding on the questionnaire and the group that did not. For example, some funding statements are quite long and detailed and others more concise. Also, some funding statements use language clearly indicating direct support of the research at hand by NSF and others indicate piecemeal individual support held by each author. Overall, a clear pattern was not detected through manual review alone.

Another possible explanation for the discrepancy is that the corresponding author who completed the survey was not immediately cognizant of all grants held by coauthors. In astronomy and other fields, the corresponding author designation is a respected role with special responsibility, and this individual is expected to be highly familiar with the research. Especially considering the freshness of sampled grants and papers at the time the questionnaire was completed, it is difficult to dismiss outright the impressions of the corresponding authors about which major funding agencies supported their papers. A binary variable was created to indicate whether a respondent from the NSF-acknowledged sample also selected NSF as a funder in response to the survey question, and significance tests were conducted (with 95% confidence interval) to further deconstruct authorship characteristics that may contribute to the discrepancy, reported briefly below.

H1: The discrepancy in reporting the funding for a paper is related to the number of authors on the paper. If a paper has many authors, it may be more difficult for the corresponding author to be aware of funding held by co-authors. The number of authors on sampled papers ranged from 2 to 122; a new log transformed variable was created and a Welch’s two sample t-test was conducted. The result was not significant at the .05 level but may be considered marginally significant and worthy of further exploration ($t = 1.8333$, $df = 58.82$, $p\text{-value} = 0.07181$).

H2: The discrepancy in reporting the funding for a paper is related to international collaboration. If a paper involves international collaboration between U.S. and non-U.S. authors, it may be more difficult for the corresponding author to be aware of funding held by coauthors. A Pearson’s Chi-squared test was conducted, and the result was not significant ($\chi\text{-squared} = 2.5885$, $df = 1$, $p\text{-value} = 0.1076$).

H3: The discrepancy in reporting the funding for a paper is related to the career stage of the corresponding author. If the corresponding author is very early in their career, they may be less familiar with funding reporting practices and other norms. To determine whether career stage of the corresponding author is significantly related to the discrepancy, a Welch’s two sample t-test was conducted, and the result was not significant ($t = 0.74725$, $df = 65.841$, $p\text{-value} = 0.4576$).

H4: The discrepancy in reporting the funding for a paper is related to whether the corresponding author is also the first author. For the present study, a majority of corresponding author respondents - 100 out of 107 - are first authors as well, which does not support statistical inference. It may be worth noting that four out of the seven non-first-authors fall into the discrepancy category (57%). This percentage is higher than the larger dataset (28.6%), but there are too few observations to draw a conclusion.

DISCUSSION AND CONCLUSION

The incidental findings and brief exploration presented here demonstrate that funding streams in astronomy are fluid and not always apparent to authors. This conclusion aligns with the argument of Rigby (2011) that links between papers and acknowledged funding are complicated and often indirect, where research funding essentially supports an ecosystem of processes rather than specific papers. By directly obtaining authors’ perspectives through a questionnaire about specific papers, the present study has illuminated discrepancies in awareness and reporting of funding sources. That said, possible reasons were identified for the discrepancy – the most promising explanation (within the limitations of the study) being that having more authors on a paper may contribute to enhanced ambiguity about research support. Features of astronomical research are unique and could contribute to the discrepancy as well, as authors may or may not choose to acknowledge the grant numbers of NSF-funded facilities and instruments, and where investments in primary research are balanced with research support-related funding for training, major facilities, instrumentation, and software (Stahlman & Heidorn, 2020).

Beyond the small study in one discipline presented here, the overall complexity of funding acknowledgements demonstrated in the broader literature warrants continued caution when using funding statements to generalize. This issue raises the question of what sorts of research can or should be addressed with funding statements, where conceptually tying papers directly to grants is not straightforward. Future work will further explore of the quality and integrity of acknowledgements as a data source, to avoid natural language pitfalls with automatic extraction and ensure accurate reporting of research outputs to justify funding and assess the overall value and impact of scientific research. These issues also point to a need for further qualitative research on the nuances of acknowledgement behavior and impressions of authors across disciplines about direct and indirect funding support for papers.

REFERENCES

- Kurczynski, P. L., & Milojevic, S. (2020). Enabling discoveries: A review of 30 years of advanced technologies and instrumentation at the National Science Foundation. *Journal of Astronomical Telescopes, Instruments, and Systems*, 6(3), 030901. <https://doi.org/10.1117/1.JATIS.6.3.030901>
- Larivière, V., & Sugimoto, C. R. (2018). Do authors comply when funders enforce open access to research? *Nature*, 562(7728), 483–486. <https://doi.org/10.1038/d41586-018-07101-w>
- Paul-Hus, A., Desrochers, N., & Costas, R. (2016). Characterization, description, and considerations for the use of funding acknowledgement data in Web of Science. *Scientometrics*, 108(1), 167–182. <https://doi.org/10.1007/s11192-016-1953-y>
- Rigby, J. (2011). Systematic grant and funding body acknowledgement data for publications: New dimensions and new controversies for research policy and evaluation. *Research Evaluation*, 20(5), 365–375. <https://doi.org/10.3152/095820211X13164389670392>
- Stahlman, G. R. (2020). *Exploring the long tail of astronomy: A mixed-methods approach to searching for dark data* (Doctoral dissertation, The University of Arizona). <https://www.proquest.com/docview/2435763825/abstract/8E65945F493F45AAPQ/1>
- Stahlman, G. R., & Heidorn, P. B. (2020). Mapping the “long tail” of research funding: A topic analysis of NSF grant proposals in the division of astronomical sciences. *Proceedings of the Association for Information Science and Technology*, 57(1), e276. <https://doi.org/10.1002/pr2.276>
- Tatsioni, A., Vavva, E., & Ioannidis, J. P. A. (2010). Sources of funding for Nobel Prize-winning work: Public or private? *The FASEB Journal*, 24(5), 1335–1339. <https://doi.org/10.1096/fj.09-148239>

United for Justice? A Critical Review of Social Justice Research in Information Science and Technology

Tacheva, Zhasmina
Namvarrad, Sepideh

School of Information Studies, Syracuse University, USA | ztacheva@syr.edu
Department of Psychology, Syracuse University, USA | snamvarr@syr.edu

ABSTRACT

As calls for justice in science and technology grow louder, the number of studies that seek to center equity and justice continues to grow. While the resulting body of knowledge is commonly assumed to constitute a monolith in terms of its theories, methods, priorities, and communities, the purpose of this study is to demonstrate the diversity of the extant justice-oriented research in information science and technology. Specifically, we show that, similar to other emerging research streams, as certain justice-oriented works become well-known and enter the mainstream, the margins once again shift and new, even more progressive ideas appear on the epistemological horizon, fighting for visibility. This study argues that instead of becoming too comfortable with what has now become the justice-oriented mainstream, the information science and technology literature should learn from the vanguard ideas in this area, some of which are presently being silently neglected.

KEYWORDS

Social Justice; Justice-oriented Research; Bibliometric Analysis; Critical Theory

INTRODUCTION

The field of information science and technology (IST) is becoming increasingly aware of the importance that the introduction of social justice as both a research topic and a tool into the core of the discipline has for the survival of the entire field (Cooke, Sweeney, & Noble, 2016). The question this study seeks to elucidate is therefore, *What are the different components of justice-oriented work in information science and technology and how are they related?*

Addressing this question can help not only situate justice-oriented research as a pivotal part of the discipline, but also provide critical insights into its future development, since a more nuanced understanding of the work each community under the umbrella of “IST justice” is engaged in is an important indicator of the future research direction of the field. Using bibliometric analysis enables us to illuminate the structural aspects of justice-oriented research by examining how different works and subject areas are related (Ellegaard & Wallin, 2015).

METHOD

The study sample was collected by performing a series of keyword searches on Web of Science pertaining to social justice-oriented work in information science and technology. A team of an IST professor and doctoral student in computational psychology assembled a list of 29 key phrases such as “critical data studies,” “feminist STS,” and “neuroqueer HCI,” based on a survey of both the extant literature and other subject matter experts. The list entries were purposely concrete and yet broad enough to capture as many of the justice-oriented works in the field as possible. While alternative data collection approaches exist, the number of relevant studies retrieved using the process outlined above was much greater than, for instance, searching for phrases such as “justice AND information science.” Specifically, the number of unique justice-oriented studies in our sample is 1,150 (these comprise the *focal* nodes in the network) with a total of 41,025 cited references (i.e., *secondary* nodes). Table 1 summarizes the key descriptive statistics of the resulting directed network where focal nodes serve as “sources” and the works they cite as “targets,” connected by a total of 49,955 weighted edges.

Network Metric	Min	Mean	Max
In-degree	0	1.18	86
Out-degree	0	1.18	350
Betweenness	0	2.16	8,316.17
Closeness	0	0.02	1
Connected component size	1	278	39,665

Table 1. Justice-oriented Research Network Statistics

The work referenced by the largest number of studies (86), i.e. the node with greatest in-degree centrality, is boyd and Crawford’s (2012) “Critical Questions for Big Data” – one of the first studies to explicitly critique the computational turn in IST (boyd & Crawford, 2012). Conversely, the work citing the most studies, i.e. the node with highest out-degree centrality, is Harris and Anthis’ (2021) comprehensive literature review titled “The moral

85th Annual Meeting of the Association for Information Science & Technology | Oct. 29 – Nov. 1, 2022 | Pittsburgh, PA. Author(s) retain copyright, but ASIS&T receives an exclusive publication license.

consideration of artificial entities” which has a total of 350 references (Harris & Anthis, 2021). The node with highest betweenness centrality, or the one best positioned to connect different communities within the network, is Floridi’s (2016) “Faultless Responsibility: On the Nature and Allocation of Moral Responsibility for Distributed Moral Actions” (Floridi, 2016), followed by Mittelstadt and Floridi’s (2016) “The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts” (Mittelstadt & Floridi, 2016), which also happens to be the paper with highest closeness centrality, i.e. the one with highest average closeness to all other papers in the network. Although there are 151 connected components in the network, the largest one accounts for 94 percent of all nodes, serving as a main component with a very small number of papers outside its orbit. This may suggest a highly cohesive and endogenous network, so to investigate the structure of justice-oriented literature in IST further, we added another level of analysis by aggregating the papers in the network to the thematic categories which we used for data collection from Web of Science.

RESULTS

The thematic level of analysis allows us to explore the connections between different themes within justice-oriented research and test our expectation of a diverse and distributed rather than a centralized and homogenous bibliometric network. Figure 1 shows a graph of the theme-level network where each node represents one of the 29 themes in our search list, node color corresponds to connected component membership, and node size is indicative of degree centrality.

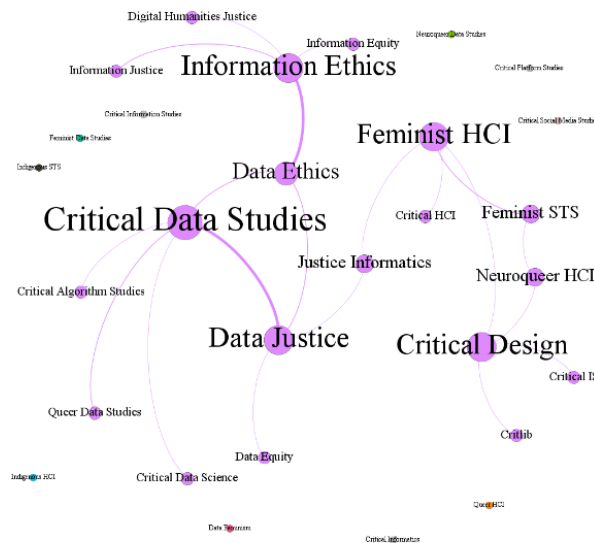


Figure 1. Theme-level Justice-oriented Research Network

While 19 of the 29 thematic areas (65 percent of all themes) are indeed connected, i.e. authors working in one area tend to reference authors active in other areas or even co-list these additional areas in their own work, certain areas in justice-oriented IST work, such as queer and Indigenous HCI, remain largely underexplored and disconnected from the mainstream IST literature. Even the main component itself does not exhibit a high degree of centralization, with data and information ethics and critical data studies forming one pole, and the justice-oriented branches of human-computer interaction (HCI), along with critical design, a second one, with very few interactions in-between.

CONCLUSION

This preliminary analysis can serve as initial evidence that far from being a monolith, justice-oriented research in the information sciences is fragmented, mirroring the underlying silos in IST. Although this is not surprising, justice-oriented researchers can benefit from venturing out of their traditional understanding of justice work and learning from adjacent communities. It is, of course, not feasible to expect that scholars from one subject area be fluent in other, often vastly different IST subfields; however, there are principles around which many of us in IST can unite. These principles can serve as a foundation for “re-imagining an information-resilient society,” and existing ASIS&T special interest groups can facilitate inter-SIG dialogue to build that foundation collectively.

Future work will broaden the scope of the study by expanding both the search terms used for data collection and the research databases surveyed. Additionally, a close content analysis of the papers in each thematic area can surface the key commitments to justice in the surveyed research communities which can help chart the future directions of the field.

REFERENCES

- boyd, d., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, 15(5), 662-679.
- Cooke, N. A., Sweeney, M. E., & Noble, S. U. (2016). Social justice as topic and tool: An attempt to transform an LIS curriculum and culture. *The Library Quarterly*, 86(1), 107-124.
- Ellegaard, O., & Wallin, J. A. (2015). The bibliometric analysis of scholarly production: How great is the impact? *Scientometrics*, 105(3), 1809-1831.
- Floridi, L. (2016). Faultless responsibility: On the nature and allocation of moral responsibility for distributed moral actions. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2083), 20160112.
- Harris, J., & Anthis, J. R. (2021). The moral consideration of artificial entities: a literature review. *Science and engineering ethics*, 27(4), 1-95.
- Mittelstadt, B. D., & Floridi, L. (2016). The ethics of big data: current and foreseeable issues in biomedical contexts. *The ethics of biomedical big data*, 445-480.

Towards a Diversity, Equity, and Inclusion (DEI) Competency Framework: A Content Analysis of Visible DEI Efforts on LIS Associations' Websites

Tang, Rong

Simmons University, USA | rong.tang@simmons.edu

Hu, Zhan

Simmons University, USA | zhan.hu@simmons.edu

Pallo, Bianca

Simmons University, USA | bianca.pallo@simmons.edu

ABSTRACT

In this poster, we report the findings of an investigation of 28 LIS associations' DEI efforts as evidenced through their websites. The DEI efforts were documented in terms of whether DEI was mentioned in associations' missions and goals, their code of conducts, and their competency frameworks, whether the associations provided DEI definitions, whether there were position statements, whether there were dedicated units leading the DEI efforts, whether there were specific DEI resources, and whether there were relevant contents on the homepage or as menu items. While we discovered the common patterns of DEI practice, we also uncovered notable gaps. Our findings contribute to a better understanding of the role of associations in promoting and fostering DEI, it also shows the need for a consistent and visible effort for LIS associations to lead their professional community toward a robust work in DEI and developing a DEI competency framework for LIS professionals.

KEYWORDS

Diversity, Equity, and Inclusion (DEI); LIS Associations; Mission, Vision, and Strategic Plans; Committees and Task Forces; Competencies

INTRODUCTION

According to *The SAGE Encyclopedia of Higher Education*, a professional association is “an entity comprised of individuals who are connected with one another through a shared mission, vision, or purpose concerning a specific profession. Professional associations exist to advance the interests of a profession, the individuals engaged in that profession, and to serve the public good” (Soto, 2020). In recent years, a variety of library and information science (LIS) associations have reaffirmed their commitment to and/or declared their positions in supporting and promoting efforts concerning Diversity, Equity, and Inclusion (DEI). Nevertheless, the exact role of the associations in fostering elements of DEI in its membership community, and incorporating or integrating DEI elements into professional competency standards, has largely been unexamined. As noted by the Public Library Association (PLA), “We acknowledge that libraries have been—and still are—complicit in systems that oppress, exclude, and harm Black people, indigenous people, and people of color (BIPOC). The library profession remains overwhelmingly white, despite decades of emphasis on diversity and inclusion. We see incredible examples of self-determination and resilience by BIPOC librarians and educators, yet the profession has largely failed to improve conditions and ensure pathways for advancement among library workers of color.” Poole et al. (2021) observe that “the field of LIS continues to face a vexing paradox. Its longstanding ideal of and concomitant commitment to serving diverse communities and users equally has failed to translate into diversity, equity, and inclusion (DEI) in the profession or in LIS education” (p. 258). While several research studies have focused on DEI in LIS education (e.g., Montiel-Overall, 2009; Jaeger et al., 2014; Poole, et al., 2021), few have investigated the role that LIS associations play in promoting and supporting DEI in LIS practice. A study by Adkins et al. (2020) focuses on DEI efforts by ALA, the ethnic caucuses of ALA, and SAA. Adkins et al.'s work did not explore how DEI knowledge, skills, or abilities (KSA) are a part of competency standards set by a given association for professional practices. We believe it is not only crucial for associations to declare it as a part of overarching missions and core values, but it is also essential that DEI KSA is explicitly outlined in professional competency frameworks developed by these associations. Our research questions (RQs) are as follows:

RQ1. What parts of their websites do associations use to visibly declare the importance of DEI for their organizations?

RQ2. In terms of organizational structure, what are the dedicated units in the associations that lead/direct the DEI initiatives/efforts? What associations seem to miss those dedicated units?

RQ3. What are the notable gaps in the Associations' DEI efforts judging from their public-facing websites?

METHODOLOGY

Upon examining multiple resources of LIS associations, we based our study samples primarily on the list provided by the University of North Texas College of Information. By removing all Texas-based associations and adding several key associations, we had 28 LIS associations located in the US and Canada. Examples of these associations include American Library Association (ALA), American Medical Informatics Association (AMIA), and Canadian

Association of Research Libraries (CARL). We reviewed association websites and recorded the details of the DEI related content in the aspects of (1) Mission, vision, core values, and strategic plans, (2) DEI definitions, (3) DEI position statements, (4) Code of ethics/conducts/ethical principles, (5) Dedicated units, (6) Competency framework, (7) Resources, and (8) DEI information presence on the homepage or as menu items. The sampling, data collection, and coding took place from November 22, 2021 to May 23, 2022.

RESULTS

The Presence of DEI Efforts on Websites

Of 28 associations, the most frequent way of addressing DEI issues was through the associations' Mission Statement, Vision Statement, Strategic Plan, or Core Value Statement (n=24, 85.71%). Twenty-three associations (82.15%) had diversity statements, whereas 19 (67.86%) had a dedicated unit to coordinate the DEI efforts. Nineteen (67.86%) associations had resources on DEI, while 17 (60.71%) associations' code of ethics contained DEI related wording. However, when it comes to competencies, less than half (n=13, 46.43%) of the associations' competency standards mentioned or had a subarea relevant to DEI. Other efforts such as having a menu item, having a definition, or having a dedicated section on their home pages are much less common, ranging from 11% to 43%.

Dedicated DEI Units within Associations

Nineteen (70.37%) associations had dedicated units leading the DEI efforts and of them, 8 (42.11%) had multiple dedicated units, whereas 11 (57.89%) had one. Twelve (63.16%) had a committee for DEI, 8 (42.11%) had a task force, whereas 6 (31.58%) had a SIG, section, or community for DEI. Examples of committees include PLA's Committee on Equity, Diversity, Inclusion and Social Justice, and MLA's Diversity, Equity, and Inclusion Committee. Examples of task forces are ARL's Task Force on Structural Equity and Inclusion, and AMIA's DEI Task Force. Examples of a SIG or community are SLA's Diversity Inclusion Community Equity (DICE) Community, and ALISE's Equity and Social Justice SIG.

Notable Gaps in DEI Efforts

Notable gaps appear in several areas, including lacking a "homepage dedicated section" to publicize associations' DEI efforts or resources (n=25), missing explicit "definitions of DEI terms" (n=19), not listing DEI initiatives as "menu items" (n=16), and no explicit "DEI competencies" (n=15). Three associations had a *dedicated homepage section* on DEI, including ALA, ARL, and IEEE. For instance, ARL featured a DEI block on the homepage that directly links to the dedicated DEI page that lists all DEI-related news, programs, and resources. Nine associations provided *explicit definitions* of DEI: ALA, ACRL, ALISE, AASL, MLA, SAA, ACM, IEEE, and CARL. With the absence of an explicit and meaningful definition of DEI, the associations may face problems of bringing their members on the same page about the fundamental concepts of DEI.

Meanwhile, 12 association websites featured *DEI-related menu items* that link to further information or resources. While some associations had DEI at the first level of the top menu, others had DEI on the third level of the main menu. For instance, ACM has "Diversity, Equity, and Inclusion" items on its homepage top menu, whereas nine associations (e.g., ALA, ACRL, ARL, AMIA, SLA) placed "DEI" as second-level menu items and two (IFLA, AALL) as the third level item. Thirteen associations integrated DEI in their *professional competency standards* or had separate standards in varying topics related to DEI. Five (AASL, ALA, ALCTS, CARL, YALSA) had an independent DEI Competency area. An example for separate standards is the "Cultural Proficiencies for Racial Equity," co-developed by PLA, ARL, ACRL, and ALA. ACRL has diversity standards. AASL has a "Shared Foundation: Include" framework which entails "an understanding of and commitment to inclusivity and diversity in the learning community."

DISCUSSION & CONCLUSION

As asserted on the PLA EDISJ Committee's webpage, "We commit to dismantling white supremacy in libraries and librarianship. We recognize the urgency of this collective work, and commit to hold ourselves, our colleagues, and our institutions accountable when we fall short." In this investigation, we uncovered the patterns of DEI efforts as shown through the LIS associations' websites. We also revealed notable gaps in associations' DEI efforts. While it is not surprising that many associations (85.71%) had DEI explicitly stated in their Mission, Vision, Goals, or Values, it is alarming that there were four associations that did not include DEI in any of their central statements. While the most commonly dedicated unit for leading the DEI effort is through committees (42%), other associations used more short termed task forces or working groups. Not all the associations had a dedicated competency area in DEI, which, we declare, should be essential for the development of LIS professionals. The findings of our research contribute to a better understanding of the role of associations in promoting and fostering DEI. Through this study, we urge the LIS professional associations to make a consistent and visible effort to lead their professional community towards robust work in DEI and developing a DEI competency framework for LIS professionals.

