International Journal on Digital Libraries (2024) 25:237–240
https://doi.org/10.1007/s00799-024-00407-3

Editorial to the special issue on JCDL 2022

Philipp Mayr[1] · **Annika Hinze**[2] · **Philipp Schaer**[3]

Published online: 6 June 2024

Abstract
This special issue features the selected works of authors who have presented papers at the 2022 iteration of the Joint Conference on Digital Libraries (JCDL) in Cologne, Germany. The motto of the conference was "Bridging Worlds" and was run as a fully hybrid event. Ten papers covering all aspects of Digital Libraries, namely Natural Language Processing, Information Retrieval, User Behavior, Scholarly Communication, Classification, Information Extraction are included in this issue.

Keywords Digital libraries · Natural language processing · Information retrieval · User behavior · Scholarly communication · Classification · Information extraction

1 Recap of JCDL 2022

This special issue features the selected works of authors who have presented papers at the 2022 iteration of JCDL[1]—the twenty-second meeting of the ACM/IEEE Joint Conference on Digital Libraries. All published articles are available in the JCDL 2022 proceedings [1]. The conference theme of JCDL 2022 was "Bridging Worlds" and aimed to bring together people from the Digital Library (DL) community again. As digital libraries are traditionally strong in connecting the digital and the physical world, we have picked up this heritage and established a hybrid format[2] that is as open as possible to everyone.

The JCDL 2022 call for papers attracted over 160 submissions from 35 countries on five continents. The program committee reviewed and accepted 24 research papers (24% acceptance rate, 100 submissions), 11 late-breaking research papers (34.4% acceptance rate, 32 submissions), 3 datasets (6 submissions), 7 demonstrations (15 submissions), 2 tutorials, and 6 workshops. The doctoral consortium for young Digital Library scholars accepted 5 contributions for presentation.

The program chairs of JCDL 2022 introduced some changes to the scientific review process. For the very first time, JCDL used a double-blind peer-review process. A range of new submission types were created: variable-length research papers, late-breaking, demos, and datasets. This led to interesting observations, such as that the average paper length in the research track was 8.05 pages, with accepted papers having an average of 8.87 pages and rejected papers 7.84 pages. Accepted submissions covered a wide range of topics across the field of digital libraries along with the four main topics of the call for papers: (1) users and interactions, (2) search and recommendation, (3) digital libraries in practice, and (4) content and structures.

The main conference program was organized into 11 plenary sessions on the following topics: information retrieval and access, search and recommendation, natural language processing, web archives, bibliometrics and altmetrics, information extraction, classification, and scholarly communication. Most of the presented papers (on-site and online) are available in the following Youtube channel.[3]

While scheduling the sessions, the speakers' time zones were considered as much as possible. We aimed for a good balance between online and on-site participants whenever

✉ Philipp Mayr
philipp.mayr@gesis.org

✉ Annika Hinze
hinze@waikato.ac.nz

✉ Philipp Schaer
philipp.schaer@th-koeln.de

1 Knowledge Technologies for the Social Sciences, GESIS – Leibniz Institute for the Social Sciences, Cologne, Germany

2 Computing and Mathematical Sciences, University of Waikato, Hamilton, New Zealand

3 Institute of Information Science, TH Köln – University of Applied Sciences, Cologne, Germany

[1] https://2022.jcdl.org/.

[2] https://2022.jcdl.org/hybrid/.

[3] https://www.youtube.com/playlist?list=PLn0nrSd4xjjZYAs4Wp52mzYBqeg0W1uQa.

possible. Therefore, in contrast to pure on-site conferences, all questions to the presenters were collected online. The session chairs, who were on-site, moderated questions from online and on-site participants in parallel. To strengthen the exchange between online and on-site participants, we offered a novel "Meet-The-Experts" format. Selected on-site experts have been available for dialogue with the online participants during sessions parallel to the main program.

The program chairs of JCDL 2022 invited all the best papers and best paper nominees of the conference to contribute to this IJDL special issue. In addition to this, we published an open call for papers on the IJDL website and shared this information with all authors at JCDL 2022. In total, we received 12 submissions, all by authors who had papers at JCDL 2022. All authors were briefed that a minimum of 30% additional materials have to be added to the extended follow-up papers. Guest Editors and reviewers checked this for all special issue submissions. After an in-depth peer review, we could accept 10 of these submissions. One paper was rejected and one paper was withdrawn by the authors. In the following, you will find a short description of all papers in this special issue.

2 Special issue papers

The following section provides an overview of the ten papers organized into categories. The papers appear in this issue in the same sequence. We compiled a comprehensive overview of the ten papers in Table 1 concerning the application area, used corpora, tasks, and methods in each publication. We see a heterogeneous set of publications that reflects the broad and interdisciplinary nature of the field of digital libraries.