REFERENCES

- Adkins, D., Buchanan, S. A., & Alston, J. K. (2020). LIS association activities in promoting and sustaining an inclusive profession. *Library Quarterly*, *90*(2), 162–172. <https://doi-org.ezproxy.simmons.edu/10.1086/707672>
- College of Information, University of North Texas. (n.d.). *Professional Organizations*. <https://informationscience.unt.edu/professional-organizations>
- Jaeger, P. T., & Franklin, R. E. (2007). The virtuous circle: Increasing diversity in LIS faculties to create more inclusive library services and outreach. *Education Libraries*, *30* (1), 20-26.
- Montiel-Overall, P. (2009). Cultural Competence: A conceptual framework for library and information science professionals, *Library Quarterly* *79* (2): 189–90.
- Soto, M. (2020). Professional and research associations. In M. E. David, & M. J. Amey (Eds.), *The SAGE encyclopedia of higher education*. Sage UK.
- Poole, A. H., Agosto, D., Greenberg, J., Xia Lin, & Erjia Yan. (2021). Where Do We Stand? Diversity, Equity, Inclusion, and Social Justice in North American Library and Information Science Education. *Journal of Education for Library & Information Science*, *62*(3), 258–286. <https://doi-org.ezproxy.simmons.edu/10.3138/jelis.2020-0018>
- Public Library Association, Committee on Equity, Diversity, Inclusion and Social Justice. (April 7, 2017). <http://www.ala.org/pla/about/people/committees/pla-tfedi>

Rethinking Algorithmic Fairness in the Context of Information Access

Udoh, Emmanuel Sebastian

University at Albany, New York, USA | eudoh@albany.edu

Yuan, Xiaojun

University at Albany, New York, USA | xyuan@albany.edu

Rorissa, Abebe

University of Tennessee, Knoxville, USA | arorissa@utk.edu

ABSTRACT

The internet has revolutionized the information seeking landscape, heightening expectations that access to information will hasten the attainment of several of the United Nation (UN)'s sustainable development goals (SDGs), although digital and content divides persist and information access is demonstrably steeped in inequalities, partly due to underlying intricate, complex, and often opaque AI/ML algorithms. While equity in information access remains a long shot goal, this poster asks the fundamental question: can access to information at least be fair? And what exactly does algorithmic fairness mean in the context of information access? Drawing from themes in information search and social informatics, especially the political valence in technologies such as online decision and filtering algorithms, this poster aims to analyze current nuances of algorithmic fairness in order to identify those intricate issues that should inform and constitute any working definition and framework to study algorithmic fairness in the context of information access in an increasingly globalized world.

KEYWORDS

Algorithmic fairness, information access, algorithmic bias, fair ranking, fair recommendation system

INTRODUCTION

Recent advances in artificial intelligence (AI) and machine learning (ML) have thrust algorithmic fairness into the front burner of the discourse on algorithms and algorithmic decision systems. The new technologies tend to digitalize human bias and digitally redline minority demographics, people of color, disabled persons, women, older adults (Benjamin, 2019; Noble, 2018; Eubanks, 2016; O'Neil, 2016), and in an information access sense, the global south. So far, while much of the algorithmic fairness discourse strives to conceptualize or demystify the underlying algorithmic opacity, one recurrent theme in the discourse is the simplistic conceptualization and definition of fairness as the absence of intentional bias in the implementation of algorithmic decision systems and technologies (Mehrabi et al., 2022). In the light of new realities and novel studies on specific domains of application, it is worth rethinking the traditional definition of algorithmic fairness and asking whether the definition is adequate and appropriate across multiple domains. Unfortunately, there is not yet enough discussion about algorithmic fairness in the context of information access. To address the above issues, this poster addresses the **Research Question**: How should we comprehensively define and study algorithmic fairness in the context of information access? This poster utilized a PRISMA-style review (Moher et al., 2009) to analyze current nuances of algorithmic fairness with the goal of identifying those intricate issues and overlapping themes that should inform and constitute any definition of and framework for algorithmic fairness in the context of information access.

METHOD

The Web of Science, ACM Digital Library, and Academic Search Complete (Ebscohost) databases were searched through three rounds to unearth peer-reviewed articles written in English and relevant for the review. We searched predetermined keywords with the query: Information AND (retriev* OR system*) AND (search* OR engine) AND Fair* AND ("Information access" OR "Information Search*" OR "Information Seek*" OR "Information use*"). This was followed by screening of retrieved titles and abstracts, using predetermined inclusion/exclusion criteria, and the full text of selected articles to ensure that they met the same inclusion/exclusion criteria. Our search produced a list of 11, 59, and 34 results from the Web of Science, ACM Digital Library (Full-text collection), and the Academic Search Complete (EBSCOHOST), respectively. Four of the articles were duplicates, leaving a total of 102 final list of non-duplicate results. The authors then screened the titles and abstracts of the final list independently, a process that resulted in the removal of 62 articles that were outside of the scope of this review, leaving only 40 articles to be subjected to further screening that produced 38 articles in our final sample.

RESULTS

The findings revealed at least seven themes: fairness issues involved in the discourse, fairness solutions proffered, technologies involved in the fairness discourse, the dimensions of fairness, metrics used to assess fairness, the stakeholders and persons or groups impacted by the fairness issue in question, and the implementation context. The first major theme is the type of fairness issue addressed. The most widely mentioned types of (un)fairness include bias, representational harms, misrepresentation, stereotyping, gender bias, allocational harms (Rekabseaz et al. 2021), and oppression. The second group of fairness issues concerns access to the desired resources and includes

85th Annual Meeting of the Association for Information Science & Technology | Oct. 29 – Nov. 1, 2022 | Pittsburgh, PA. Author(s) retain copyright, but ASIS&T receives an exclusive publication license.

the concepts of unfair ranking, relevance judgments, internet neutrality, digital divide, disability-based digital divide (Yang & Chen, 2015), content divide, inaccessibility, access to the internet (Yang & Chen, 2015), web accessibility for older adults (Yang & Chen, 2015), and censorship. One author (Joachims, 2021) also distinguished between exogenous and endogenous biases. These sub-themes are related in their common tendency to exclude some persons or groups or make it hard for such groups to participate in some benefits. We also found the third genre of literature comprising suggested models that promote algorithmic fairness: SoftRank models, Stochastic ranking, Human centered and Proactive search systems design that uses the implicit primary task context to model the user's search intent (Koskela et al., 2018), task-based information retrieval (Koskela et al., 2018), and the CLEVER (Client-side EigenVector Enhanced Retrieval) search system developed at IBM (Kumar et al., 2006). In this group also belongs an approach based on Latent Dirichlet Allocation (LDA) for recommending tags of resources to improve search (Krestel et al), Cost-per-action (CPA) Equalization (Rey & Kannan, 2010), the generation of mediatory summarization that facilitates users' assessments of the credibility of information they find on the web (Shibuki et al., 2010), and interactive retrieval model (Koskela et al., 2018). The fourth group comprises those focused on the technologies involved in the fairness discourse. Some of these include search engines, information retrieval systems, commercial platforms, library databases, social platforms, and tagging systems (Krestel et al., 2009). Other technologies include social bookmarking systems (Krestel et al., 2009), anchor text, hyperlinks, linear algebra, link analysis, and web search (Kumar et al, 2006), distributed clustering systems (Lamprier et al., 2010), web content indexing and ranking (Li et al., 2014), and topical crawlers (Menczer et al., 2004). While these technologies share fairness vulnerabilities, sometimes their peculiar environments also suggest variations in the types of (un)fairness possible. Our fifth literary genre addresses the metrics used to determine or evaluate algorithmic fairness. common fairness metrics in literature include group fairness (Rekabsaz et al., 2021), individual fairness, protected attributes fairness, provider group fairness, fairness relating to disability laws (Yang & Chen, 2015), expected exposure, and exposure allocation (Joachims, 2021). Also in this group are precision, effectiveness, recall, the level of document preprocessing and reproducibility (Roy et al, 2018), retrieval results fairness, website readability, and information quality on website (Hamwela et al., 2018). There are also articles on the probability ranking principle (Joachims, 2021), and the cluster hypothesis that relevant documents tend to be more like each other than to non-relevant ones (Lamprier et al., 2010). This genre of articles manifest and share various dimensions of algorithmic fairness which forms the focus of our sixth literary genre comprising of articles that more explicitly mention or address the dimensions of algorithmic fairness discourse. Such dimensions include the legal (Joachims, 2021; Yang & Chen, 2015), public policy, effects on the items or how rankings arbitrate exposure and thus economic opportunity, how the ranking systems shape incentives and the long-term effectiveness of markets (Joachims, 2021), and retrieval effectiveness (Lewandowski & Spree, 2011). We can also speak of individual fairness (where similar individuals are treated similarly) and group fairness where groups identified by sensitive or protected attributes (e.g., gender, ethnicity, age, etc.) are treated similarly as well. Finally, our seventh literary genre is composed of articles addressing the context of information access. Accordingly, the common themes include information behavior, information search, information seeking, information retrieval, information use, and the information user.

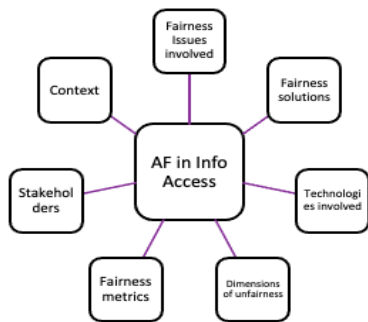


Figure 1. Main groups of fairness concepts in the reviewed literature

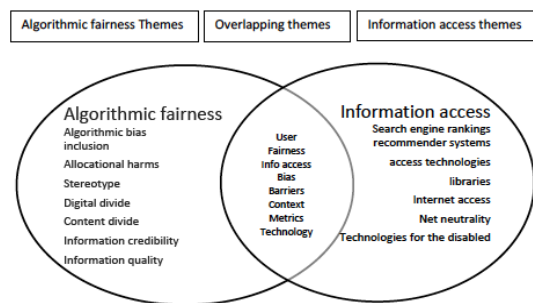


Figure 2. Overlapping concepts in the reviewed literature

Results show the following overlapping concepts and themes overlap between algorithmic fairness and information access: The user, fairness, access to information, barriers to information, context, bias, fairness assessment metrics, and information technology (See Figure 2).

CONCLUSION

This poster used a PRISMA literature review to assess the concepts employed by authors in the discourse on algorithmic fairness in relation to the domain of information access. Results indicate that current conceptualizations of algorithmic fairness in this domain revolve around seven literary genres. In future work, we intend to build on these early findings by designing a framework for comprehensively defining algorithmic fairness in the context of information access.

REFERENCES

- Benjamin, R. (2019). *Race after Technology: Abolitionist Tools for the New Jim Code*. Cambridge, Mass.: Polity Press.
- Bryson, J. (2004). What to do when stakeholders matter: Stakeholder identification and analysis techniques. *Public Management Review* 6(1); pp.21-53.
- Eubanks, V. (2016). *Automating Inequality: How High-Tech Tools profile, Police, and Punish the Poor*. New York: St. Martin's Press.
- Gao, R., Shah, C. (2020). Toward creating a fairer ranking in search engine results. *Information Processing & Management* Vol. 57 Issue 1, 102138.
- Hamwela, V., Ahmed, W., Bath, P. A. (2018). Evaluation of websites that contain information relating to malaria in pregnancy. *Public Health (Elsevier)* Vol. 157, pp. 50-52.
- Joachims, T. (2021, July). Fairness and Control of Exposure in Two-Sided Markets. *Proceedings of the 2021 ACM SIGIR International Conference on Theory of Information Retrieval*, p.1.
- Koskela, M., Luukkonen, P., Ruotsalo, T., Sjöberg, M., Floreen, P. (2018). Proactive Information Retrieval by Capturing Search Intent from Primary Task Context. *ACM Transactions on Interactive Intelligent Systems* Vol.8 Issue 3, pp. 1–25.
- Krestel, R., Fankhauser, P., Nejdl, W. (2009). Latent Dirichlet Allocation for Tag Recommendation. *Proceedings of the Third ACM Conference on Recommender Systems*, pp. 61–68.
- Kumar, R., Raghavan, P., Rajagopalan, S., Tomkins, A. (2006). Core Algorithms in the CLEVER System. *ACM Trans. Internet Technology* Vol 6 Issue 2, pp. 131–152.
- Lamprier, S., Amghar, T., Saubion, F., Levrat, B. (2010). Traveling among Clusters: A Way to Reconsider the Benefits of the Cluster Hypothesis. *Proceedings of the 2010 ACM Symposium on Applied Computing*, pp. 1774–1780.
- Lewandowski, D., Spree, U. (2011). Ranking of Wikipedia articles in search engines revisited: Fair ranking for reasonable quality? *Journal of the American Society for Information Science & Technology* Vol. 62, Issue 1, pp.117-133.
- Li, Nan; Anderson, Ashley A.; Brossard, Dominique; Scheufele, Dietram A. (2014). Channeling Science Information Seekers' Attention? A Content Analysis of Top-Ranked vs. Lower-Ranked Sites in Google
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2022). A Survey on Bias and Fairness in Machine Learning. <https://arxiv.org/pdf/1908.09635.pdf>
- Menczer, F., Pant, G., Srinivasan, P. (2004). Topical Web Crawlers: Evaluating Adaptive Algorithms. *ACM Trans. Internet Technology* Vol.4, Issue 4, pp. 378–419.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). The PRISMA Group (2009) Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA statement. *PLOS Medicine*, 6(7): e1000097. Retrieved from: <https://doi.org/10.1371/journal.pmed.1000097>.
- Mukherjee, S., Weikum, G. (2015). Leveraging Joint Interactions for Credibility Analysis in News Communities. *Proceedings of the 24th ACM International Conference on Information and Knowledge Management*, pp. 353–362.
- Noble, S. (2018). *Algorithms of Oppression: Race, Gender and Power in the Digital Age*. New York: NYU Press.
- O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.
- PRISMA review. PRISMA (prisma-statement.org). Moher, D., Altman, D. G., Liberati, A., & Tetzlaff, J. (2011). PRISMA statement. *Epidemiology*, 22(1), 128.
- Rekabsaz, N., Kopeinik, S., Schedl, M. (2021). Societal Biases in Retrieved Contents: Measurement Framework and Adversarial Mitigation of BERT Rankers. *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 306–316.
- Rey, B., Kannan, A. (2010). Conversion Rate Based Bid Adjustment for Sponsored Search. *Proceedings of the 19th International Conference on World Wide Web*, pp. 1173–1174.
- Shibuki, H., Nagai, T., Nakano, M., Miyazaki, R., Ishioroshi, M., Mori, T. (2010). A Method for Automatically Generating a Mediator Summary to Verify Credibility of Information on the Web. *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pp. 1140–1148.
- Sun, Z., Di, L., Heo, G., Zhang, C., Fang, H., Yue, P., Jiang, L., Tan, X., Guo, L., Lin, L. (2017). GeoFairy: Towards a one-stop and location based Service for Geospatial Information Retrieval. *Computers, Environment and Urban Systems*, Vol. 62, March 2017, pp. 156-167.
- Yang, Y. T., Chen, B. (2015). Web Accessibility for Older Adults: A Comparative Analysis of Disability Laws. *Gerontologist* Vol. 55 Issue 5, pp. 854-865.

How Does Awe Fuel Information Seeking? A Mixed-methods, Virtual Reality Study

Urban, Alex

University of Missouri - Columbia, USA | urbanac@umsystem.edu

ABSTRACT

Awe experiences may increase awareness of knowledge gaps and promote inquiry. However, the relationship between awe and exploratory behavior, such as information seeking, remains unclear. To explore awe as an antecedent for information seeking, participants ($n = 17$) were exposed to a variety of virtual reality (VR) awe elicitors, and their quantitative and qualitative responses were gathered. One day later, participants responded to follow-up questionnaires. Then, within one week, a stratified sample ($n = 4$) completed interviews. To determine which awe elicitors were the most curiosity-provoking, the study utilized descriptive statistics as well as mixed-effects logistic regressions. The study also included interpretative phenomenological analysis of the qualitative laboratory responses and interview transcripts. The findings suggest that awe-inspiring VR content may be a valuable technique to promote intrinsic motivations for inquiry, but future design research is necessary to bridge the gap between intent and behavior.

KEYWORDS

Awe, Emotions, Learning, Information seeking behavior, Virtual reality

INTRODUCTION

Awe is a sense of enormity that alludes comprehension. Because awe elicitors prevent assimilation of information into pre-existing schema (Keltner & Haidt, 2003), this emotion may prompt exploration of the wonder in question—whether it be a tornado, a painting, or a mathematical theory. This study explores the relationship between awe and curiosity as well the use of VR to foster information seeking. Although the link between awe and curiosity appears in both informal and formal learning environments, little empirical research has focused on how specific awe elicitors promote self-directed exploration.

This preliminary study asks the following question: *How, if at all, do virtual awe elicitors foster curiosity and exploration?*

This study provides insight into the specific qualities of VR content that activate awe, curiosity, and intentions to satisfy knowledge gaps. If virtual awe elicitors foster intrinsic motivation for information seeking, educators may find value in incorporating awe-inspiring VR content within their curricula.

LITERATURE REVIEW

Because of awe's stimulus-driven processing and outward orientation (Fiedler, 2001; Keltner & Haidt, 2003; Piaget, 1970, 1973; Shiota et al., 2007), it may be considered part of a broader group of knowledge emotions. Knowledge emotions are affective states involved in the knowledge acquisition process (Keltner & Shiota, 2003; Morton, 2009; Schindler et al., 2017). Anderson and colleagues (2020) propose that a primary function of awe is to produce curiosity, another knowledge emotion. Moreover, Anderson and colleagues suggest that the link between awe behaviors and emotional states is information seeking and learning about physical and social environments. In sum, there are theoretical connections between awe and curiosity but limited empirical investigation on how or if awe leads to exploratory behavior, such as information seeking.

METHODS AND ANALYSIS

In a laboratory setting, undergraduate participants ($n = 17$) viewed six 360° videos with a VR head-mounted display. These videos represented a variety of awe-inspiring qualities through topics like (1) the sun's magnetic fields, (2) destruction from a hurricane, (3) a grove of mysteriously bent trees, (4) a data storage bunker, (5) the aftermath of a nuclear disaster, and (6) gothic boxwood sculptures. All videos (totaling 10 minutes of content) contained narration as well as brief contextual information. After exposure to the stimuli, participants responded to two seven-point scales to first determine which scenes prompted the greatest levels of awe (Shiota et al., 2007) and curiosity. As there are no well-validated measures for state curiosity (Litman, 2017), the study used a simple instrument that asked participants how curious they were about each stimulus. Then, after describing which scenes were the most interesting, participants reported their thoughts and questions about the elicitors and whether they had any intentions of searching for answers.

Then, one day later, participants completed the same scales and responded to two qualitative prompts regarding (1) which thoughts or questions persisted over time and (2) whether they searched for information. Next, within one week of the follow-up questionnaire, a stratified sample of these participants ($n = 4$) completed interviews.

To determine which elicitors provoked the greatest curiosity, the study utilized descriptive statistics as well as mixed-effects logistic regressions. To examine the qualitative responses and interview transcripts, analysis followed the iterative, six-step, interpretative phenomenological procedures outlined by Smith and colleagues (2009).

FINDINGS

Participants found the grove of mysteriously bent trees video as the most curiosity-provoking (avg. score = 6.47, SD = 1.06). The mixed-effects logistic regressions also suggest that the bent trees video was significantly different from all other videos—except the nuclear disaster video—in its ability to prompt high reports of curiosity. For example, participants were 34.5 times more likely to score high on feelings of curiosity from this video compared to the sun’s magnetic fields video. The interviews indicate that this video fostered self-diminishment—a key component of awe—due to its physical as well as conceptual qualities (e.g., the vastness of time and nature). Furthermore, when focusing on these trees, the lab and interview responses embody how a sense of mystery can capture attention. The video’s lack of a distinct explanation for why the trees were bent fostered this mystery.

All participants expressed intentions to explore the awe elicitors in further detail. When asked what questions emerged from their viewing experience, eight of the seventeen reactions concerned the elicitors’ origins, making it the most prevalent response. For example, participants desired to know more about the different theories on the trees’ curvature.

Prior commitments and lack of immediate relevance, however, often stymied curiosity-sating behavior. Additionally, feelings of awe and curiosity from the VR experience reduced on average by 7% and 4%, respectively, over 24 hours. As such, only six participants confirmed that they searched (via Google) for information on the awe elicitors. The four interviewees similarly expressed that the bent tree and boxwood sculpture videos piqued their curiosity, but only two searched for further information. The theme of immediate relevance also appeared for these two interviewees; both searched for information on the day of the laboratory session.

When asked what would have pushed them to search for more information, the most prevalent laboratory response centered on greater mystery and context. Similarly, when asked what they would incorporate into VR designed to foster awe and curiosity, all interviewees stressed the importance of fantastical elements and drew from science fiction or fantasy titles.

DISCUSSION

Although awe-inspiring VR scenes may not always lead to information seeking, the study points to (1) how awe re-adjusts mental schemas and (2) design considerations for educational VR content.

The results suggest that Piaget’s (1970, 1973) notion of a need for accommodation does occur with virtual awe elicitors, especially through thoughts on the origins of particular stimuli. Unable to assimilate new information in mental schemata, the mind may reel back to the imagined creation of the elicitor. This cognitive experience echoes feelings of aura (Benjamin, 1935/2008), resonance and wonder (Greenblatt, 1991), or numinosity (Cameron & Gatewood, 2000, 2003; Latham, 2013). The viewer begins stimulus-driven processing, first diverting their attention to the object’s material qualities and then envisioning cultural or natural connections that produced it.

Participants’ desires for awe-inspiring VR content with greater mystery and fantasy may also reveal qualities of awe as a phenomenon. Questioning the use of VR to instill awe, Schneider (2017) suggests that real-life, slow-simmer forms of awe provoke mystery that is “beyond schematization” (p. 105). Participants’ wishes for VR content with more mystery and fantasy may indicate a desire for this slow-simmer awe, which is unexplainable and perhaps more akin to childlike wonder.

This study also uncovers the challenge of balancing emotionally-fueled awe experiences with didactic learning content. If VR designers wish to promote awe while providing learning content, they may benefit from incorporating suspenseful storytelling. Suspense, similar to epistemic curiosity, increases the desire to know (Ryan, 2008). Because suspense is a future-oriented state that motivates people to foresee possibilities, epistemic plots may foster goal-oriented actions like information seeking (Ryan, 2008). Furthermore, incorporating these goals into curricula will require immediate relevance with learning objectives.

CONCLUSION

This study considered VR-induced awe—a perception-expanding, epistemic state—as a tactic to increase motivation for information seeking. Other information researchers have tested novel tools for increasing motivation in student inquiry projects, such as video games (Urban, 2019) and creative nonfiction (Urban et al., 2020). This preliminary study’s findings suggest that awe-inspiring VR content may be another valuable technique, but future design research is necessary to uncover the gap between intent and behavior.

REFERENCES

- Anderson, C. L., Dixon, D. D., Monroy, M., & Keltner, D. (2020). Are awe-prone people more curious? The relationship between dispositional awe, curiosity, and academic outcomes. *Journal of Personality, 88*(4), 762–779. <https://doi.org/10.1111/jopy.12524>
- Benjamin, W. (2008). *The work of art in the age of mechanical reproduction* (J. A. Underwood, Trans.). Penguin. (Original work published 1935)
- Cameron, C. M., & Gatewood, J. B. (2000). Excursions into the un-remembered past: What people want from visits to historical sites. *The Public Historian, 22*(3), 107–127. <https://doi.org/10.2307/3379582>
- Cameron, C. M., & Gatewood, J. B. (2003). Seeking numinous experiences in the unremembered past. *Ethnology, 42*(1), 55. <https://doi.org/10.2307/3773809>
- Fiedler, K. (2001). Affective states trigger processes of assimilation and accommodation. In L. L. Martin & G. L. Clore (Eds.), *Theories of mood and cognition: A user's guidebook*. Lawrence Erlbaum Associates, Inc.
- Greenblatt, S. (1991). Resonance and wonder. In I. Karp & S D Lavine (Eds.), *Exhibiting cultures: The poetics and politics of museum display* (pp. 42–56). Smithsonian Institution Press.
- Keltner, D., & Haidt, J. (2003). Approaching awe, a moral, spiritual, and aesthetic emotion. *Cognition and Emotion, 17*(2), 297–314. <https://doi.org/10.1080/02699930302297>
- Latham, K. F. (2013). Numinous experiences with museum objects. *Visitor Studies, 16*(1), 3–20. <https://doi.org/10.1080/10645578.2013.767728>
- Litman, J. (2017, May 10). *Anyone have a state measure of curiosity?* [Comment]. ResearchGate. https://www.researchgate.net/post/Anyone_have_a_state_measure_of_curiosity
- Morton, A. (2009). *Epistemic emotions*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199235018.003.0018>
- Schindler, I., Hosoya, G., Menninghaus, W., Beermann, U., Wagner, V., Eid, M., & Scherer, K. R. (2017). Measuring aesthetic emotions: A review of the literature and a new assessment tool. *PLOS ONE, 12*(6), e0178899. <https://doi.org/10.1371/journal.pone.0178899>
- Piaget, J. (1970). Piaget's theory. In H. Mussen (Ed.), *Carmichael's manual of child psychology* (3rd ed., pp. 703–710). Wiley.
- Piaget, J. (1973). *The child and reality*. Grossman Publishers.
- Schneider, K. (2017). The resurgence of awe in psychology: Promise, hope, and perils. *The Humanistic Psychologist, 45*(2), 103–108. <https://doi.org/10.1037/hum0000060>
- Shiota, M. N., Keltner, D., & Mossman, A. (2007). The nature of awe: Elicitors, appraisals, and effects on self-concept. *Cognition and Emotion, 21*(5), 944–963. <https://doi.org/10.1080/02699930600923668>
- Smith, J. A., Flowers, P., & Larkin, M. (2009). *Interpretative phenomenological analysis: Theory, method and research*. SAGE.
- Ryan, M.-L. (2008). Interactive narrative, plot types, and interpersonal relations. In U. Spierling & N. Szilas (Eds.), *Interactive Storytelling* (pp. 6–13). Springer. https://doi.org/10.1007/978-3-540-89454-4_2
- Urban, A. C. (2019). Serious games for information literacy: A scoping review and design recommendations. *Library Hi Tech, 37*(4), 679–698. <https://doi.org/10.1108/LHT-01-2019-0010>
- Urban, A. C., Edgar, W. B., & Bossaller, J. S. (2020). Narrative immersion for information literacy: A pilot study. *Proceedings of the Association for Information Science and Technology, 57*(1). <https://doi.org/10.1002/pr2.371>

Are Virtual Reference Services Color and Gender Blind? Service Equality Revisited

Vladoiu, Megan Indiana University, United States | mvladoiu@indiana.edu
Fichman, Pnina Indiana University, United States | fichman@indiana.edu

ABSTRACT

Following the rise of the Black Lives Matter Movement in 2020, there has been an increased focus on racial equality in library services, including recent modifications of the ALA Code of Ethics. Online library services may eliminate race-based service discrimination by obscuring the racial identity of the patron. However, racial assumptions based on patrons' names can lead to online discrimination. While prior research shows mixed results, we examine if there is evidence of racial bias in virtual reference in American academic libraries. Initial findings based on content analysis of 946 responses from 255 libraries suggest that Tyrone Jackson received better service than others did.

KEYWORDS

Information needs, racial and gender bias, virtual reference services.

INTRODUCTION AND BACKGROUND

In July 2021 the American Library Association updated their code of ethics, adding a ninth principle in support of diversity and inclusion, which reads:

We affirm the inherent dignity and rights of every person. We work to recognize and dismantle systemic and individual biases; to confront inequity and oppression; to enhance diversity and inclusion; and to advance racial and social justice in our libraries, communities, profession, and associations through awareness, advocacy, education, collaboration, services, and allocation of resources and spaces.

Service discrimination on the basis of race or gender has been a major social concern in the face-to-face environment and reports on discrimination in public accommodation are not rare (Bertrand & Duflo, 2017). Research has shown that because computer-mediated communication decreases social cues and reduces social presence, it may decrease discrimination. Online discrimination may be less common also due to the ability of potential targets to eliminate social and group identification cues and to remain anonymous; the relative absence of social cues may mean greater equality of services in the virtual environment. Yet, toxic online disinhibition may increase behaviors that are less socially acceptable, such as harassment and trolling, and may give rise to online discrimination (Suler, 2004). Results from prior research on online discrimination in library services in the USA are mixed (Shachaf & Horowitz, 2006; Shachaf et al., 2008), and there is a lack of empirical research for over a decade. However, not only has recent evidence of bias in library reference practices in the UK been documented (Hamer, 2021), evidence of bias in American public services, including public libraries, toward African Americans was documented (Giulietti et al., 2019). It has also been documented that gender impacts online services, where female are more likely to be discriminated against (Schiffrin et al., 2021). There is a need to examine if online library reference services are biased or equal nowadays, especially given the increased awareness of these issues among librarians. We aim to address the question: Are library services color and gender blind?

METHOD

	Week 1	Week 2	Week 3	Week 4
First set of institutions	Question 1 from female Caucasian	Question 2 from female African American	Question 3 from male Caucasian	Question 4 from male African American
Second set of institutions	Question 2 from female Caucasian	Question 3 from female African American	Question 4 from male Caucasian	Question 1 from male African American
Third set of institutions	Question 3 from female Caucasian	Question 4 from female African American	Question 1 from male Caucasian	Question 2 from male African American
Fourth set of institutions	Question 4 from female Caucasian	Question 1 from female African American	Question 2 from male Caucasian	Question 3 from male African American

Table 1. Counter balanced method of email requests

We designed and received IRB approval to conduct an unobtrusive study to determine whether email reference services provide equivalent service to different ethnic and gender groups when salience of diversity is not an obvious factor, and when all other factors are constant. A 2 by 2 experimental design, using scenarios of information

needs, was conducted, manipulating user’s gender and ethnicity (Table 1). The experimental feature of the study is that the requests that were emailed to 255 academic libraries in the USA differed only in the implicit ethnicity of the user involved in the scenarios, as indicated by the user’s name. We have used four versions of each request; each version of the request represented one ethnic (African American/Caucasian) by gender (Female/Male) group. Each targeted library received a version of the same request, but with a different user name, indicating a different gender and ethnicity, using the counterbalanced method. This method is particularly important when attitude and behavior variability are measured on sensitive variables, such as ethnicity, because service providers may monitor their discriminatory behavior more closely if they believe the study is related to gender, race, or ethnicity (Bushman & Bonacci, 2004). Thus, the use of names perceived to be of a particular ethnicity is a common method to examine possible bias (Bertrand & Mullainathan, 2004) and specifically in virtual reference services by librarians (Hamer, 2021; Shachaf et al., 2008). The names used were Emily Baker (female Caucasian), Latoya Jones (female African American), Todd Kelly (male Caucasian), and Tyrone Jackson (male African American).

The questions sent were as follows:

1. Dissertation query: Can you tell me the title of [name]'s dissertation? [They] finished [their] degree at [institution name] in [year]. Do you have it in your library? How can I obtain a copy of this dissertation?
2. Sports team query: How did [sports team name] become the name for [institution name]'s sports teams? Can you refer me to a book or article that discusses it?
3. Population query: Could you tell me the population of [institution's city name] in 1963 and 1993?
4. Policy query: Can I use the computer at the library if I’m not a student?

Data collection occurred in February-March 2022, and 946 email replies were coded in Nvivo. Automated replies were not included in data analysis. Responses were coded using a modified version of Shachaf et al.’s (2008) codebook and included: Answered Yes/No; Asks Question; Closing; Complete Yes/No; Follow Up; Further Response; Greeting; Name; Policy Explanation; Referral; and Show of Interest.

FINDINGS

As can be seen in Table 2, we found that Latoya Jones had the highest response rate. She also received the highest amount of follow-up responses, as indicated by “# of Replies Total”. However, Tyrone Jackson also received a higher response rate than either Emily Baker or Todd Kelly. Tyrone Jackson had the highest percentage of Answered: Yes, as well as the highest percentage of Complete Answer. However, given that a high number of the responses that did not contain an answer (Answered: No) and many of the incomplete answers were in response to the population query, which was not sent by Tyrone, this is not surprising. We excluded the population query in week 4 in order to preserve the integrity of the study, as we learned that some libraries were communicating on a listserv in an attempt to answer this challenging question.

	Emily Baker	Latoya Jones	Todd Kelly	Tyrone Jackson
Greeting	203 (82.5%)	217 (78.9%)	196 (81.7%)	158 (85.4%)
Answered: Yes	202 (92.2%)	215 (91.9%)	195 (92.0%)	165 (97.6%)
Complete Answer	62 (42.8%)	70 (45.2%)	58 (40.3%)	57 (53.3%)
# of Libraries Responding	219 (85.9%)	234 (91.8%)	212 (83.1%)	169 (88.5%)
Total # of Replies	246	275	240	185

Table 2. Code frequency (and percent of answers) by name

While the initial findings are mixed, further analysis is needed to determine if these differences are significant, and to determine if an examination of the answers by question type may yield additional insights.

DISCUSSION AND CONCLUSION

Now more than ever before it is important to examine if online library services are equal, especially as more services are provided online, driven by the COVID-19 lockdown, and more awareness to equality is brought to the front, driven by the Black Lives Matter movement and the addition to the ALA Code of Ethics. Our mixed findings can be explained in light of the study limitations (the population question was more challenging than we had planned), the changing times (increased awareness on one hand and increased reliance on online interactions on the other), and theories of computer mediated communication (dual manifestations of online disinhibitions, benign and toxic). However, only further analysis can shed light on service equality on the basis of gender and race, in online reference services, that are provided by academic libraries in the USA.

REFERENCES

American Library Association (2021). ALA adopts new Code of Ethics principle on racial and social justice. Retrieved from: <https://www.ala.org/news/member-news/2021/07/ala-adopts-new-code-ethics-principle-racial-and-social-justice>

- Bertrand, M. & Duflo, E. (2017). Field experiments on discrimination. *Handbook of Economic Field Experiments*, 1, 309–393.
- Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *The American Economic Review*, 94(4), 991-1013.
- Bushman, B.J., & Bonacci, A.M. (2004). You've got mail: Using e-mail to examine the effect of prejudiced attitudes on discrimination against Arabs. *Journal of Experimental Social Psychology*, 40(6), 753– 759.
- Giulietti, C., Tonin, M., & Vlassopoulos, M. (2019). Racial Discrimination in Local Public Services: A Field Experiment in the United States. *Journal of the European Economic Association*, 17(1), 165–204. <https://doi.org/10.1093/jeea/jvx045>
- Hamer, J. (2021). Colour blind: Investigating the racial bias of virtual reference services in English academic libraries. *Journal of Academic Librarianship*, 47(5). <https://doi.org/10.1016/j.acalib.2021.102416>
- Shachaf, P., & Horowitz, S. (2006). Are virtual reference services color blind? *Library & Information Science Research*, 28(4), 501– 520.
- Shachaf, P., Oltmann, M. S., & Horowitz, S. (2008). E-quality in virtual reference services: Evaluation of service equality. *Journal of the American Society for Information Science and Technology*, 59(4), 535-550.
- Schiffirin, A., Koc-Michalska, K., & Ferrier, M. (2021). Women in the digital world. *Information, Communication & Society*, 24(14), 1991-1997.
- Suler, J. (2004). The online disinhibition effect. *Cyberpsychology & Behavior*, 7, 321–326.

Making a Cocoon: The Social Factors of Pandemic Misinformation Evaluation

Wan, Yi

University of South Carolina, USA | YW19@email.sc.edu

Thompson, Kim M.

University of South Carolina, USA | KimThompson@sc.edu

ABSTRACT

This study explores the social factors that may impact individuals' evaluation process of pandemic-related misinformation through a socio-cognitive lens. We conducted eight semi-structured interviews to collect data from individuals. Content analysis was guided by framework analysis of the interview transcripts. The social factors revealed in the study are social identity, social groups, social authorities, social spaces, social media, and social algorithms. These factors work together and isolate individuals from heterogeneous information. Social identity may decide other factors; correspondingly, the information filtered by social groups, authorities, spaces, media, and algorithms reinforces individuals' social identity. The tendency may reinforce bias on pandemic information and put people at risk. The research may provide an implication to information platforms to reconsider their algorithm designs and a direction for information literacy training programs to break the deficit assumption on individuals.

KEYWORDS

Misinformation; COVID-19; Information practices; Information world

INTRODUCTION

Pandemic misinformation is incorrect or inaccurate information related to the COVID-19 pandemic regardless of intent (Gabarron et al., 2021). The World Health Organization (WHO) indicated that the prevalence of COVID-19 is accompanied by misinformation (WHO, 2020). Widely spread misinformation may exacerbate the pandemic, put people at risk, and threaten democracy (Ali, 2020; Hansson et al., 2021; Islam et al., 2020).

Understanding what factors may impact individuals' perceptions of pandemic misinformation is essential to reduce its negative outcomes. Pandemic misinformation studies usually emphasize individuals' health or/and information literacy impacted by the principle of autonomy, a paradigm that encourages citizens to take individual responsibilities for their health (Morley et al., 2020). However, information phenomena are embedded in social, organizational, and professional contexts (Hartel, 2019). This study focuses on how the surrounding environment can create barriers for an individual to identify pandemic misinformation. The researchers explore the following questions: 1) What social factors may impact individuals' evaluation process of pandemic-related misinformation? 2) How may social factors impact individuals' evaluation process of pandemic-related misinformation?

METHODOLOGY

This exploratory study adopts a qualitative research design since individuals' perception of pandemic misinformation is heavily impacted by their various backgrounds. Eight participants were recruited by convenience and snowball sampling. All the participants are above 18 years old and currently living in Columbia, SC, USA. Four participants are native-born US citizens, and the other four were born in China.

The semi-structured interviews were conducted in person between March and May 2022. The interviews include two main parts: the context of pandemic misinformation exposure and participants' evaluation process of pandemic misinformation. Participants could decide if they wanted to use pseudonyms. The interviews were audio-recorded and transcribed verbatim with participants' permission. We developed the initial codebook by inductive content analysis using NVivo. At first, we separately read the transcripts, highlighted the sentences related to the research questions, and extracted keywords from participants' statements. Then, we compared the highlights and the keywords, classified the keywords, and created codes together. We defined each code by the transcripts and exiting literatures to resolve the coding discrepancies. The coding process was ongoing during data collection, which lasted about three months.

FINDINGS AND DISCUSSION

Pandemic Misinformation Evaluation

Participants used various strategies to evaluate the pandemic information. They tended to verify the suspicious information with different information resources. Embodied experiences play an important role in the evaluation process. For example, Yuting followed a diet suggestion that claimed it would strengthen people's immune system but eventually found it was misinformation. Participants considered costs and benefits when facing pandemic information: "There's a lot of tools that you can use to do this, but they take more time than it's worth" (Katie).

Making a Cocoon

Social Identity as a Starting Point

Participants hold specific social identities when interacting with pandemic information. They defined themselves and other people by nationality, region, political affiliation, religious affiliation, personal belief, and social status. These social identities profoundly impact their choices of social groups, social authorities, social spaces, and social media. When talking about news, Ollie said, “Like news sources that are generally considered more conservative, then I generally would be more skeptical. That probably, you know, is my own bias.” The participants born in China but living in Columbia acquired pandemic information from Chinese media; as Hannah said, “I just feel it is easy for me to understand and read.”

Spinning: The Flow of Pandemic Information

Pandemic information flows to an individual from different directions. *Social group members* are essential pandemic information resources: Stacy heard pandemic misinformation from her grandparents, Joey heard it from co-workers, etc. Pandemic information also directly comes from *social authorities*; a typical example is news about the pandemic or government statements. *Social spaces* facilitate access to pandemic information, increasing both opportunities and risks. Adrienne’s workspace provided them free access to the New York Times; meanwhile, Joey encountered misinformation in her lab. *Social media* are another primary pandemic information source; however, participants critiqued them as lacking evidence (Adrienne) and profit-oriented (Hannah).

These factors often create intertwined scenarios for pandemic information flow instead of working separately. Adrienne indicated that their *partner* listened to a lot of *podcasts* and read a lot of *news* during the pandemic, so they would ask for pandemic information from them. In this scenario, pandemic information flowed from social groups to social media and social authorities. Joey’s *colleagues* showed her misinforming *TikTok* videos when she went to her *lab* after getting vaccinated. In this example, pandemic misinformation was flowing to her via her social group using social media in a social space. A potential underlying issue identified in our analysis is that the information flowing within these social dynamics might tend to be homogenous rather than heterogeneous since individuals construct information worlds based on their social identity, which may create a consensus among people in the information flow circle (e.g., Chatman, 1991; Granovetter, 1973).

Social Algorithms, The Last Layer

What was clear in our interviews is that participants’ interactions with pandemic information was deeply intertwined with the Internet, leaving the space for algorithms to control information flow. “I’m not sure you’re familiar with TikTok, if you watch one video on that topic, then it just gives you more about the same topic, so you can keep watching it. It’s like constantly they pushing you with all the information, like, the bad part of the technology” (Joey). Social algorithms tailor pandemic information to participants based on their social identity, which might be reflected by their interaction with social groups, authorities, spaces, and social media. When participants clicked a New York Times link, liked the speech of Dr. Fauci, or shared statements from the Centers for Disease Control and Prevention or the World Health Organization, the social algorithms automatically decreased the priority of or even filtered information from news sources that opposed these perspectives. Social algorithms may relieve people from information overload but also create filter bubbles to isolate us from different opinions. They work with social groups, social authorities, social spaces, and social media, tightening the cocoon for an individual to limit the heterogeneous information flow in one’s information world.

CONCLUSION

Individuals may try to verify pandemic information using various literacies and strategies when they think it is necessary. However, the resources they relied on may be determined by their social identities. Information literacy training usually emphasizes checking personal biases. Biases, however, can be invisible when we live in a world fulling of heterogeneous information filtered algorithms. Therefore, information literacy training programs will be more effective if they include training about social algorithms and how social decisions affect what information one is fed online. In a perfect world, information delivered through social media platforms and search engines might be reconsidered and used to empower the user rather than making decisions for them based on fragments of individuals’ social identities. As COVID continues, misinformation about the pandemic will likely continue to be prevalent. This pilot study may raise more attention to social factors in pandemic misinformation research. The research might lack universality due to the sample size and sampling techniques. Future studies can recruit participants from a wider variety of backgrounds to explore more social factors impacting misinformation evaluation. Quantitative studies also can be conducted to verify the influence of the discussed factors on individual misinformation evaluation practices.

ACKNOWLEDGEMENTS

We gratefully appreciate the participants of this study and Dr. Mónica Colón-Aguirre and Dr. Vanessa Kitzie for suggestions on previous versions of the document.

REFERENCES

- Ali, S. (2020). Combatting against COVID-19 & misinformation: A systematic review. *Human Arenas*, 1-16. <https://doi.org/10.1007/s42087-020-00139-1>
- Chatman, E. A. (1991). Life in a small world. *Journal of the American Society for Information Science*, 42(6), 438-449. [https://doi.org/10.1002/\(SICI\)1097-4571\(199107\)42:6<438::AID-ASI6>3.0.CO;2-B](https://doi.org/10.1002/(SICI)1097-4571(199107)42:6<438::AID-ASI6>3.0.CO;2-B)
- Gabarron, E., Oyeyemi, S. O., & Wynn, R. (2021). COVID-19-related misinformation on social media: a systematic review. *Bulletin of the World Health Organization*, 99(6), 455-463A. <http://dx.doi.org/10.2471/BLT.20.276782>
- Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6), 1360-1380. <https://www.jstor.org/stable/2776392>
- Hansson, S., Orru, K., Torpan, S., Bäck, A., Kazemekaityte, A., Meyer, S. F., ... & Pigrée, A. (2021). COVID-19 information disorder: six types of harmful information during the pandemic in Europe. *Journal of Risk Research*, 24(3-4), 380-393. <https://doi.org/10.1080/13669877.2020.1871058>
- Hartel, J. (2019). Turn, Turn, Turn. *Proceedings of CoLIS*, 24, paper colis 1901. <http://www.informationr.net/ir/24-4/colis/colis1901.html>
- Islam, M. S., Sarkar, T., Khan, S. H., Kamal, A. H. M., Hasan, S. M., Kabir, A., ... & Seale, H. (2020). COVID-19-related infodemic and its impact on public health: A global social media analysis. *The American Journal of Tropical Medicine and Hygiene*, 103(4), 1621-1629.
- Morley, J., Cows, J., Taddeo, M., & Floridi, L. (2020). Public health in the information age: Recognizing the infosphere as a social determinate of health. *Journal of Medical Internet Research*, 22(8), e19311.
- WHO. (2020). *Countering misinformation about COVID-19*. <https://www.who.int/news-room/feature-stories/detail/countering-misinformation-about-covid-19>

Investigating the Relationship between In-Situ User Expectations and Web Search Behavior

Wang, Ben School of Library and Information Studies, University of Oklahoma, USA | benw@ou.edu

Liu, Jiqun School of Library and Information Studies, University of Oklahoma, USA | jiqunliu@ou.edu

ABSTRACT

Pre-adoption *expectations* often serve as an implicit reference point in users' evaluation of information systems and are closely associated with their goals of interactions, behaviors, and overall satisfaction. Previous studies have involved simulated user expectation as a feature in user modeling to model biased search actions. However, there is still little direct evidence revealing the relationships between users' expectations and their actual search behaviors. To address the gap, we collected 448 query sessions from participants in a controlled-lab user expectation study and gathered direct query-level feedback on their *expected information gains* (e.g., number of useful pages) and *expected search efforts* (e.g., clicks and dwell time) under each query. To our knowledge, this is the first attempt to explicitly examine the connections between different aspects of *in-situ* search expectations and user behaviors. Findings on user expectation advance our understanding of users' search decision-making and evaluation strategies and will also facilitate the design and evaluation of expectation-aware user models, metrics, and IR systems.

KEYWORDS

User search expectation; Interactive information retrieval; Web search; User cognitive bias

INTRODUCTION

With growing research in cognitive psychology and behavioral economics, investigating the cognitive bias is attracting tremendous attention to users' search behaviors in information retrieval (Azzopardi, 2021; Gomroki, Behzadi, Fattahi, & Salehi Fadardi, 2021; Lau & Coiera, 2007; White, 2013). Users estimate possible information gain/cost based on existing gain/cost and make the decision to maintain an optimal rate of gain and cost instead of pursuing the highest gain (Kahneman, 2003; Pirolli & Card, 1999; Tversky & Kahneman, 2019). In this process, users' expectation can represent their estimation and perception of gain and cost, help them form reference points, and affect their decision in information seeking (Abeler, Falk, Goette, & Huffman, 2011; Backus, Blake, Masterov, & Tadelis, 2022, 2017; Brown & Liu, 2022; Cox & Fisher, 2009; Kahneman, 2003; Liu & Han, 2020). Thus, their expectations create cognitive bias and lead to biased results in user modeling and evaluation. Previous studies have been engaged in modeling users' behaviors under influence of their expectations (N. Chen, Zhang, & Sakai, 2022; Moffat, Bailey, Scholer, & Thomas, 2017). However, there is little direct evidence showing relationships between users' expectations and actual search behaviors. Therefore, we conducted a user study to collect direct evidence about users' in-situ expectations regarding information gains and search efforts and their associated search behaviors at query level. This poster presents late-breaking results of our study, aiming to answer following research question:

RQ: What are the relationships between users' in-situ query level expectations and their actual search behaviors?

METHOD

To answer the RQ, we conducted a user study to collect data on both search interactions and users' explicit feedback on their expectations in each query segment. This user study assigned predefined complex search tasks of four types to 60 participants (undergraduate students from a U.S. research university). We adopted the four journalism tasks applied in previous studies and recruited participants not from the journalism major to control prior experience in these task types. The four task types include copy editing, story pitching, relationship, and interview preparation (Liu, Mitsui, Belkin, & Shah, 2019). These tasks have been scientifically demonstrated to be beneficial for motivating multi-round search interactions and regulating the possible impacts of a variety of contextual factors (Cole, Hendaheva, Belkin, & Shah, 2015; Li & Belkin, 2008). In addition, we chose two uncommon topics (1. coelacanth, 2. methane clathrates and global warming) to further control the possible variation in topic familiarity, as neither topic is likely to be familiar to our participant pool. When the participants submit a query, they are asked to complete a quick survey about their expectations of information gains and search efforts before browsing the retrieved documents. Survey questions are listed in Table 1. Here we define the expectation as users' perceived gain (e.g., useful pages) and cost (e.g., spending time) in each query segment. Besides their expectation labels, we also collected search behavioral data under each query for which user expectation feedback was collected. The behaviors include clicking, usefulness annotation, and spending/dwell time. The behavioral data is collected through a browser extension based on an open-source user study toolkit (J. Chen et al., 2021) with essential modifications for this study. After the data collection, we examined the normality of data distribution and chose statistical tests accordingly to find the associations between their expectations and actual behaviors. We further visualized the

distributions of their expectations and behaviors and implemented a locally weighted linear regression (Loess) (Atkeson, Moore, & Schaal, 1997) to investigate the impacts of expectations on their search behaviors.

Expectation type	Question
Useful pages	How many useful pages do you expect to find? (Numeric)
Clicking results	How many results do you expect to click before obtaining the expected number of useful pages? (Numeric)
Spending/dwell time on content pages	How much time do you expect to spend on this search? (Ordinal) <ul style="list-style-type: none"> • fewer than 30 s (I can find the useful result instantly) • 30 s to 1.5 min (I can find the useful result quickly after inspecting it) • 1.5 min to 3 min (I need some time to read the results, but it won't take so long.) • 3 min to 5 min (I need some time to read the results.) • more than 5 min (I need more time to read the results carefully.)

Table 1. Pre-query questionnaire about expectations

RESULTS

As the result of the user study, we recruited 60 undergraduates as participants and collected 448 queries with expectation feedback on search gains and efforts and search behavioral data. The distributions of users' expectations are shown in Figure 1 as grey bars. Overall, multiple aspects of user expectations are positively correlated with users' behaviors of finding useful pages, clicks, and dwell time, and Spearman's rank coefficients are 0.203, 0.280, and 0.333 with $p < 0.01$, respectively. The positive correlations demonstrate that users with high expectations also have similar levels and trends in search behaviors to match their pre-search expectations. The blue line in Figure 1 represents the average values of search behaviors with a 95% confidence interval error bar. In general, the value of users' search behaviors increases with higher expectation values. However, when the expectations reach a point, there is a high variation in users' actual behaviors, and their behaviors may not increase accordingly. This inconsistency is also indicated by the red line, which is the result of Loess. Loess fits the behavior trends in different subranges of expectations. The trend is more linear when the number of expected useful pages is lower than three, the number of expected clicks is lower than six, or the expected spending time is high than 1.5 minutes. For the useful pages and clicks, when the expectations are higher than these values (for the spending time, when the expectation is lower than 1.5 minutes), the trends become flat because of the less correlation between their high expectations and high variance of behaviors. Although users have high expectations (expectations of high numbers of useful pages, clicks, and low spending time), they may not have compatible behaviors to match their expectations. This inconsistency reflects users' bounded rationality (e.g., optimism bias) in estimating their gain and cost in Web search. The heterogeneity in expectation-behavior correlations across varying ranges would not have been revealed without collecting explicit feedback and labels on in-situ search expectations in our user study.

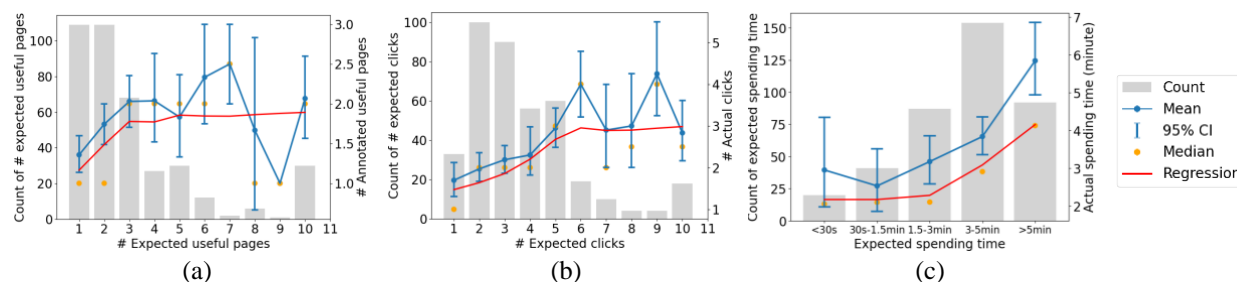


Figure 1. Relationships between users' expectations and actual search behaviors

CONCLUSION

To investigate reference dependence and pave the way toward comprehensively studying cognitive biases in IR, we conducted a user study to collect data about user search expectations and actual behaviors in Web searching, including finding useful information, clicks, and dwell time on pages. The results indicate that users' pre-search expectations generally tend to be conservative (e.g., less than three expected useful pages and higher expected spending time) and that their expectations are positively correlated with their actual behaviors in general. Furthermore, the trends of their behaviors are more linear until users have high expectations (high number of useful pages, clicks, or low dwell time). However, when users have higher expectations, their behaviors may deviate from their pre-search expectations, leading to expectation disconfirmation scenarios in search and evaluation. These results can help us better understand users' search behaviors with the knowledge of their expectations and explore the impacts of implicit reference points and other cognitive biases. Future work can further examine the relationships between users' expectations and other factors, such as the search intention, emotional state, and query formulation strategies, and incorporate the knowledge about search expectations into user-centered IR evaluation.

ACKNOWLEDGMENT

This work is supported by the National Science Foundation (NSF) grant IIS-2106152. We also want to thank our undergraduate assistant, Zeming Hu, who helped us run some of the user study sessions and analyze data.

REFERENCES

- Abeler, J., Falk, A., Goette, L., & Huffman, D. (2011). Reference points and effort provision. *American Economic Review*, 101(2), 470–492. <https://doi.org/10.1257/aer.101.2.470>
- Atkeson, C. G., Moore, A. W., & Schaal, S. (1997). Locally weighted learning. *Lazy Learning*, 11–73.
- Azzopardi, L. (2021). Cognitive Biases in Search: A Review and Reflection of Cognitive Biases in Information Retrieval. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval* (pp. 27–37). <https://doi.org/10.1145/3406522.3446023>
- Backus, M., Blake, T., Masterov, D., & Tadelis, S. (2022). Expectation, Disappointment, and Exit: Evidence on Reference Point Formation from an Online Marketplace. *Journal of the European Economic Association*, 20(1), 116–149. <https://doi.org/10.1093/jeea/jvab033>
- Backus, M., Blake, T., Masterov, D. V., & Tadelis, S. (2017). *Expectation, Disappointment, and Exit: Reference Point Formation in a Marketplace*.
- Brown, T., & Liu, J. (2022). A Reference Dependence Approach to Enhancing Early Prediction of Session Behavior and Satisfaction; A Reference Dependence Approach to Enhancing Early Prediction of Session Behavior and Satisfaction. In *Proceedings of the 22nd ACM/IEEE Joint Conference on Digital Libraries* (pp. 1–5). <https://doi.org/10.1145/3529372>
- Chen, J., Mao, J., Liu, Y., Zhang, F., Zhang, M., & Ma, S. (2021). Towards a better understanding of query reformulation behavior in web search. In *Proceedings of the World Wide Web Conference, WWW 2021* (pp. 743–755). <https://doi.org/10.1145/3442381.3450127>
- Chen, N., Zhang, F., & Sakai, T. (2022). Constructing Better Evaluation Metrics by Incorporating the Anchoring Effect into the User Model. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 2709–2714). <https://doi.org/10.1145/3477495.3531953>
- Cole, M. J., Hendaheba, C., Belkin, N. J., & Shah, C. (2015). User activity patterns during information search. *ACM Transactions on Information Systems (TOIS)*, 33(1), 1–39.
- Cox, A., & Fisher, M. (2009). An expectation-based model of web search behaviour. In *Proceedings of the 2nd International Conferences on Advances in Computer-Human Interactions, ACHI 2009* (pp. 49–56). <https://doi.org/10.1109/ACHI.2009.47>
- Gomroki, G., Behzadi, H., Fattahi, R., & Salehi Fardadi, J. (2021). Identifying effective cognitive biases in information retrieval. *Journal of Information Science*, 01655515211001777. <https://doi.org/10.1177/01655515211001777>
- Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, 93(5), 1449–1475. <https://doi.org/10.1257/000282803322655392>
- Lau, A. Y. S., & Coiera, E. W. (2007). Do People Experience Cognitive Biases while Searching for Information? *Journal of the American Medical Informatics Association*, 14(5), 599–608. <https://doi.org/10.1197/jamia.M2411>
- Li, Y., & Belkin, N. J. (2008). A faceted approach to conceptualizing tasks in information seeking. *Information Processing & Management*, 44(6), 1822–1837.
- Liu, J., & Han, F. (2020). Investigating Reference Dependence Effects on User Search Interaction and Satisfaction: A Behavioral Economics Perspective. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 1141–1150). <https://doi.org/10.1145/3397271.3401085>
- Liu, J., Mitsui, M., Belkin, N. J., & Shah, C. (2019). Task, information seeking intentions, and user behavior: Toward a multi-level understanding of web search. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval* (pp. 123–132). <https://doi.org/10.1145/3295750.3298922>
- Moffat, A., Bailey, P., Scholer, F., & Thomas, P. (2017). Incorporating user expectations and behavior into the measurement of search effectiveness. *ACM Transactions on Information Systems*, 35(3), 1–38. <https://doi.org/10.1145/3052768>
- Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review*, 106(4), 643.
- Tversky, A., & Kahneman, D. (2019). Advances in prospect theory: Cumulative representation of uncertainty. *Choices, Values, and Frames*, 5(4), 44–66. <https://doi.org/10.1017/CBO9780511803475.004>
- White, R. W. (2013). Beliefs and biases in web search. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 3–12). <https://doi.org/10.1145/2484028.2484053>

Using Textual Analysis Tools to Identify Trends in Privacy Concern Research

Weinberger, Maor
Bouhnik, Dan

Bar-Ilan University, Israel | maor89@gmail.com
Jerusalem College of Technology, Israel | bouhnik@g.jct.ac.il

ABSTRACT

Text mining is a research method that uses interdisciplinary techniques to identify patterns and trends in unstructured data and can be used to detect research trends and developments in various fields. Thus, this study attempts to identify research trends in the issue of privacy concern (PC), focusing on the differences between the first two decades of the 21st century. This is done by applying an automated textual analysis tool (Voyant Tools) on a vast corpora of academic papers conducted on this field. We found that while the first examined decade of PC research focused on commerce and marketing issues, there was much more research emphasis on the social dimension of privacy in the second decade. This difference was explained by the technological developments that were prominent in the respective decades. The research provided support for the use of textual analysis of academic literature as a tool to identify research trends in various fields, including technologically-oriented fields.

KEYWORDS

Privacy concern, textual analysis, Voyant Tools, comparative study, research trends.

INTRODUCTION

Text mining is a research method based on a variety of techniques, originating in the fields of statistics, machine learning and linguistics. Text mining makes use of interdisciplinary techniques to identify patterns and trends in unstructured data, mostly textual. The goal of text mining is "to be able to process large textual data to extract "high quality" information, which will be helpful for providing insights into specific scenario to which the text mining is being applied" (Ryan et al., 2014, p. 4). This method also aids the analytical examination of academic papers (Miller, 2018). Tu and Hsu (2016) surveyed the different uses of text mining in research and applied some of its features in their own research, such as using collections of words to gather data, calculating similarities between articles and identifying keywords. They concluded that text mining can be used to detect research trends and developments in many fields. For example: Hung (2012) used text mining to examine research trends in academic papers written about distant learning; another research used this method to detect trends in papers about cancer (Spasić et al., 2014); drug use (Chou et al., 2020); and even about waste recycling (Garechana, 2015).

The notion of privacy is a main research subject in various fields, such as philosophy, law, sociology and psychology. Studies that investigated the public attitudes towards privacy, found that users are concerned about preserving their personal privacy (Paine et al., 2007; Wills & Zeljkovic, 2011). Thus, privacy concern (PC) is a major research sub-theme within the general research about user information (Dinev & Hart, 2005; Paine et al., 2007; Wills & Zeljkovic, 2011). Fornaciari (2014) examined the evolution of the term "privacy," while using framing and textual analysis techniques applied to American news sources. Yun et al. (2019) studied the research development on personal information privacy, by conducting interviews and content analysis. In addition, Del Alamo et al. (2022) used natural language processing (NLP) methods to analyze privacy policies. Yet, not many papers have used text mining to investigate the issue of PC using automatic textual analysis tools. This research will strive to identify research trends in the issue of PC, while focusing on the differences between two decades: 2000-2009 and 2010-2019. This will be done by using an automatic textual analysis tool (Voyant Tools). The paper addresses whether there are differences in the themes discussed in the academic papers written about PC, between the first and second decades of the 21st century, and if so, how are these differences manifested?

METHODS

The research applies a textual analysis research method, based on an automatic tool: Voyant Tools (<http://voyanttools.org/>). This is a free, web-based tool, used for text analysis and. Researchers use this tool to analyze various digital texts, which can be uploaded from external files or web pages, or copied directly into the tool (Welsch, 2014). For the purposes of analysis, two corpora were created consisting of the 50 most cited papers in Web of Science database (as for June 2020), with at least 50 citations, relating to "privacy concern," published between the years 2000 and 2019:

- A. 25 most cited PC academic papers, published between 2000 and 2009.
- B. 25 most cited PC academic papers, published between 2010 and 2019.

The two comparative decades will be referred to throughout the paper as: P1 (2000-2009) and P2 (2010-2019). To refine the list to papers dealing directly with the themes discussed in this study, the search was limited to the title

85th Annual Meeting of the Association for Information Science & Technology | Oct. 29 – Nov. 1, 2022 | Pittsburgh, PA. Author(s) retain copyright, but ASIS&T receives an exclusive publication license.

field only, using the keywords: "privacy concern*." The two corpora were uploaded to the automatic analysis tool separately and the output was analyzed in the findings section below. Then, a comparison was conducted between them. Notably, after the file upload, terms that did not reflect the content of the papers were filtered out, e.g. et al., http, serial numbers and stop words (the, a, to, etc.). Also, the terms "privacy" and "concern" (including variations, such as "concerns") were filtered out from the PC word frequency list. In addition, as there was no limit on the paper length, there might be bias in the word frequency, due to the possible preference of long papers (containing more words than others).

RESULTS

We examined the differences between the PC research papers that were published in the first (P1) and second (P2) decades of the 21st century. Interestingly, the term "social" claimed the second place in the P2 word frequency list, while positioned only at the 14th place in the P1 list. Another prominent difference is the placement of the terms "consumers", "marketing," and "consumer," positioned at the eighth, ninth and tenth places of the P1 word frequency list, while only found around the 60th and 90th places of the P2 list. Table 1 presents the most frequent terms in the PC academic papers that were published between the years 2000-2009 and 2010-2019.

Rank	2000-2009		2010-2019	
	Term	No. of occurrences	Term	No. of occurrences
1	information	2,540	information	2,949
2	online	1,026	social	1,425
3	internet	991	research	1,032
4	use	809	online	1,007
5	personal	792	personal	968
6	users	650	data	951
7	research	643	use	939
8	consumers	603	trust	835
9	marketing	583	self	807
10	consumer	579	model	782
Total		9,216		11,695

Table 1. Most frequent terms in the PC academic papers that were published between the years 2000-2009 and 2010-2019

As for relations between terms, the automatic tool did not detect any significant correlations between the terms found in the PC P1 word frequency list. However, significant correlations were found in the P2 list between the term: "research" and several other terms, such as "personal" ($r=0.30$, $p<0.001$), "data" ($r=0.26$, $p<0.001$), and "behavior" ($r=0.21$, $p<0.01$). Notably, many of those terms were among the most frequent terms in this decade's word frequency list.

CONCLUSIONS

This paper attempted to identify research trends in the issue of privacy concern. This was done by using an automated textual analysis tool (Voyant Tools) that was used to compare two decades of research: 2000-2009 and 2010-2019. The interest in personal privacy is growing over the years and it seems that the academic literature is aligned with this trend. As found by Yun et al. (2019), we noticed that in the first decade of the 21st century there was a much more emphasis on PC in commerce (probably electronic), while in the second decade there is more focus on the social dimension. This can be explained by the vast emergence of social networks (e.g. Facebook) in the first decade of the 21st century and the research that was conducted about them in the following decade. Accordingly, the research on commerce and marketing in the first decade may be explained by the dramatic increase of the use in electronic commerce in this decade and several years before. Thus, it is safe to say that textual analysis of academic literature may be used as an efficient tool to identify research trends in various fields. The automatic textual analysis tool has proven its value for the digital humanities research, however it should be used sensibly and sparingly as a complementary tool in a broad research. In addition, it seems that additional tools should be used for textual and content analysis in order to obtain clearer and more comprehensive examination of the subject matter. Finally, future research may apply a smaller breakdown of research periods (e.g. every five years) for gaining more accurate and punctual insights regarding research trends.

REFERENCES

- Chou, L. W., Chang, K. M., & Puspitasari, I. (2020). Drug abuse research trend investigation with text mining. *Computational and Mathematical Methods in Medicine*, Article ID 1030815, 1-8. <https://doi.org/10.1155/2020/1030815>
- Del Alamo, J. M., Guaman, D. S., García, B., & Diez, A. (2022). A systematic mapping study on automated analysis of privacy policies. *Computing*. <https://doi.org/10.1007/s00607-022-01076-3>

- Dinev, T., & Hart, P. (2005). Internet privacy concerns and social awareness as determinants of intention to transact. *International Journal of Electronic Commerce*, 10(2), 7-29 (2005). <https://doi.org/10.2753/JEC1086-4415100201>
- Fornaciari, F. (2014). Mapping the territories of privacy: Textual analysis of privacy frames in American mainstream news. In the proceedings of: *the 47th Hawaii International Conference on System Science (HICSS '14)*, Waikoloa, HI. <https://doi.org/10.1109/HICSS.2014.230>
- Garechana, G., Rio-Belver, R., Cilleruelo, E., & Sarasola, J.L. (2015). Clusterization and mapping of waste recycling science. Evolution of research from 2002 to 2012. *Journal of the Association for Information Science and Technology*, 66(7), 1431-1446. <https://doi.org/10.1002/asi.23264>
- Hung, J. L. (2012). Trends of e-learning research from 2000 to 2008: Use of text mining and bibliometrics. *British Journal of Educational Technology*, 43(2), 5-16. <https://doi.org/10.1111/j.1467-8535.2010.01144.x>
- Miller, A. (2018). Text mining digital humanities projects: Assigning content analysis capabilities of Voyant Tools. *Journal of Web Librarianship*, 12(3), 169-197. <https://doi.org/10.1080/19322909.2018.1479673>
- Paine, C. Reips, U.-D., Steiger, S., Joinson, A., & Buchanan, T. (2007). Internet users' perceptions of 'privacy concerns' and 'privacy actions'. *International Journal of Human-Computer Studies*, 65(6), 526-536. <https://doi.org/10.1016/j.ijhcs.2006.12.001>
- Ryan, M., Talabis, M., Miyamoto, I., & Martin, J. L. (2014). *Information security analytics: Finding security insights, patterns and anomalies in big data*. Elsevier Publishing, Amsterdam, Netherlands.
- Spasić, I., Livsey, J., Keane, J. A., & Nenadić, G. (2014). Text mining of cancer-related information: Review of current status and future directions. *International Journal of Medical Informatics*, 83(9), 605-623. <https://doi.org/10.1016/j.ijmedinf.2014.06.009>
- Tu, Y-N. & Hsu, S-L. (2016). Constructing conceptual trajectory maps to trace the development of research fields. *Journal of the Association for Information Science and Technology*, 67(8), 2016-2031. <https://doi.org/10.1002/asi.23522>
- Welsh, M. E. (2014). Review of Voyant Tools. *Collaborative Librarianship*, 6(2), 96-97. <https://digitalcommons.du.edu/collaborativelibrarianship/vol6/iss2/8>
- Yun, H., Lee, G., & Kim, D. J. (2019). A chronological review of empirical research on personal information privacy concerns: An analysis of contexts and research constructs. *Information & Management*, 56(4) 570 -601. <https://doi.org/10.1016/j.im.2018.10.001>

Academic Casualization, Precarity, and Information Practices: A Scoping Review

Willson, Rebekah

McGill University, Canada | rebekah.willson@mcgill.ca

Stewart-Robertson, Owen

McGill University, Canada | owen.stewart-robertson@mcgill.ca

Julien, Heidi

University at Buffalo, USA | heidijul@buffalo.edu

Given, Lisa

RMIT University, Australia | lisa.given2@rmit.edu.au

ABSTRACT

In the first phase of a larger research project exploring the information practices of contract academic staff, a review of the wider literature around academic casualization was conducted. This research begins to address the need for greater understanding of the experiences of these workers in the information-intensive environments of higher education. A scoping review methodology was applied, and both academic and grey literature from multiple disciplines was reviewed. The literature points to casualization in academia as a growing phenomenon that has important consequences for staff, students, and universities themselves. This poster discusses the initial themes emerging from the literature, including: the precarity and marginalization of academic staff, the unrecognized emotional labor shaping information practices, and the impacts of multiple temporalities on the work, careers, and lives of such staff. The findings suggest the need for empirical research to address the lack of knowledge around the information environments of contract academics and set the stage for the next phases of the research project.

KEYWORDS

casualization, precarity, information practices, information marginalization, academic staff

INTRODUCTION

Contract academic staff (CAS) make vital contributions to teaching and research in higher education, yet their working conditions are challenging and increasingly precarious (Kezar et al., 2019; Zheng, 2018). While academic work is complex and information-intensive, requiring specialized knowledge, skills, and supports (Willson, 2018), academics on short-term contracts face added difficulties, often feeling insecure, undervalued, undercompensated, and prevented from working at full capacity (e.g., Birdsell Bauer, 2018; Brady & Briody, 2016; Foster & Birdsell Bauer, 2018; Jolley et al., 2013). Frequently marginalized within universities (e.g., Lopes & Dewan, 2014; Willson, 2016), CAS often lack the workplace information required to carry out day-to-day activities and advance their careers (Dolan, 2011; Kezar, 2013). Increasingly relied upon by universities to teach (al-Gharbi, 2020), the COVID-19 pandemic has intensified disparities and uncertainty for CAS (e.g., CAUT, 2020; Tenure for the Common Good, 2020). Because of the importance of this issue, for both staff and their institutions, the literature in this area is broad, including academic dialogues, personal narratives, and disciplinary discussions in the academic and grey literature. However, because of the wide-ranging discussions and the relative lack of empirical research into the experiences of CAS themselves, much remains unclear about the information experiences of CAS.

METHODOLOGY AND METHODS

To build understanding around the experiences and information practices (IP) of contract academic staff (CAS) (also known also as adjunct, contingent, part-time faculty, casual academics, sessionals, etc.), the first phase of a larger research project involved a scoping literature review (Arksey & O'Malley, 2005). A search strategy was developed by the research team and carried out in a wide selection of scholarly databases (Scopus, Academic Search Complete, ProQuest Research Library, Web of Science Core Collection, ProQuest Dissertations & Theses, LISA, LISTA, Worldwide Political Science Abstracts, ERIC, JSTOR, International Bibliography of the Social Sciences, and EconLit), Google News, and through searching of various online grey literature sources. Backward and forward citation tracking was also used, and searching was completed once the team agreed that a saturation point had been reached. Documents from English-language academic (2000-present) and grey (the past 10 years) literatures, from the United States, Canada, the United Kingdom, New Zealand, Ireland, and Australia, which discussed topics and trends related to the experiences of CAS and casualization and precarity in academia were selected for review. A total of 223 documents were included in the scoping review, which included journal, magazine and newspaper articles, conference proceedings, books, reports, and doctoral dissertations. The following research questions guided this review: 1) What are the experiences and information environments of CAS as described in both scholarly and grey literatures? 2) How do CAS and academic institutions interact and share information?

PRELIMINARY FINDINGS

This poster presents initial findings from the review of selected documents, to better understand the discourses around and information experiences of contract academic staff.

Precarity and Uncertainty Lead to Information Marginalization

The literature around CAS reinforces views of precarity as something that is differentially experienced (e.g., Rose, 2020), dehumanizing (Mason & Megoran, 2021), and self-perpetuating (Schofield, 2022). The uncertainty CAS experience centers around not only work contracts, but also roles, status within departments, how to find information, and career development, making professional (and personal) planning incredibly difficult (Loveday, 2018; Willson & Julien, 2020). The complex and intersecting realities of many CAS suggest the need to engage approaches such as information marginalization (Gibson & Martin, 2019) and information precarity (Stewart-Robertson, 2022) which highlight the systemic and institutional processes (e.g., Willson, 2018, 2019) inhibiting access to information and leading to marginalization. Furthermore, the differentially experienced conditions described in the literature suggest the need to conceptualize precarity in IP as not only tied to particular work circumstances but as interconnected with gender, race, ability, and wider economic and political forces.

Emotional Labor Complicates Information Practices

The literature points to both the emotional labor involved in the work of CAS and the emotional impacts of such precarious work as often suppressed or neglected. CAS have been seen as providing considerable undocumented emotional labor through student contact time and support (Lopes & Dewan, 2014) and in masking their contingent status from students (Read & Leathwood, 2020). Elsewhere, universities have been seen as repressing, ignoring, or commodifying the emotional labor required of CAS (Gannon et al., 2015), and the inability of institutions to meet the emotional and information needs of these staff has been suggested as linked to lack of belonging, isolation, and marginalization (e.g., Leigh, 2014; Ryan, 2017). CAS may thus be unable to receive or effectively provide the affective information often involved in carrying out their roles while also expending considerable emotional energy to manage the stress, anxiety, and depression often resulting from their work (e.g., Reevy & Deason, 2014). Despite noted links between specific IP (e.g., information avoidance or non-use) and affective experiences (e.g., Willson & Given, 2020), further research is needed to better understand how increased and underrecognized emotional labor may shape the ability of CAS to deal with their complex information environments.

Time as an Important Contextual Factor

In the literature reviewed, several documents suggest CAS face a variety of temporal realities that are often incongruous with the traditional duties of academics. A lack of time, both for teaching preparation and research activities (e.g., Ivancheva, 2015), an inability to plan for or even anticipate the future, on personal, professional, and financial levels (e.g., Lopes & Dewar, 2014; Mason & Megoran, 2021), and fears around running out of time and falling into a perpetual cycle of part-time roles (Roy et al., 2021), have been suggested. Furthermore, pressures of time have been seen as particularly impactful for CAS with disabilities (e.g., Butler-Rees & Robinson, 2020). These findings suggest the need for research addressing the impacts of different temporal realities, and the employed tactics for negotiating those temporalities, on the IP of CAS. Furthermore, while time has been acknowledged as important to the context of IP (Savolainen, 2006), and the need to recognize multiple understandings of time in IP (e.g., at individual, social, and project levels) has been noted (McKenzie & Davies, 2015, 2021), little research in the area has addressed temporality.

Large Gaps Remain in Our Understanding

Despite their complex information needs, the wider literature lacks discussions of the IP-related activities of CAS, such as information seeking, sharing, creation, or use. Preferences for (e.g., Loh, 2004) and circumstances which require (e.g., Langan & Morton, 2009; Mapes, 2019) seeking of information through informal, social sources have been noted. Yet, for some CAS, even these sources were seen as unavailable (e.g., Mason & Megoran, 2021) or challenging to identify (Willson & Julien, 2020), creating more isolation from important workplace information. Additionally, direct discussion of official policies and documents involved in interactions between CAS and higher education institutions is rare, and detailed information from institutions about their contract appointments is often lacking (Brownlee, 2015). While this may result from a need for greater transparency (Gelman et al., 2022) surrounding the hiring and inclusion of CAS, further research around the types of information conveyed (or missing) could suggest opportunities for better supporting the information needs of this group.

CONCLUSION

While casualization in academia is a growing issue and the surrounding literature is broad, many documents reviewed were largely conceptual and relatively few focused directly on the perspectives of those most directly affected. Additionally, despite the complex and disparate information environments of academic workers, understanding of the IP of CAS is quite limited. In particular, the relationships between IP and conditions of precarity, unseen emotional labor, and multiple temporalities seem crucial areas for further exploration. To further understand the current experiences of CAS, subsequent phases of this research project will involve quantitative and qualitative analysis of social media postings and in-depth, semi-structured interviews with both CAS and university department chairs.

ACKNOWLEDGEMENTS

This research project is supported in part by funding from the Social Sciences and Humanities Research Council of Canada. Assistance with literature searching was provided by Sara Cantler (McGill University) and Noel Sander (University at Buffalo).

REFERENCES

- al-Gharbi, M. (2020, May 1). Universities run on disposable scholars. *The Chronicle of Higher Education*.
<https://www.chronicle.com/article/universities-run-on-disposable-scholars/>
- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19–32. <https://doi.org/10.1080/1364557032000119616>
- Birdsell Bauer, L. (2018). *Precarious professionals: Non-tenure-track faculty in Southern Ontario universities* [Doctoral dissertation, University of Toronto]. TSpace Repository. <https://hdl.handle.net/1807/89903>
- Brady, M., & Briody, A. (2016). Strategic use of temporary employment contracts as real options. *Journal of General Management*, 42(2), 31–56.
- Brownlee, J. (2015). Contract faculty in Canada: Using access to information requests to uncover hidden academics in Canadian universities. *Higher Education*, 70(5), 787–805. <https://doi.org/10.1007/s10734-015-9867-9>
- Butler-Rees, A., & Robinson, N. (2020). Encountering precarity, uncertainty and everyday anxiety as part of the postgraduate research journey. *Emotion Space and Society*, 37. <https://doi.org/10.1016/j.emospa.2020.100743>
- CAUT (Canadian Association of University Teachers). (2020). *What impact is the pandemic having on post-secondary teachers and staff?* https://www.caut.ca/sites/default/files/covid_release-impacts_of_pandemic-en-final-ready_to_go.pdf
- Dolan, V. L. B. (2011). The isolation of online adjunct faculty and its impact on their performance. *The International Review of Research in Open and Distributed Learning*, 12(2), 62–77. <https://doi.org/10.19173/irrodl.v12i2.793>
- Foster, K., & Birdsell Bauer, L. (2018). *Out of the shadows: Experiences of contract academic staff*. CAUT. https://www.caut.ca/sites/default/files/cas_report.pdf
- Gannon, S., Kligyte, G., McLean, J., Perrier, M., Swan, E., Vanni, I., & van Rijswijk, H. (2015). Uneven relationalities, collective biography, and sisterly affect in neoliberal universities. *Feminist Formations*, 27(3), 189–216.
- Gelman, C., Gandel, J., & Bausman, M. (2022). A multi-faceted, adjunct-centered initiative to support part-time faculty. *Journal of Teaching in Social Work*, 42(1), 82–99. <https://doi.org/10.1080/08841233.2021.2013000>
- Gibson, A. N., & Martin, J. D. (2019). Re-situating information poverty: Information marginalization and parents of individuals with disabilities. *Journal of the Association for Information Science and Technology*, 70(5), 476–487. <https://doi.org/10.1002/asi.24128>
- Ivancheva, M. P. (2015). The age of precarity and the new challenges to the academic profession. *Studia Europaea*, 60(1), 39–47.
- Jolley, M., Cross, E., & Bryant, M. (2013). A critical challenge: The engagement and assessment of contingent, part-time adjunct faculty professors in United States community colleges. *Community College Journal of Research and Practice*, 38(2–3), 218–230. <https://doi.org/10.1080/10668926.2014.851969>
- Kezar, A. (2013). Examining non-tenure track faculty perceptions of how departmental policies and practices shape their performance and ability to create student learning at four-year Institutions. *Research in Higher Education*, 54(5), 571–598. <https://doi.org/10.1007/s11162-013-9288-5>
- Kezar, A., DePaola, T., & Scott, D. T. (2019). *The gig academy: Mapping labor in the neoliberal university*. Johns Hopkins University Press. <https://doi.org/10.1353/book.68032>
- Langan, D., & Morton, M. (2009). Through the eyes of farmers' daughters: Academics working on marginal land. *Women's Studies International Forum*, 32(6), 395–405. <https://doi.org/10.1016/j.wsif.2009.09.002>
- Leigh, J. S. (2014). "I still feel isolated and disposable": Perceptions of professional development for part-time teachers in HE. *Journal of Perspectives in Applied Academic Practice*, 2(2). <http://jpaap.napier.ac.uk/index.php/JPAAP/article/view/105>
- Loh, K. (2004). *Socialization experiences of part-time faculty: A study of socialization programs and employment longevity* (Publication No. 305211847) [Doctoral dissertation, American University]. ProQuest Dissertations & Theses Global.
- Lopes, A., & Dewan, I. (2014). Precarious pedagogies? The impact of casual and zero-hour contracts in Higher Education. *Journal of Feminist Scholarship*, 7(8), 28–42.
- Loveday, V. (2018). The neurotic academic: Anxiety, casualisation, and governance in the neoliberalising university. *Journal of Cultural Economy*, 11(2), 154–166. <https://doi.org/10.1080/17530350.2018.1426032>
- Mapes, M. (2019). Unjust precarity: Contingent faculty and the introductory communication course. *Communication Education*, 68(2), 246–252. <https://doi.org/10.1080/03634523.2019.1571213>
- Mason, O., & Megoran, N. (2021). Precarity and dehumanisation in higher education. *Learning and Teaching*, 14(1), 35–59. <https://doi.org/10.3167/latiss.2021.140103>
- McKenzie, P., & Davies, E. (2015). Multiple temporalities in personal information management. *Proceedings of the Annual Conference of CAIS / Actes Du Congrès Annuel de l'ACSI*. <https://doi.org/10.29173/cais908>
- McKenzie, P. J., & Davies, E. (2021). Documenting multiple temporalities. *Journal of Documentation*, 78(1), 38–59. <https://doi.org/10.1108/JD-11-2020-0196>

- Read, B., & Leathwood, C. (2020). Casualised academic staff and the lecturer-student relationship: Shame, (im)permanence and (il)legitimacy. *British Journal of Sociology of Education*, 41(4), 539–554. <https://doi.org/10.1080/01425692.2020.1748570>
- Reevy, G. M., & Deason, G. (2014). Predictors of depression, stress, and anxiety among non-tenure track faculty. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.00701>
- Rose, D. (2020). A snapshot of precarious academic work in Canada. *New Proposals-Journal of Marxism and Interdisciplinary Inquiry*, 11(1), 7–17.
- Roy, R., Oldfield, L., Simpson, A. B., Jolliffe Simpson, A., & Salter, L. (2021). Academic activism in the wake of a pandemic: A collective self-reflection from Aotearoa/New Zealand. *International Perspectives in Psychology: Research, Practice, Consultation*, 10(4), 215–227. <https://doi.org/10.1027/2157-3891/a000027>
- Ryan, C. A. (2017). *Part-time faculty and their sense of belonging* (Publication No. 2173846615) [Doctoral dissertation, University of Southern California]. ProQuest Dissertations & Theses Global.
- Savolainen, R. (2006). Time as a context of information seeking. *Library & Information Science Research*, 28(1), 110–127. <https://doi.org/10.1016/j.lisr.2005.11.001>
- Schofield, L. N. (2022). The workings and effects of precarious employment on Black women educators in development studies: An autoethnographic account of an international fieldtrip. *Progress in Development Studies*, 22(3), 1–7. <https://doi.org/10.1177/14649934221089085>
- Stewart-Robertson, O. (2022). Embracing theories of precarity for the study of information practices. *Journal of Documentation*, ahead-of-print. <https://doi.org/10.1108/JD-04-2021-0084>
- Tenure for the Common Good. (2020, April 30). A very stable and secure position? *Inside Higher Ed*. <https://www.insidehighered.com/views/2020/04/30/covid-19-shows-how-precarious-positions-contingent-faculty-actually-are-opinion>
- Willson, R. (2016). *Information in transition: Examining the information behaviour of academics as they transition into university careers* [Doctoral dissertation, Charles Sturt University]. Strathprints. <https://strathprints.strath.ac.uk/59763/>
- Willson, R. (2018). “Systemic managerial constraints”: How universities influence the information behaviour of hss early career academics. *Journal of Documentation*, 74(4), 862–879. <https://doi.org/10.1108/JD-07-2017-0111>
- Willson, R. (2019). Transitions theory and liminality in information behaviour research: Applying new theories to examine the transition to early career academic. *Journal of Documentation*, 75(4), 838–856. <https://doi.org/10.1108/JD-12-2018-0207>
- Willson, R., & Given, L. M. (2020). “I’m in sheer survival mode”: Information behaviour and affective experiences of early career academics. *Library & Information Science Research*, 42(2). <https://doi.org/10.1016/j.lisr.2020.101014>
- Willson, R., & Julien, H. (2020). Precarious academics: Information practices and challenges. *Proceedings of the Annual Conference of CAIS / Actes Du Congrès Annuel de l’ACSI*. <https://doi.org/10.29173/cais1131>
- Zheng, R. (2018). Precarity is a feminist issue: Gender and contingent labor in the academy. *Hypatia*, 33(2), 235–255. <https://doi.org/10.1111/hypa.12401>

Parasitic Knowledge Infrastructures: Data Reuse by Anthropogenic Climate Change Skeptics

Wofford, Morgan

University of Michigan, United States | mwofford@umich.edu

ABSTRACT

Stakeholders from academia, industry, funding agencies, and scholarly publishing are increasingly investing in open data partially in the hope that it will democratize science and promote more diverse data reuse. However, fewer studies examine how unconventional communities outside academia and industry use open data. Through an investigative digital ethnography, I observed the data practices of anthropogenic climate change (ACC) skeptics, specifically how they discuss, evaluate, and reuse open data. This poster focuses on the knowledge infrastructure that affords the data practices of ACC skeptics. I argue that ACC skeptics are building a parasitic knowledge infrastructure on the back of the climate science knowledge infrastructure it often seeks to discredit. Understanding the infrastructure that supports skeptics' data reuse can inform how we design policies and infrastructure to actualize open data's promises while minimizing its perils.

KEYWORDS

Data practices; knowledge infrastructures; open science; climate science; climate change skepticism

INTRODUCTION

One of the purported promises of the open data movement is its democratizing potential, affording people outside of orthodox research communities the ability to take advantage of data (Baack, 2015; Cavalier & Kennedy, 2016; Espinosa et al., 2014; Nielsen, 2011; Ricker et al., 2020; Zuiderwijk & Janssen, 2014). Open data activists envision data allowing individuals to generate their own knowledge representing a “democratization of information” (Baack, 2015). Citizen groups like the Anti-Eviction Mapping Project and the Mapping Police Violence Database use open government data to support the needs of the traditionally marginalized (*Anti-Eviction Mapping Project*, 2021; *Mapping Police Violence*, 2021). Tech-savvy unconventional data reusers view their activities in their communities as philanthropic endeavors (Kassen, 2021).

However, open data supporting the traditionally marginalized is not guaranteed, nor is it the only eventuality. Those with the most privilege and power are more likely to take advantage of open data, reinforcing extant hegemonic structures (Mirowski, 2018). Even if open data facilitates broader and more diverse data reuse, this can sometimes result in misuse or misinterpretation. For instance, local governments and private companies frequently misuse climate data in determining their financial climate-related risk (Fiedler et al., 2021). In another example, Lee et al. showed that anti-maskers use orthodox visualization techniques on open government data to support their unorthodox beliefs of removing mask mandates (Lee et al., 2021).

Broadly, unconventional data reusers could include citizen scientists, students, conspiracy theorists, community organizers, activists, teachers, and many more. Understanding how these data reusers leverage open data is essential to inform infrastructural development and support appropriate unconventional data reuse. The poster is based on an investigative digital ethnography of a single group of unconventional data reusers—ACC skeptics—examining their data practices, specifically how they discuss, evaluate, and reuse open climate data. The artifacts are the knowledge they construct from the data in the trimmings of scholarly discourse, including statistics, figures, graphics, and other computational models. Following conventional digital ethnography methods, I account for the infrastructures that afford this data reuse (Pink et al., 2015). I contextualize these as knowledge infrastructures that are “robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds” (Edwards, 2010).

METHODS

Through an investigative digital ethnography, I qualitatively traced how open data travels through the ACC skepticism community and becomes “knowledge,” following cascades of documents. Investigative digital ethnography combines the search for specific information or actors with longer-term observation (Friedberg, 2020). This methodology draws on various works on digital ethnography to understand how artifacts move through online communities (Donovan, 2019; Lewis & Marwick, 2017; Pink et al., 2015).

I used “deep lurking,” which draws on Clifford Geertz’s concept of “deep hanging out,” to participate by observing and systematically documenting the data practices of this community (Geertz, 1998; Lee et al., 2021). The units of analysis are data from repositories and the “knowledge artifacts” skeptics produce from those data. These knowledge artifacts include graphs, statistics, tables, charts, maps, computational models, code, and more. Because

of the nature of digital media, I collected data with an added focus on the knowledge infrastructure that supports the data reuse (Pink et al., 2015).

I began by identifying relevant publications, social media platforms, organizations, and influencers to define an ACC skepticism knowledge infrastructure. The preliminary list seeded my initial monitoring environment for various social media platforms. I then engaged with platform affordances, viewing, liking, and reposting content to influence recommendation algorithms. The recommendation algorithms offered new accounts to follow through algorithmic-assisted snowball sampling.

For my monitoring strategy, I began each observation period deep lurking on social media platforms. I archived these web pages if open data were discussed or reused. If these posts linked to blogs or organizations that mentioned open data, those web pages were archived. I observed the community for 75 hours from September 2021 through November 2021, archiving the relevant web pages and taking 15 pages of memos. All documents were uploaded into NVivo for qualitative coding. I synthesized themes across my data using grounded theory (Glaser & Strauss, 1967) adapted for social media analysis (Postill & Pink, 2012), inductively coding the data for emergent themes. These themes were then coded into higher-level concepts, constructing theories grounded in the data (Charmaz, 2014). I analyzed 125 files consisting of 2,867 pages of text over 67 hours with 19 pages of memos.

FINDINGS

I propose that ACC skeptics are building a parasitic knowledge infrastructure upon the back of the mainstream climate science knowledge infrastructure. A parasitic knowledge infrastructure generates, shares, and maintains its knowledge using components of another knowledge infrastructure while simultaneously weakening that infrastructure it relies upon. The “hypertransparency [of] open data, open code, commodity software tools, and alternative publication venues” allows skeptics to selectively use these components where feasible and advantageous (Edwards, 2019, p. 21). The digital ethnography shows that the parasitic knowledge infrastructure relies on open climate science data and tools to produce much of their “knowledge.” For instance, ACC skeptics use global temperature data from NASA and NOAA to attempt to disprove the very temperature trends these agencies publish.

The parasitic knowledge infrastructure utilizes mainstream tools when they provide affordances that skeptics cannot replicate. For instance, the World Meteorological Organization’s KNMI Climate Explorer allows users to investigate various climate data, including time series, model scenario runs, and more. KNMI is only mentioned as an available tool to visualize climate model scenario runs. This is an affordance that skeptical and agnostic tools do not offer. When suggesting tools to explore time series data, skeptics mention their alternatives rather than the KNMI Climate Explorer. Their devices more easily show data trends that support skeptical conclusions than mainstream tools. Many sites provide resource lists of mainstream climate science data, tools, and organizations alongside their skeptical counterparts, presented as equally legitimate.

Where infrastructural components are not available but could prove beneficial for their goals, skeptics critique the infrastructure for its lack of transparency. One skeptic explains that while National Centers for Environmental Information published a paper on how the Global Historical Climatology Network (GHCN) dataset is processed, the code used for data processing is held “specifically by the US government that we cannot test externally [or] even replicate it.” He concludes that the code must be made public to understand how the GHCN is calculated. In other instances, alternative infrastructural components are created to mimic mainstream climate science to maximize its perceived legitimacy. For example, one skeptical organization calls its lists of world temperature datasets, a data repository, and its weekly newsletter a journal.

This parasitic knowledge infrastructure masquerades as a trustworthy knowledge infrastructure trying to spread its knowledge in the trappings of scholarly discourse. As we open more components of knowledge infrastructures to the public, parasitic infrastructures are more likely to arise. How do we deal with the largely unavoidable misuse of open data? Is the misuse significant enough to address through changes in open science policies or the design of infrastructural components, such as data repositories?

CONCLUSION

The ACC skeptics’ data practices are made possible through their parasitic knowledge infrastructure built on the back of the mainstream climate science knowledge infrastructure. The parasitic knowledge infrastructure gains strength from the mainstream knowledge infrastructure’s increased openness. ACC skeptics’ data practices call into question the often uninterrogated assumption that open data is a universal and democratizing social good. We need a more comprehensive picture of what open data leads to in practice and how various groups, including unconventional communities, reuse data. This picture can inform how we should design open data policy and infrastructure to actualize open data’s promises while minimizing its perils.

ACKNOWLEDGMENTS

I am grateful to Andrea Thomer, Libby Hemphill, and Irene Pasquetto for their guidance and support on this research for my pre-candidacy project.

REFERENCES

- Anti-Eviction Mapping Project*. (2021). Anti-Eviction Mapping Project. <https://antievictionmap.com>
- Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data & Society*, 2(2), 2053951715594634. <https://doi.org/10.1177/2053951715594634>
- Cavalier, D., & Kennedy, E. B. (2016). *The rightful place of science: Citizen science*. Consortium for Science, Policy & Outcomes.
- Charmaz, K. (2014). *Constructing Grounded Theory*. SAGE.
- Donovan, J. (2019). Toward a Militant Ethnography of Infrastructure: Cybercartographies of Order, Scale, and Scope across the Occupy Movement. *Journal of Contemporary Ethnography*, 48(4), 482–509. <https://doi.org/10.1177/0891241618792311>
- Edwards, P. N. (2010). *A vast machine: Computer models, climate data, and the politics of global warming*. MIT Press.
- Edwards, P. N. (2019). Knowledge infrastructures under siege: Climate data as memory, truce, and target. In D. Bigo, E. F. Isin, & E. S. Ruppert (Eds.), *Data politics: Worlds, subjects, rights* (pp. 21–42). Routledge, Taylor & Francis Group.
- Espinosa, R., Garriga, L., Zubcoff, J. J., & Mazón, J.-N. (2014). Linked Open Data mining for democratization of big data. *2014 IEEE International Conference on Big Data (Big Data)*, 17–19. <https://doi.org/10.1109/BigData.2014.7004479>
- Fiedler, T., Pitman, A. J., Mackenzie, K., Wood, N., Jakob, C., & Perkins-Kirkpatrick, S. E. (2021). Business risk and the emergence of climate analytics. *Nature Climate Change*, 11(2), 87–94. <https://doi.org/10.1038/s41558-020-00984-6>
- Friedberg, B. (2020). Investigative Digital Ethnography: Methods for Environmental Modeling. In *Media Manipulation Casebook*. <https://mediamanipulation.org/research/investigative-digital-ethnography-methods-environmental-modeling>
- Geertz, C. (1998). Deep hanging out. *The New York Review of Books*, 45(16), 69–72.
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Aldine Publishing.
- Kassen, M. (2021). Understanding motivations of citizens to reuse open data: Open government data as a philanthropic movement. *Innovation*, 23(1), 44–70. <https://doi.org/10.1080/14479338.2020.1738940>
- Lee, C., Yang, T., Inchoco, G. D., Jones, G. M., & Satyanarayan, A. (2021). Viral Visualizations: How Coronavirus Skeptics Use Orthodox Data Practices to Promote Unorthodox Science Online. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–18. <https://doi.org/10.1145/3411764.3445211>
- Lewis, B., & Marwick, A. E. (2017). Media Manipulation and Disinformation Online. *Data & Society*. <https://datasociety.net/library/media-manipulation-and-disinfo-online>
- Mapping Police Violence*. (2021). Mapping Police Violence. <https://mappingpoliceviolence.org>
- Mirowski, P. (2018). The future(s) of open science. *Social Studies of Science*, 48(2), 171–203. <https://doi.org/10.1177/0306312718772086>
- Nielsen, M. A. (2011). *Reinventing discovery: The new era of networked science*. Princeton University Press.
- Pink, S., Horst, H., Postill, J., Hjorth, L., Lewis, T., & Tacchi, J. (2015). *Digital Ethnography: Principles and Practice*. SAGE.
- Postill, J., & Pink, S. (2012). Social Media Ethnography: The Digital Researcher in a Messy Web. *Media International Australia*, 145(1), 123–134. <https://doi.org/10.1177/1329878X1214500114>
- Ricker, B., Cinnamon, J., & Dierwechter, Y. (2020). When open data and data activism meet: An analysis of civic participation in Cape Town, South Africa. *The Canadian Geographer / Le Géographe Canadien*, 64(3), 359–373. <https://doi.org/10.1111/cag.12608>
- Zuiderwijk, A., & Janssen, M. (2014). Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly*, 31(1), 17–29. <https://doi.org/10.1016/j.giq.2013.04.003>

Persuasion in Audio-based Social Media: A Clubhouse Interview Study

Xiao, Lu Syracuse University, USA | lxiao04@syr.edu
Wang, Yueyue Syracuse University, USA | ywang153@syr.edu
Wei, Xinyi Syracuse University, USA | xwei21@syr.edu

ABSTRACT

Clubhouse is an audio-based social media platform launched in March 2020. In this platform, users create chat rooms for discussions around a topic of their choice. Organizers of these discussions can decide which listener in the room to be given the right to speak. To explore their experiences of persuading others and of being persuaded in Clubhouse discussions, we interviewed 16 Clubhouse users. One's credibility is important to the perceived persuasiveness of one's argument. In Clubhouse, one may establish their credibility through their profile pages where they introduce themselves such as their backgrounds, experiences, affiliations, passions, etc. Hence, our interviews also probed whether speakers' profile pages are used to check their credibility.

KEYWORDS

Persuasion, Clubhouse, credibility

INTRODUCTION

People engage in complex social interactions in social media nowadays (Dey et al., 2017; Nguyen, Dabbish, & Kiesler, 2015). One of such interactions is persuasion, commonly considered as "human communication that is designed to influence others by modifying their beliefs, values, or attitudes" (Simons, 1976, p.21). A lot of research has been conducted to understand the processes and mechanisms in online persuasion, e.g., to explore the linguistic indicators that signify an online message's perceived persuasiveness (Tan et al., 2016), or the types of persuasion strategies applied in the message (Chen, Xiao, & Mao, 2021). The contextual factors in online communication are also found to influence the perceived persuasiveness of a comment, e.g., how committed one is to their view (Mensah, Xiao, & Soundarajan, 2019), the entry order of the comments (Xiao & Khazaei, 2019), and the interactions among the participants (Jo et al., 2018). These studies are contextualized in text-based social media.

In March 2020, Clubhouse, an audio-based social media platform, was launched. Although the application was not available to non-iOS operating systems till May 2021 and the participation is by invitation only, Clubhouse has gained popularity quickly in social media use. It allows users to organize a chat room about a topic and invite people to join the room or make it open to any Clubhouse user. In such a chat room, participants are invited to speak to the room by the host through audios as opposed to text-based chats. Clubhouse users in general take one or more of the three roles in a chat room: listener, speaker, and organizer. These features of Clubhouse make it a platform for developing business (Medvedenko & Neusikhin, 2021). Researchers also find that Clubhouse helps people establish social relationships through voice communications (Jung et al., 2022). In addition, prior studies suggest that online environment affects the persuasion strategies and persuasion power in communication (Guadagno & Cialdini, 2002). These studies imply that online persuasion happens in Clubhouse and its processes and mechanisms may differ from the persuasion in text-based social media environments. Yet, to our best knowledge, there hasn't been an online persuasion study in the context of audio-based social media such as Clubhouse.

As a step to address this research gap, we conducted a semi-structured interview study to understand people's experiences in Clubhouse chat rooms, specifically their experiences of persuading others and being persuaded. Also, the credibility of the author is found to be important in the perceived persuasive power of the statement (Murphy, 2001). In Clubhouse, one may establish their credibility through their profile pages where they introduce themselves such as their backgrounds, experiences, affiliations, passions, etc. Hence, our interviews also probed whether speakers' profile pages are used to check their credibility.

INTERVIEW STUDY – DATA COLLECTION

Through the combination of the researchers' social network and the advertisement in social media, we recruited seven female and nine male Clubhouse users. All provided their informed consent before the interview and received compensation afterwards. Their age ranges from 19 to 39. Eight interviewees are from China, four from India, and four from USA. There are eight graduate students, three just graduated from college, one will start his/her college education in the fall, and four college students. Participants used various social media tools for online discussion and the three most mentioned tools are Instagram, Facebook, and Weibo (a Chinese social media platform). In terms of the Clubhouse role, most of them have only been a listener. One was mainly a speaker, and one has played all three roles. The interviews were conducted in Zoom lasting from 25 to 70 minutes with the average being 40 minutes.

FINDINGS

Persuasion and Being Persuaded – Experiences in Clubhouse

Half of the interviewees said they recalled being persuaded when they used Clubhouse. The responses suggest that in Clubhouse one potentially high persuasion power for the discussion is that multiple or many people share their similar personal experiences. For instance, interviewee #16 said, “But having heard stories of all these people who are actually making it, or even failing in it, but still giving it, still not giving up and still trying and trying to do as much as they could, that did make me think a lot that ...”. In addition, by having people of different regions to be together and talk about the related geopolitical issues, this has created a space for people to better understand and be persuaded. Interviewee #3 said, “at that time the room has about 3000 people and there are young adults from mainland China and Taiwan, many people probably has changed how to view the Taiwan issue in that night.”. Another persuasion power from the discussion is at the affective level. Interviewee # 14 explained how she/he felt being persuaded – “I was scared to watch the show ... But because I entered the room and I talked to a lot of people regarding it, it kind of helped me just enjoy the show and learn more about it”.

Only three interviewees said that they had tried to persuade others in Clubhouse. They mainly shared their personal experiences as a strategy to convince the others – interviewee # 14 said, “I shared my personal experience about the shows. And I was like, you know, if you watch this genre, it's not always about love stories... So basically, I encourage the person or maybe I can say I persuaded the person to watch gay dramas and to enjoy them in a certain way”. The rest interviewees commented that they did not have the persuasion experiences in Clubhouse because they were only listeners.

Speakers' profile pages for credibility check

Seven interviewees said they often checked the speakers' profiles, but their reasons of doing so vary. Feeling that this is important information that helps them check the speakers' credibility, interviewee #4 said, “I want to know this speaker's background and the level of domain knowledge. This helps me judge whether what the person said is believable or not”. Interviewee # 14 also commented that – “when I started talking to this person, I made sure that whatever he's written in his profile actually matched to what he thought that his opinions actually match”. Interestingly, interviewee #7 stated that as Clubhouse collects less personal information compared to the other social media platforms, it is hard for him to form a mental image of the speaker without checking their profiles. The other two reasons are to connect with the speaker and to satisfy one's curiosity about the speaker. Interestingly, interviewee #12 said while he does not check the speakers' profiles, he does check the room organizers' profiles – “(I check) how many followers are there, how many (they are) following so I see their neighbors in Graph theory” and he uses that information to help decide whether he should join that chat room.

When asked whether and how they would verify the information in people's profiles, all said they do not do so except interviewees #3, #14, and #16. Interviewee #3 did not give the reason of doing so. Interviewee #16's action seemed to be more driven by his curiosity to know the person than to check if the provided information is true or not – “if they've given their Instagram handle every they've given their LinkedIn ID, I probably go there, just go through whatever they have done”.

Ten interviewees said it doesn't matter to them if the speakers do not have information on their profile pages. Most of them do not think the content of the profile is relevant to them and they care more about what speakers say. For example, Interviewee#5 states, “I actually think it's fine, because if you know, it's rare, but if the presenter he clicks in, and he's like, his profile is empty, but he's saying something that I think makes a lot of sense that's fine, I don't really mind”. However, six interviewees said the content on the speakers' profile pages matter to them because the content helps them evaluate the speaker's credibility or decide whether the speaker is an interesting person to listen to. Interviewee#10 said that “I would tend to maybe give less credit to what they're saying if you know, sort of a more anonymous looking profile page”. Interviewee#12 thinks that profile pages matter to him because he needs to see the description to decide whether he wants to listen to that speaker.

CONCLUSION

We explore people's experiences of persuading others and being persuaded by others in Clubhouse. We also probe whether people use speakers' profile pages to check their credibility. Our findings suggest that people mainly use their personal stories as a persuasion strategy. Voice is effective for establishing social relationships bringing in more interactivity and intimacy (Jung et al., 2022). Compared to text-based online discussions, there are two special aspects about the persuasion mechanisms in Clubhouse discussions. First, as the platform allows many people to be present, an argument may be perceived stronger when many share similar experiences. Second, the presence of many people and their verbal support can add an affective aspect to the persuasion.

REFERENCES

Chen, S., Xiao, L., & Mao, J. (2021). Persuasion strategies of misinformation-containing posts in the social media. *Information Processing & Management*, 58(5), 102665

- Dey, S., Duff, B., Karahalios, K., & Fu, W. T. (2017). The art and science of persuasion: not all crowdfunding campaign videos are the same. *In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pp. 755-769
- Guadagno, R. E., & Cialdini, R. B. (2002). Online persuasion: An examination of gender differences in computer-mediated interpersonal influence. *Group Dynamics: Theory, Research, and Practice*, 6(1), pp. 38–51
- Jo, Y., Poddar, S., Jeon, B., Shen, Q., Rosé, C. P., & Neubig, G. (2018). Attentive interaction model: Modeling changes in view in argumentation. *arXiv preprint arXiv:1804.00065*
- Jung, K., Park, Y., Kim, H., & Lee, J. (2022). Let's Talk@ Clubhouse: Exploring Voice-Centered Social Media Platform and its Opportunities, Challenges, and Design Guidelines. *In CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pp.1- 6
- Medvedenko, D., & Neusikhin, V. (2021). Clubhouse as a new platform for developing business.
- Mensah, H., Xiao, L., & Soundarajan, S. (2019). Characterizing susceptible users on reddit's changemyview. *In Proceedings of the 10th International Conference on Social Media and Society*, pp.102-107
- Murphy, P. K. (2001). What makes a text persuasive? Comparing students' and experts' conceptions of persuasiveness. *International Journal of Educational Research*, 35(7–8). pp. 675–698.
- Nguyen, D. T., Dabbish, L. A., & Kiesler, S. (2015). The perverse effects of social transparency on online advice taking. *In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 207-217.
- Simons, H. W. (1976). Persuasion. *Reading: Addison-Wesley*, 21
- Tan, C., Niculae, V., Danescu-Niculescu-Mizil, C., & Lee, L. (2016). Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. *In Proceedings of the 25th International Conference on World Wide Web*, pp.613-624.
- Xiao, L., & Khazaei, T. (2019). Changing Others' Beliefs Online: Online Comments' Persuasiveness. *In Proceedings of the 10th International Conference on Social Media and Society*, pp.92-101.

Performance, Usability, and User Experience of Rayyan for Systematic Reviews

Yu, Fei

University of North Carolina at Chapel Hill, USA | feifei@unc.edu

Liu, Chuanrui

Cornell University, USA | cl2328@cornell.edu

Sharmin, Selina

University of North Carolina at Chapel Hill, USA sselina@email.unc.edu

ABSTRACT

This study investigated performance, usability, and user experience with Rayyan, a systematic review (SR) tool. A remote usability inspection was conducted via the institutional Zoom platform. Each participant was randomly assigned to two groups and completed two citation screening sessions independently. Data were collected through pre- and post-test questionnaires, screening sessions, and Systems Usability Scale (SUS) surveys. Thirteen participants completed the study. We found significant accuracy difference between Web Rayyan and its mobile app ($t(12)=1.877$ $p=.042$), but no statistically sequential effect on screening performance. The average SUS scores of both Rayyan versions were below 68. More participants (77%) preferred Web Rayyan to the mobile app. The mobile Rayyan has potential if its user interface design can prevent errors or allow easy action reversal. As the first controlled experiment investigating both Web and mobile Rayyan, this study provides empirical evidence to guide the design, development, and selection of SR tools.

KEYWORDS

Usability, Performance, User experience, Rayyan, Systematic Reviews

INTRODUCTION

Playing a vital role in evidence-based decision making, systematic reviews (SRs) help identify available evidence for a specific research question through comprehensive literature search, citation screening, evidence evaluation, and result synthesizing. However, conducting an SR is time- and labor- intensive, especially the citation screening process (Wang, Sharmin, Wang, & Yu, 2021). A variety of applications have been developed to support or expedite SRs and many are embedded artificial intelligence (AI) features. However, the performance and usability of these SR tools lack validation and evaluation, which potentially prohibit their adoption.

Rayyan (<https://www.rayyan.ai/>) is a free SR application developed to facilitate collaborative citation screening. With built-in machine learning and AI, Rayyan developers aim to reduce screening time and workload. Although two studies previously tested and validated Rayyan AI algorithms (Olofsson et al. 2017; Ouzzani et al. 2016), no formal controlled user experiment has been conducted on Rayyan. This study was designed to investigate the performance, usability, and user experience with Rayyan of both Web and Mobile versions in a controlled experiment.

METHODS

A remote usability inspection was conducted via our institutional Zoom platform. Participants were recruited from the *User Interface Design* class and all students were trained for user interface (UI) design principles and evaluation prior to this study. Participants were randomly assigned to two groups and each group completed two screening sessions with the same number of citations (i.e., 25). In Session 1, Group 1 participants used the Web Rayyan on their computers while Group 2 participants used Rayyan mobile app pre-loaded to their phones. After finishing the first session, Group 1 proceeded to screen with Rayyan mobile app in Session 2 while Group 2 used the Web Rayyan. Although the citations for screening in each session have the same number, but records were different.

Data were collected through (1) a pre-test questionnaire for participants demographics, (2) screen session logs in Rayyan application (time & screening decisions), (3) System Usability Scale (SUS) surveys (Sauro, n.d.), and (4) a post-test survey and discussion for user experience. Quantitative data analysis was performed using IBM SPSS statistics and major themes were summarized from participants' comments and discussion.

RESULTS

A total of 13 participants completed the usability inspection entirely. Seven of them were 18-24 years old (YR), 4 were 25-34 YR, and 2 were 45-54 YR. Eleven participants are native English speakers and 2 are not. Ten participants did not have any SR experience and never used an SR tool, but 3 of them conducted less than 3 SRs previously. No participants have used Rayyan prior to this study.

Performance

There was no sequential effect on all performance metrics (time, accuracy, recall, and precision) between Session 1 and 2 ($p >.05$). In addition, there was no difference in time, recall, and precision between Web Rayyan and the

mobile app ($p > .05$) (Table 1). However, there was statistical difference in accuracy between Web Rayyan ($mean = .59, std = .23$) and mobile app ($mean = .66, std = .19$) at one-sided ($t(12)=1.877 p=.042$).

Metrics	Session 1 (m, std)	Session 2 (m, std)	Web (m, std)	Mobile (m, std)
Time	13.85, 6.85	10.92, 5.69	12.69, 6.56	12.08, 6.38
Accuracy	0.63, 0.24	0.61, 0.19	0.59, 0.23	0.66, 0.19
Recall	0.46, 0.52	0.77, 0.44	0.54, 0.52	0.69, 0.48
Precision	0.10, 0.15	0.07, 0.05	0.058, 0.07	0.11, 0.14

Table 1. Performance Measurement (m = mean; std = standard deviation)

Usability

The average SUS scores of Web Rayyan and mobile app were 63.85 ($std = 15.5$) and 56.73 ($std = 23.04$) respectively. Paired t-tests show the SUS scores of Statement 1 (“I think that I would like to use this system frequently”) and Statement 3 (“I thought the system was easy to use”) have statistical difference between Web Rayyan and the mobile app ($p < .05$). However, no other statement responses associated with the two Rayyan versions were statistically different ($P > .05$).

User Experience

Ten participants preferred Web Rayyan to the mobile app while 3 preferred the mobile app. Similarly, 9 out of 13 participants would choose Rayyan in the future for an SR while 2 chose “Maybe” and 1 “No.” The positive user experience with Web Rayyan included “Easier to read and distinguish between items,” “more familiar with the format,” and “the interface was more natural” while participants also liked the mobile app for being “quicker to screen articles,” “easier to navigate,” and “easier to read text.” The major negativity associated with the mobile app was that “easily selecting the wrong decision,” and “accidentally clicked on the wrong button.”

DISCUSSION

Although the participants have varied demographic characteristics, within-subject comparison (Session 1 vs. 2; Web vs. Mobile) shows that (1) user performance was not affected by the sequence of which Rayyan version to use first, and screening with the mobile app produced slightly higher accuracy than with Web Rayyan. (2) Neither the SUS average scores of Web (63.85) nor the mobile app (56.73) was above 68, a benchmark SUS score (Sauro, n.d.). (3) The majority of participants preferred Web Rayyan to the mobile app due to ease to use. Nevertheless, the mobile Rayyan has great potential if its UI design can be improved by implementing error-prevention or allowing users to easily reverse actions and correct mistakes.

This study has a couple of limitations. First, the participants included students at different academic levels. Although the majority (70%) were graduate or doctoral students, undergraduates were also included (30%), who are usually less likely to perform a SR study. Second, this usability inspection took place remotely. Variables that could impact online user performance and experience such as internet speed, devices, operation systems, and web browsers were not explored in this project, but warrant future studies.

CONCLUSION

This is the first controlled experiment exploring the difference of an SR application’s Web and mobile versions regarding performance, usability, and user experience. It provides empirical evidence to guide the design, development, and selection of SR tools.

REFERENCES

- Olofsson, H., Brolund, A., Hellberg, C., Silverstein, R., Stenström, K., Österberg, M., & Dagerhamn, J. (2017). Can abstract screening workload be reduced using text mining? User experiences of the tool Rayyan. *Research Synthesis Methods*, 8(3), 275–280.
- Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan-a web and mobile app for systematic reviews. *Systematic Reviews*, 5(1), 210.
- Sauro, J. (n.d.). Measuring Usability with the System Usability Scale (SUS) – MeasuringU. Retrieved from (<https://measuringu.com/sus/>).
- Wang, M., Sharmin, S., Wang, M., & Yu, F. (2021). A Mixed-Method Usability Study on User Experience with Systematic Review Software. In *Proceedings of Association of Information Science and Technology*, 58, 346-356.

Perception and Use of COVID Contact Tracing Mobile Applications in New York State (NYS)

Yuan, Xiaojun

University at Albany, New York, USA | xyuan@albany.edu

Bennett Gayle, DeeDee

University at Albany, New York, USA | dmbennett@albany.edu

Dadson, Yvonne

University at Albany, New York, USA | ydadson@albany.edu

Jung, Ellie

University at Albany, New York, USA | sjung@albany.edu

ABSTRACT

The study explores the use of COVID-19 related apps for contact tracing deployed in New York State (NYS). The project seeks to understand potential differences in perception, adoption, or privacy concerns among racial and ethnic populations and across age groups. Using the Antecedent-Privacy Concerns-Outcomes (APCO) framework and the perceived usefulness construct, this study explores factors influencing the individual level adoption of these apps. Data collected from 120 Amazon Mechanical Turkers located in NYS was analyzed. The results indicate that race and gender are important factors to consider in expanding the Antecedent-Privacy Concerns-Outcomes (APCO) framework. Specifically, race impacted the perception of the seriousness of the pandemic, with Asians and Black being serious about the pandemic. Age played a role in privacy and security concerns. The youngest group of respondents, aged 18-24, did not have many privacy and security concerns about mobile apps. These results provided empirical results and evidence that can contribute to the expansion of the APCO model and help further the model's development.

KEYWORDS

Contact tracing; Privacy concerns; Perceived usefulness; Mobile applications; COVID-19

INTRODUCTION

Many countries resorted to and mandated contact tracing apps to help mitigate the spread of COVID-19. Despite the varying public policy efforts, these tracking apps have not been fully effective due to public concerns over data privacy (Russo et al., 2021). Major concerns included questions about the usefulness of the apps and the privacy of the individual data, which can affect mass acceptance (Akinbi, Forshaw, & Blinkhorn 2021). Contact tracing has been employed in extraordinary ways to mitigate the transmission of infectious diseases (Fairchild et al., 2020; Wacksman, 2021). Throughout the pandemic, contact tracing emerged as a fundamental mitigation strategy to minimize the spread of COVID-19. A key challenge to the effective use of such apps relates to social policies that determine an individual's autonomy to participate. Cultural differences and social norms often govern citizens' choice to adopt and use these apps. The New York public health officials and healthcare providers were hit fast, with the most significant spike in cases and subsequent deaths between late March through early May 2020 (Rothfeld et al., 2020; New York Times, 2020). This led to a public health initiative responding to the COVID-19 public health emergency. Test & Trace Corps (2022) began on June 1, 2020, as a coordinated effort of doctors, public health professionals, and community advocates to contain the transmission of COVID-19 in NYC (NYHealthandhospitals.org). The COVID Alert NY mobile application was launched by New York State in 2021 and worked alongside the Exposure Notification (launched by Google and Apple). The important societal issues resulting from this study on the increased use of contact tracing devices and other technologies include the perceived usefulness of mobile apps for COVID-19 mitigation and privacy and security concerns. Furthermore, this study focuses on racial and ethnic minority population groups who may be more reluctant to adopt and, therefore, not participate in mobile application mitigation strategies.

THEORETICAL FRAMEWORK

Using the Antecedent-Privacy Concerns-Outcomes (APCO) framework (Figure 1) (Polites & Karahanna, 2013) and the perceived usefulness construct, this study explores what factors influence the individual level adoption of these apps. Also, we investigate what factors influence the individual level privacy of COVID-19 mobile applications by including a more comprehensive set of demographic considerations such as race/ethnicity, income, age, education, and political affiliations. Additionally, because the technology is related to the pandemic, we included questions about trust and awareness regarding the pandemic and contact tracing. The following research questions (RQs) were used to examine these factors: RQ1: How does the perceived usefulness of mobile apps for COVID-19 contact tracing differ by race and age? RQ2: How do privacy and security concerns influence the adoption of COVID-19-related mobile apps? Are there differences by race and age?

METHODOLOGY

A survey was administered via Amazon Mechanical Turk to explore the perceived usefulness perceptions of NYS mobile apps for COVID-19 and privacy and security concerns of the technologies used. We first recruited 120

Turkers located in New York State to answer this survey. In the survey, Turkers were shown pictures of the mobile apps (COVID Alert NY and Exposure Notification) and then asked a series of questions regarding the apps.

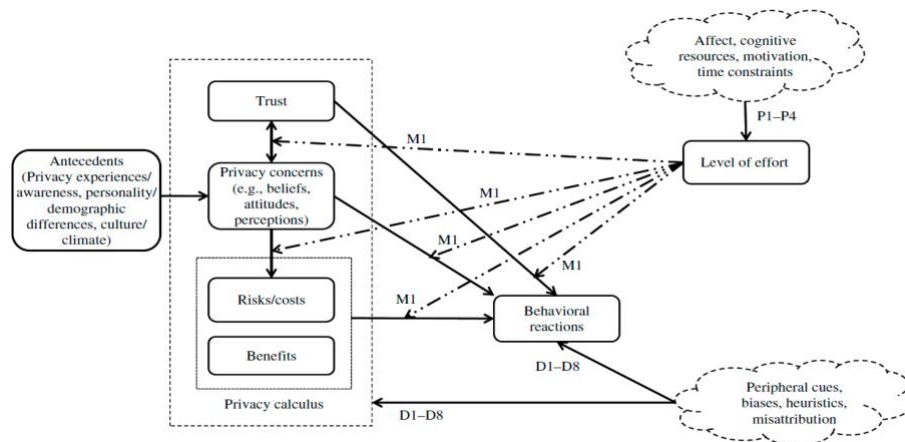


Figure 1. The Antecedent-Privacy Concerns-Outcomes (APCO) framework (Polites & Karahanna, 2013)

RESULTS AND DISCUSSION

There was an equal distribution of male and female respondents, with one person identifying as transmasculine and one who preferred not to identify their gender. Nearly 73% of the respondents self-identified as White, 12.5% Asian, 11% African American or Black, 6% other. Additionally, 15% of respondents also identified as Hispanic or Latino, where all participants were asked about their race and their Hispanic ethnicity. Most of the respondents were between the ages of 25 and 44 (68%). Most respondents had at least a bachelor's degree (74%). Over half of the respondents supported the democratic party in handling COVID (52%), 20% supported the republican party in handling the pandemic, and 18% did not support any party. Others did not know in terms of political identity; half of them identified as democratic, nearly a quarter of respondents identified as independent, and 20% of them identified as republican. A small number of respondents identified as something else politically, and others did not know their political identity. Among 119 respondents who reported their annual income from the last 12 months, a bit over 10% responded that their income was less than \$25,000.

Results show that the seriousness in which the pandemic was perceived varied by race, with Asian and Black respondents indicate more serious perception. Individual perceptions about the pandemic were asked to establish a baseline related to risk perception that may relate to overall trust and privacy concerns regarding technologies used primarily to mitigate the pandemic. Other questions were asked to directly identify how individuals perceive contact tracing and contact tracing mobile applications. The perceived helpfulness of contact tracing to minimize the risk from COVID and the perceived usefulness of contact tracing mobile applications are closely aligned by race. While the perceived helpfulness ranged from neutral to helpful, respondents saw the most benefits from using the contact tracing applications hovered mostly around neutral for each race and across different activities, such as work, education, and recreation. It may relate to the fact that the COVID-19 outbreak was first in Wuhan, China, and that Black/African American populations in the US have been disproportionately affected (in terms of infection and mortality) since the first outbreak.

With respect to privacy concerns, Asian respondents indicate less privacy concerns with use of the contact tracing mobile application, where Black respondents indicated more privacy concerns with the mobile app. Populations outside of Asian, Black or White, indicate privacy concerns causing them to change the permissions of the apps they download to their mobile devices. The responses show more significant variation based on age. The youngest age group, 18 to 24, showed the least willingness to use the contact tracing mobile app, even if it was mandatory to use. Concern about privacy and security was more likely to prevent respondents aged 25 - 44 from using contact tracing apps.

CONCLUSION

This study examined the perceived usefulness of the NY contact tracing apps and related privacy concerns of users. Our preliminary results indicate that race and age impacted the perception of mobile apps and the privacy and security concerns of users. Regarding privacy concerns, race and age may play a role. Asian respondents indicate more minor privacy concerns with the contact tracing mobile application than those Black respondents. Populations outside of Asian, Black, or White claimed that they changed the permission of mobile apps because of privacy concerns. In terms of age, the youngest age group (18–24-year-old) should be paid special attention since they are more likely to change permissions of the apps, and their privacy perceptions of mobile apps would prevent their use of a mobile app.

REFERENCES

- Akinbi, A., Forshaw, M., & Blinkhorn, V. (2021). Contact tracing apps for the COVID-19 pandemic: A systematic literature review of challenges and future directions for neo-liberal societies. *Health Information Science and Systems*, 9(1), 18. <https://doi.org/10.1007/s13755-021-00147-7>
- Fairchild, A. L., Gostin, L. O., & Bayer, R. (2020). Contact tracing's long, turbulent history holds lessons for COVID-19. (n.d.). *Contact Tracing's Long, Turbulent History Holds Lessons for COVID-19*. Retrieved April 15, 2022, from <https://news.osu.edu/contact-tracings-long-turbulent-history-holds-lessons-for-covid-19/>
- Polites, G. L., & Karahanna, E. (2013). The embeddedness of information systems habits in organizational and individual level routines: Development and disruption. *Mis Quarterly*, 221-246.
- Rothfeld, M., Sengupta, S., Goldstein, J., & Rosenthal, B. M. (2020, March 25). 13 Deaths in a Day: An 'Apocalyptic' Coronavirus Surge at an N.Y.C. Hospital. *The New York Times*. <https://www.nytimes.com/2020/03/25/nyregion/nyc-coronavirus-hospitals.html>
- Russo, M., Ciccotti, C. C., Alexandris, F. D., Gjinaj, A., Romaniello, G., Scatorchia, A., & Terranova, G. (2021, August 2). A cross-country comparison of contact-tracing apps during COVID-19. *VoxEU.Org*. <https://voxeu.org/article/cross-country-comparison-contact-tracing-apps>
- Test & Trace Corps (2022). Test & Trace Corps | NYC Health + Hospitals. (n.d.). Retrieved June 7, 2022, from <https://www.nyhealthandhospitals.org/test-and-trace/>
- New York Times (2020). U.S. Hits Another Record for New Coronavirus Cases—The New York Times. (n.d.). Retrieved June 7, 2022, from <https://www.nytimes.com/2020/07/09/world/coronavirus-updates.html>
- Wacksman, J. (2021). Digitalization of contact tracing: Balancing data privacy with public health benefit. *Ethics and Information Technology*, 23(4), 855–861. <https://doi.org/10.1007/s10676-021-09601-2>

Creating an Undergraduate Data Science Program in a Disciplinary and Institutional Context in an iSchool

Zhang, Yin

Kent State University, U.S.A. | yzhang4@kent.edu

Zeng, Marcia Lei

Kent State University, U.S.A. | mzung@kent.edu

ABSTRACT

Data science (DS) programs have grown to address the workforce demand in an increasingly data-driven society. Given that DS is highly interdisciplinary and individual institutions have various academic structures, creating a new DS program needs to consider many factors. This poster details the process of how to create an undergraduate DS minor program with considerations of the following key questions: (1) What is the general background of DS education landscape and iSchools? (2) What is the institutional context of DS education, i.e., what are gaps in the institutional DS education? How does the proposed program fit in and distinguish itself from other DS programs on campus? (3) How to navigate the DS approval process with multiple DS programs on campus? This work will help those who plan to create and revise a DS program that benefits students and fits in the broad DS education landscape and their own institution.

KEYWORDS

Data Science Education; Data Science; Information Science; iSchool

INTRODUCTION

Data are everywhere and have become part of our daily life and work. Private sectors, government agencies, and research, academic, and cultural heritage institutions have all been major players in this emerging data economy and culture, as they share and consume massive amounts of data. Directly relevant to the research communities and academic institutions, Open science and data sharing is becoming the future of scientific research (Gewin, 2016). As the needs of a data-driven society are growing and evolving rapidly, there is a shortage of workers skilled in dealing with data challenges. In response to these needs, data-related academic programs have been developed by various disciplines and units. A recent systematic review of data science education programs shows that disciplines and units such as Mathematics and Statistics, Computer Science, Business, and Library and Information Science (LIS) are the leading disciplines/units offering such programs. Some institutions are offering an interdisciplinary data science and education program (Wu, 2019).

The purpose of this poster is to share the process of developing an undergraduate minor in DS at an iSchool that fills the gaps in the existing curriculum in an institution with existing DS programs offered by other disciplines. The process, considerations, and lessons learned would be helpful for other iSchools to create DS programs that make unique contributions in their institutional context and benefit students across disciplines.

APPROACH AND PROCESS

This DS minor is created through a collaborative approach and based on a systematic data collection, review, and analysis of the following sources: (1) data education curriculum and programs from multiple perspectives: unit, disciplinary, institutional, and professional; (2) data education literature and best practices; (3) the job market to identify key areas of knowledge and skills; and (4) data processing and management standards and practices.

Background of the DS Education Landscape and iSchools

DS is multi- and interdisciplinary in nature (Cao, 2017; Raban & Gordon, 2020). Many disciplines have started to establish a disciplinary identity in the DS research and education landscape, including LIS (Shah et al., 2021; Virkus & Garoufallou, 2019, 2020). Our analysis of DS job advertisements found 200+ subject domain areas in the degree requirement specifications, reflecting market need for a broad range of DS education programs at all levels. To address DS education in the iField, iSchools Organization established two curriculum committees: iSchools Data Science Curriculum Committee (iDSCC, 2019-2021), with core DS curriculum recommendations (iDSCC, 2019a, 2019b, 2020, 2021), and iSchools Digital Humanities Curriculum Committee (iDHCC, 2019-2021), with core DH curriculum recommendations (Walsh & Zeng, 2021; Walsh et al., 2021).

Institutional Context of DS Education

We have conducted a detailed analysis of existing DS curriculum in [the institution name] to address the following two questions: (1) What are gaps in DS education at our institution? (2) How does this proposed program fit in and distinguish itself from other DS programs on campus? As the result of such analysis, we concluded that the major foci and emphases of current DS offerings are in two areas: (a) data analytics in respective disciplinary contexts, and (b) mathematical, statistical, and computational approaches to data science. We identified the major gaps that an

iSchool program can address for a holistic understanding of data with the following program learning outcomes for students to mainly study the tools and theories dealing with data to obtain the following knowledge and abilities:

- Explain the role and impact of data on people and society as well as the interdisciplinary and disciplinary nature of data science.
- Describe the information science approach to data in an applied area.
- Recognize various stages of data in the data lifecycle and what each stage entails.
- Obtain basic knowledge and skills associated with various stages of the data lifecycle.
- Identify and describe data standards, principles, and best practices for collecting, curating, preserving, and making data more accessible, discoverable, retrievable, and usable.
- Describe research data management challenges and opportunities in archives, academic and public libraries, and public life.
- Develop the skills to create actionable data management plans to support sustainable projects.
- Identify and describe issues and best practices in the creation, management, curation, access, and reuse of research data.
- Discuss human-centered, ethical, and contextual considerations of data.

A detailed description of the curriculum for the program will be presented during the poster session.

Lessons Learned and Recommendations

It took three years from the initial conception to the final approval of the DS program at [institution name]. Many lessons have been learned in building the curriculum and obtaining support from other disciplines/units on campus. It has been a very challenging and daunting process to deal with DS politics as more disciplines and academic units are offering DS programs. Throughout the process, we have met and articulated our values to institutional administrators, colleagues, prospective employers, and students alike about what and how this program complements and distinguishes itself from other existing DS programs while contributing to the broader DS education spectrum and career opportunities for students from various disciplines. Details will be shared at the poster session.

CONCLUSION

While the interdisciplinary nature of DS offers great opportunities for iSchools to develop programs that uniquely contribute to the broader DS education landscape, an institutional context is also important to consider and deal with when building an iSchool DS program to co-exist and complement to other DS programs on campus. This poster presents a detailed case study of such a process and approaches that will be helpful to other iSchools to navigate the process.

REFERENCES

- Cao, L. (2017). Data science: A comprehensive overview. *ACM Computing Surveys (CSUR)*, 50(3), 1-42.
- Gewin, V. (2016) Data sharing: An open mind on open data. *Nature*, 529 (7584), 117-119.
- iDSCC (2019a). *Model data science curriculum for iSchools: The iSchools Data Science Committee (iDSCC) update*. Conference presentation at the iConference 2019.
- iDSCC (2019b). *Data science education in the iSchool context*. Conference presentation at the 82nd Annual Meeting, ASIS&T 2019.
- iDSCC (2020). *iSchools Data Science Curriculum Committee (iDSCC) update*. Conference presentation at the iConference 2020.
- iDSCC (2021). *iSchools Data Science Curriculum Committee (iDSCC) update*. Conference presentation at the iConference 2021.
- Raban, D. R., & Gordon, A. (2020). The evolution of data science and big data research: A bibliometric analysis. *Scientometrics*, 122(3), 1563–1581.
- Shah, C. Anderson, T., Hagen, L., & Zhang, Y. (2021). An iSchool approach to data science: Human-centered, socially responsible, and context-driven – A position paper. *Journal of the Association for Information Science and Technology*, 72(6), 793-796.
- Virkus, S., & Garoufallou, E. (2019). Data science from a library and information science perspective. *Data Technologies and Applications*, 53(4), 422-441.
- Virkus, S., & Garoufallou, E. (2020). Data science and its relationship to library and information science: A content analysis. *Data Technologies and Applications*, 54(5), 643-663.
- Walsh, J. & Zeng, M. (2021). *Report from the iSchools Digital Humanities Curricula Committee (iDHCC)*. Conference presentation at the iConference 2021.
- Walsh, J.A., Cobb, P.J., de Fremery, W., Golub, K., Keah, H., Kim, J., ... & Wang, X. (2021). Digital humanities in the iSchool. *Journal of the Association for Information Science and Technology*, 73(2), 143-147.
- Wu, D. (2019). Research on data science curriculum [Conference presentation]. 2019 Association for Information Science & Technology Annual Meeting, Melbourne, Australia.

Is it Feasible to Use BP Neural Network for Academic Journal Evaluation: an Attempt Based on 6 Single Objective Weighting Methods

Zhao, Rongying Research Center for Chinese Science Evaluation, Wuhan University, China | zhaorongying@126.com
Zhu, Weijie Research Center for Chinese Science Evaluation, Wuhan University, China | zhuwjzzz@163.com
Chen, Wenxin Research Center for Chinese Science Evaluation, Wuhan University, China | chenwenxin@whu.edu.cn

ABSTRACT

In the new era, the analysis of academic journal evaluation methods and the comprehensive comparison of the advantages, disadvantages, and stability of different methods can help to provide some reference for academic journal evaluation. In this study, single model evaluation was carried out for 6 weighting methods without comprehensive evaluation value, and fuzzy comprehensive evaluation is performed on the results passing the nonparametric test. Based on the evaluation, BP neural network is introduced, and BP neural network evaluation model is established. The results show that the fuzzy Borda evaluation can integrate the evaluation value and evaluation order of single models, and has higher accuracy compared with single evaluation models. The prediction rate of the network model based on the gradient descent optimization algorithm can reach more than 80%, and the weights obtained from the continuous self-learning of the neural network training set can reduce the subjectivity and mutual interference between indicators.

KEYWORDS

Academic evaluation; Evaluation method; Artificial neural network; Fuzzy comprehensive evaluation

INTRODUCTION

Journal evaluation theory has a long history. the emergence of Samuel Clement Bradford's law of document dispersion (1934) and Eugene Garfield's citation law (1969) in the last century drives many scholars to focus on the development of journals, and the research on academic journal evaluation indexes and evaluation methods has been on a linear rise, including proposing new indexes, such as the e index (Zhang, 2009), the h_{top} index (Yu et al., 2018); constructing new models, such as Ivanovic (2012) proposed an extended CERIF data model; proposing new methods, such as objective weighting based on the coefficient of variation method (Amiri et al., 2011), entropy method to calculate weights(Kumar et al., 2021), index difficulty assignment(Zhou et al., 2010), TOPSIS method to calculate the ideal closeness (Krohling et al., 2015), etc. However, there are often differences in the weights obtained from the evaluation methods of academic journals, which can also have an impact on the scientific validity of the evaluation results, based on which this study explores how different methods can be better combined in the evaluation work and poses the following questions: First, what is the relationship between the weights obtained by different objective weighting methods? Second, is comprehensive evaluation more advantageous than single evaluation? Third, can the powerful mapping ability of BP neural networks be used in academic evaluation?

Method

The journals studied in this paper are 359 academic journals in the subject category of "Economics" included in JCR, starting from three perspectives: article impact, dissemination, and innovation, and the academic journal citation indexes of highly recognized are obtained from JCR, Google Scholar, and Scopus. The data values of alternative metrics of economics journals were obtained through Altmetric Explorer (Priem & Costello, 2010), totaling 155,618 items. The frontier research themes were obtained from the Research Frontiers jointly published by Clarivate and the Chinese Academy of Sciences, and all the papers published in international economics journals in the past five years were downloaded, totaling 104,776 items. The similarity between the keywords of the two parts of the frontier topic and the topic of the journal paper was calculated using the method of word overlap (calculated as Jaccard coefficient) to obtain the innovation index of journal papers. (The indicators are shown in Figure 1).

In addition, the BP neural network model is constructed from the model structure and the weight threshold. A hidden layer is designed, whose neurons in the input layer are the evaluation indexes. The number of neurons in the output layer is the number of evaluation results, and the number of neurons in the hidden layer is determined to be 5 according to the empirical formula method and the trial-and-error method. The journal data are randomly divided into two groups according to the ratio of 8:2, and the maximum number of iterations is set to 500 to construct a single hidden layer BP network of 22-5-1.

Empirical Analysis

The six tested objective weighting methods are used to assign the evaluation index weights, and the fuzzy comprehensive evaluation method is used to integrate the results. The asymptotic significance of the Kendall coordination coefficient test is 0.000, which is less than 0.05, implying that the results obtained by assigning weights using six methods such as factor analysis are correlated. The factor (Sun, 2000), principal component coefficients (Jackson, 1996), entropy value (Chen et al., 2009), coefficient of variation (Gupta et al., 1985), gray correlation coefficient (Wang, 1988), and deviation partial derivative (Tan & Deng, 1995) are calculated in turn. A vector matrix of academic journal evaluation index weights is established. Then we get the affiliation function value between the evaluation methods based on the fuzzy Borda formula (Xu et al., 2017), and get the combined evaluation score after calculating the affiliation degree, fuzzy frequency, etc. As shown in Table 1.

Journal	Factor Analysis	Principal Components	Entropy value	Variance Coefficient	Gray correlation	Discrepancy maximization	Fuzzy Borda
<i>Quarterly Journal of Economics</i>	0.5644	0.5649	0.5697	0.5713	0.5652	0.5692	0.5674
<i>Journal of Economic Perspectives</i>	0.4325	0.4327	0.4341	0.4332	0.4322	0.4329	0.4329
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
<i>Journal of Economic Literature</i>	0.2427	0.2430	0.2422	0.2417	0.2428	0.2431	0.2426
<i>Journal of Financial Economics</i>	0.2206	0.2201	0.2176	0.2177	0.2206	0.2195	0.2194

Annotation: The scores are normalized using extreme values, and the omitted journals can be found in JCR.

Table 1. Comparison of single evaluation model results and combined evaluation model results

BP neural network is a multilayer feedforward neural network trained according to the error backpropagation algorithm. Based on the initialization parameters of the neural network, the training set data is input for training. The indicator data of the training sample is input, and the set network training parameters are invoked for training the neural network model. The specific model and training process are shown in Figure 1.

Calculating the average error between the actual output and the expected output, the training cycle stops after reaching the average error target. The error value (default is 0) is 0.0001 because only one hidden layer is added and the training channels are not increased. The prediction accuracy of the network model based on the gradient descent optimization algorithm reaches more than 80%, which shows that the constructed BP neural network evaluation model is effective for the evaluation of the training set. The correlation coefficient R between the expected and actual output values is 0.99996.

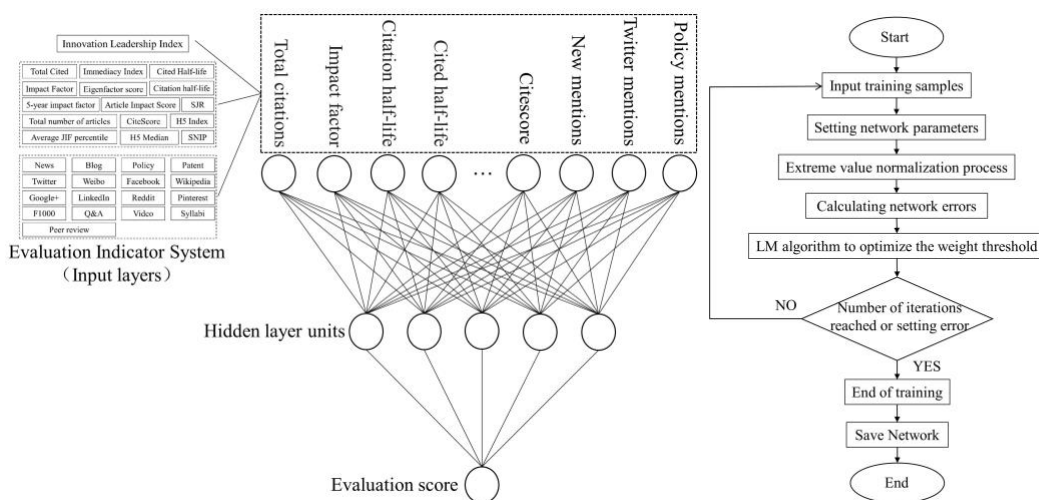


Figure 1. Neural network topology and construction process of evaluation model

CONCLUSION

The evaluation results obtained by the single objective weighting method all passed the nonparametric test, indicating that the six methods have a high degree of consistency. The fuzzy combination evaluation model established on this basis shows that the results take into account the magnitude of the evaluation value and the relative position of the evaluation order, which can make full use of the effective information, overcome the one-sidedness and limitations of the single evaluation method, and make the evaluation results more reasonable. By

introducing BP neural network with 22 evaluation indexes as the number of neurons in the input layer, the error rate of the model reaches the preset accuracy, and the actual training results have a high correlation with the expected results, indicating that the constructed evaluation model can be applied to the actual evaluation.

ACKNOWLEDGMENT

This research is funded by the National Social Science Found Major Project of China (18ZDA325).

REFERENCES

- Bradford, S. C. (1934). Sources of information on specific subjects. *Engineering*, 137: 85-86.
- Brookes, B. C. (1969). Bradford's Law and The Bibliography of Science. *Nature*, 224(5223):953-956.
- Zhang, C. T. (2009). The e-index, complementing the h-index for excess citations. *PLoS One*, 4(5), e5429.
- Yu, L. P., Wang Z G., Zhang Z J. (2018). Comprehensive amendment and research on h index: h_{top} index. *Journal of the China Society for Scientific and Technical Information*, 37(12):1188-1192.
- Ivanović, D., Surla, D., & Racković, M. (2012). Journal evaluation based on bibliometric indicators and the CERIF data model. *Computer Science and Information Systems*, 9(2), 791-811.
- Amiri, S., & Zwanzig, S. (2011). Assessing the coefficient of variations of chemical data using bootstrap method. *Journal of chemometrics*, 25(6), 295-300.
- Kumar, R., Singh, S., Bilga, P. S., Singh, J., Singh, S., Scutaru, M. L., & Pruncu, C. I. (2021). Revealing the benefits of entropy weights method for multi-objective optimization in machining operations: A critical review. *Journal of materials research and technology*, 10, 1471-1492.
- Zhou, P., Ang, B. W., & Zhou, D. Q. (2010). Weighting and aggregation in composite indicator construction: A multiplicative optimization approach. *Social indicators research*, 96(1), 169-181.
- Krohling, R. A., & Pacheco, A. G. (2015). A-TOPSIS—an approach based on TOPSIS for ranking evolutionary algorithms. *Procedia Computer Science*, 55, 308-317.
- Priem, J., & Costello, K. L. (2010). How and why scholars cite on Twitter. *Proceedings of the American Society for Information Science and Technology*, 47(1), 1-4.
- 2020 Research Frontiers. (2020). In Chinese Academy of Sciences official website. Retrieved from (<http://www.casisd.cn/zkeg/zxcg/202011/P020201114578078349185.pdf>).
- Sun, J. (2000). A note on principal component analysis for multi-dimensional data. *Statistics & probability letters*, 46(1), 69-73.
- Jackson, J. E. (1996). A user's guide to principal components. *Journal of Educational and Behavioral Statistics*, 20(1):105.
- Chen, M., Lu, D., & Zha, L. (2010). The comprehensive evaluation of China's urbanization and effects on resources and environment. *Journal of Geographical Sciences*, 20(1), 17-30.
- Gupta, R., Tripathi, R., Michalek, J., & White, T. (1985). An exact test for the mean of a normal distribution with a known coefficient of variation. *Computational Statistics & Data Analysis*, 3, 219-226.
- Wang, Y. M. (1988). Multi-Indicator decision making and ranking using the discrepancy maximization method. *China Soft Science*, (03):36-38+65.
- Tan, X. Y., & Deng, J. L. (1995). Gray correlation analysis: a new method for multi-factor statistical analysis. *Statistical Research*, 03,011.
- Xu, L. M., Lin, Z. B., Li, M. J., & Wu, S. Q. (2017). A dynamic combined evaluation method based on fuzzy borda and its application. *Chinese Journal of Management Science*, (02), 165-173.

The Influence of Open Access on the Academic Purification Effect of Retracted Publications: A Causal Inference Analysis

Zheng, Er-Te Zhejiang University, People's Republic of China | 1290625375@qq.com
Fu, Hui-Zhen Zhejiang University, People's Republic of China | fuhuizhen@zju.edu.cn

ABSTRACT

Recently, open access has gradually become an important publishing model in scientific research, while the academic purification effect of open access on flawed publications needs to be further explored. In this study, we use the methods of entropy balancing matching (EBM) and difference in difference (DID) in causal inference to examine the academic purification effect of different open access levels. The results show that in the retraction speed, limited open access papers had the slowest retraction speed, non-open access had the second and full open access had the fastest, reflecting the best performance of full open access in the detection of problematic papers. As for post-retraction citation changes, the citation of limited open access papers declined slightly slower than that of non-open access papers after retraction, while the citation of full open access papers showed an increase rather than a decrease. The next step of the research is to discover what factors may cause the above results.

KEYWORDS

Open access; Retracted publications; Bibliometrics; Academic purification; Causal inference

INTRODUCTION

Scientific publications are regarded as the cornerstone reflecting the development of the scientific community (Shah et al., 2021). Research integrity is important since the level of trust characterized science and its relationship with society (Olson & Griffiths, 1995). Misconduct and errors in publications will undermine academic development and public trust in science. The retraction is a vital way of self-purification in the scientific community, which detect the publications that lack authenticity and integrity, and reduce the negative influence of flawed researches. Recently, with the raising number of retracted publications, the issue of research integrity causes highly attention.

Meanwhile, the Open Science movement is emerging. Open Science is aimed to promote transparency and reproducibility of results, widen the diffusion of knowledge (European Commission, 2020). The post-publication content scrutiny of open access publications based on a large number of readers may accelerate the detection of misconduct and errors of flawed publications. However, with the proliferation of flawed publication practices, like the “predatory journals”, open access also has some potential threats to research integrity.

Therefore, this study aimed to explore the influence of open access (OA) on the academic purification of retracted publications. A few previous studies used basic statistical analysis and correlation analysis (Peterson, 2013; Shah et al., 2021), while this study used causal inference for a more precise estimate.

METHOD

This study collected information of retracted publications from two comprehensive databases, Web of Science (WoS) and Retraction Watch databases. To control the differences in discipline, biochemistry was selected as the object discipline of the study, for it had the largest number of retracted publications in all disciplines. 1032 biochemistry retracted publications were obtained. And the un-retracted papers published in the same issue as the retracted papers were selected as control group, totally 93496 un-retracted papers were collected. The OA levels of the papers are classified into three categories: full OA (Gold OA), limited OA (Hybrid OA, Bronze OA, Green OA) and non-OA, which used the OA classification in WoS for reference.

In this study, entropy balancing matching (EBM) and difference in difference (DID) method in causal inference were mainly applied. EBM took the increment of information entropy as distance for matching (Hainmueller, 2012). The study used this method to control for the balance of irrelevant variables between the control and treatment groups and ultimately to estimate the intervention influence of open access on the effect of academic purification.

The idea of the DID model is to compare the difference in the change between the treatment group and the control group before and after the intervention, and to use a particular output as the impact of the intervention. In this study, the treatment group was the retracted papers and the control group was the un-retracted papers, the intervention was the OA level of the papers, and the output was the post-retraction citation changes which we used five indicators to measure (see Table 2). Since the citations of un-retracted papers with different OA levels changed over time, DID was used to remove the gap in citations among un-retracted papers in order to get the exact effect of open access on the academic purification of retracted papers. A two-period panel DID model was constructed as follows:

85th Annual Meeting of the Association for Information Science & Technology | Oct. 29 – Nov. 1, 2022 | Pittsburgh, PA. Author(s) retain copyright, but ASIS&T receives an exclusive publication license.

$$Y_{it} = \alpha + \beta_1 Treat_i + \beta_2 T_i + \beta_3 Treat_i \times T_i + \gamma X_{it} + \varepsilon_{it}$$

In the above equation, Y_{it} denoted the total number of citations of the i th publication in period t . $Treat_i$ referred to the OA level of the publication. T_i was the dummy variable of retraction, indicating whether sample i was retracted publication or not, X_{it} denoted other covariates may affecting the retraction characteristics of papers.

RESULTS

By using EBM, we first controlled the variables of different OA levels of retracted publications in balance. Publish year, number of authors, journal impact factor, total citation before retraction, and citation of the year of retraction were chosen as control variables. The mean values of variables after matching were nearly the same in all groups, while the table of mean value of variables was long so we did not put it here.

After controlling for covariate balance, we calculated the mean differences of retraction time lag (i.e. the time lag between publication and retraction, which indicated the retraction speed) of different OA levels. The retraction time lag of limited OA papers was 1.36 years longer than non-OA, and that of limited OA was 2.08 years longer than full OA. We drew the survival rate curve (Figure 1) to provide a more delicate landscape of the difference of retraction speed of different OA level. It showed that in almost every year after publication, the survival rate (i.e. the proportion of flawed publications remain un-retracted) of limited OA papers was higher than non-OA papers, and that of non-OA papers was higher than full OA papers. The result indicated that full OA papers had the fastest retraction speed, which indicated it was better at timely detection and academic purification of flawed publications, while limited OA performed the worst.

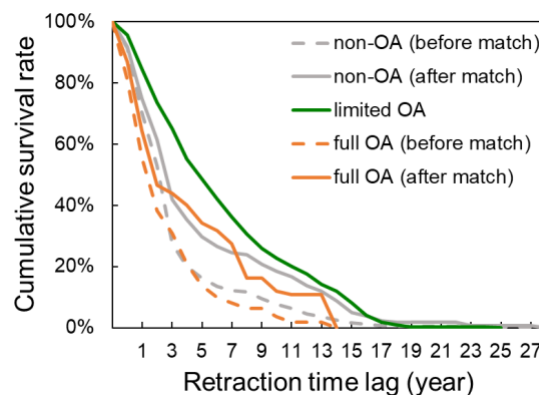


Figure 1. The survival rate curve of different OA level before and after match

We then added the retraction time lag as a covariate and applied the EBM-DID method to calculate the comparison of change in citation after retraction for different OA levels of papers (Table 1). The results showed that limited OA papers had the lowest post-retraction citation proportion, followed by non-OA papers, while full OA papers had the highest. After retraction, the number and percent change in annual citation of non-OA papers decreased the fastest, followed by limited OA, while that of full OA increased rather than decreased. The change in immediate citation (i.e. citations in the first year after retraction compared to citations in the year of retraction) of limited OA papers was lower than non-OA papers and higher than full OA papers, while the percent change in immediate citation of limited OA papers was higher than non-OA papers and full OA papers, and that of non-OA was higher than full OA.

	Post-retraction citation proportion	Change in annual citation	% change in annual citation	Change in immediate citation	% change in immediate citation
Limited OA vs non-OA	-7.429*** (-3.55)	4.073 (1.20)	5.507 (0.21)	1.515 (0.80)	-5.088 (-1.07)
Limited OA vs full OA	-54.22*** (-6.74)	-4.584*** (-4.12)	-144.0 (-1.82)	-3.242*** (-6.49)	-51.65** (-3.23)

Note: t-values in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 1. The comparison of post-retraction citation in different OA levels of retracted publications

CONCLUSION

The results showed that in terms of the retraction speed, limited open access papers had the longest retraction time lag, non-open access had the second and full open access had the shortest, reflecting the poor performance of limited open access in the detection of problematic papers, while full open access performed the best.

As for post-retraction citation changes, the citation of limited open access papers declined slightly slower than that of non-open access papers after retraction, while the citation of full open access papers showed an increase rather than a decrease. The next step of the research is to discover the reasons why limited open access papers have the longest retraction time lag and why the post-retraction citation of full open access papers shows an increased trend.

REFERENCES

- European Commission. (2020). Responsible Open Science: an ethics and integrity perspective. Retrieved April 5th, 2022 from: <https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/swafs-30-2020>.
- Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political analysis*, 20(1), 25-46.
- Olson, S., & Griffiths, P. A. (1995). *On Being a Scientist: Responsible Conduct in Research*. National Academy Press.
- Peterson, G. M. (2013). Characteristics of retracted open access biomedical literature: A bibliographic analysis. *Journal of the American Society for Information Science and Technology*, 64(12), 2428-2436.
- Shah, T. A., Gul, S., Bashir, S., Ahmad, S., Huertas, A., Oliveira, A., ... & Chakraborty, K. (2021). Influence of accessibility (open and toll-based) of scholarly publications on retractions. *Scientometrics*, 1-18.