2.1 Extended invited talk

Derived from an invited talk for JCDL 2022, Mukumbira & Winschiers-Theophilus [2] introduce this special issue with their work on "Implications of an ecospatial indigenous perspective on digital information organization and access." They explore the complexities of digitizing indigenous knowledge, highlighting the challenge posed by epistemological differences and the limited engagement of indigenous communities. They draw on experiences from projects in Namibia to shed light on how indigenous ecospatial worldviews influence the design of digital information systems for accessing indigenous knowledge. Embracing emerging technologies like augmented and virtual reality, the paper envisions more inclusive digital libraries that honor indigenous perspectives, aiming for richer, more immersive representations of their knowledge and culture.

2.2 Natural language processing

In their paper on "Cross-lingual extreme summarization of scholarly documents" Takeshita et al. [3] explore advanced neural models to create cross-lingual summaries of scholarly texts, addressing the overwhelming volume of scientific publications. They introduce the X-SCITLDR dataset, facilitating multilingual summarization by training models to process English papers and generate German, Italian, Chinese, and Japanese summaries. The study benchmarks various models, including a two-stage pipeline and direct cross-lingual approaches, while investigating methods like intermediate-stage training and knowledge distillation to enhance efficiency and reduce computational complexity during summarization inference.

The following paper on NLP is "Creating and validating a scholarly knowledge graph using natural language processing and microtask crowdsourcing" [4]. In this work, Oelen et al. introduce TinyGenius, a method validating NLP-extracted scholarly knowledge by leveraging crowdsourced microtasks to populate a paper-centric knowledge graph. They address the limitations of NLP accuracy by employing five distinct NLP methods and extend their methodology by detailing NLP tasks and the data model. A corresponding user evaluation reveals promising prospects for microtask-based statement validation despite varying participant agreement across different microtasks, indicating its potential for improving the quality of extracted scholarly knowledge.

2.3 Information retrieval, access and user behavior

Roy at al. examine retrievability within an integrated search system, analyzing the influence of retrieval systems on accessing datasets, publications, and variables in a digital library. In their paper "Retrievability in an integrated retrieval system: An extended study" [5] they employ traditional metrics like the Lorenz curve and Gini coefficient. The study visualizes and discovers a significant popularity bias, revealing that specific datasets are more frequently retrieved than others within the same category. Interestingly, retrievability scores among variables and publications exhibit more balanced distributions, indicating a greater diversity in document retrievability for datasets compared to the other document types.

The second paper in this section is on "Graduate student search strategies within academic digital libraries" [6]. Hoeber & Storie investigate graduate students' information seeking strategies in academic digital library searches, aiming to determine whether their chosen strategies align with the search scenario's complexity. Utilizing a survey method, participants recalled recent search sessions, revealing that only the lookup search strategy matched the search scenario's complexity consistently. Factors like discipline of study and

Table 1 Overview on this special issue concerning the application area, used corpora, tasks, and methods in each publication

	Application area	Corpus	Tasks	Methods
Mukumbira and Winschiers-Theophilus [2]	Information Organization	Digitalisation projects	Access of indigenous knowledge	Case studies
Takeshita et al. [3]	Document summarization	Scient. abstracts	Cross-lingual summaries	TLDR methods
Oelen et al. [4]	Scholarly Knowledge Graph	ArXiv	Crowdsourcing	NLP methods
Roy et al. [5]	Dataset Search	Social Science DL	Measuring retrievability	Lorenz curves; Gini coefficient
Hoeber and Storie [6]	Information Seeking	–	Search strategy utilization	Online survey
Hu et al. [7]	Book Reviews	Goodreads	Understanding user-generated book reviews	Case studies
Fernandes and Vaz-de-Melo [8]	Peer Review	OpenReview data	Review score prediction	Classification models
Heck et al. [9]	Systematic literature review	Educational DL	Relevance decision	Coverage analysis
Brack et al. [10]	Scholarly Document Processing	Scient. abstracts and full texts	Sentence classification	Deep Learning
Kroll et al. [11]	Information extraction workflow	Wikipedia, DL	Entity linking	Information extraction

DL digital library

search training impacted strategy choice, emphasizing the need to train students in effectively aligning strategies with task complexity and to develop interfaces that facilitate this alignment in the search process.

2.4 Scholarly communication

The first work on scholarly communication is from Hu et al. on "Complexities of leveraging user-generated book reviews for scholarly research: Transiency, power dynamics, and cultural dependency" [7]. This paper highlights the scarcity of infrastructure for studying book reception in digital libraries and underscores the overlooked complexities and ethical concerns surrounding the use of user-generated book reviews in scholarly research. Through three case studies investigating temporal shifts in rankings, sponsored reviews, and reader ratings across languages, it reveals complexities in reviews, emphasizing their transient nature, power dynamics, and cultural dependencies. The study calls for a critical and nuanced approach to utilizing user-generated book reviews in scholarly research, acknowledging challenges in data curation and proposing mitigation strategies for researchers.

The last paper in this section is by Fernandes and Vaz-de-Melo [8]. In their paper "'Enhancing the Examination of Obstacles in an Automated Peer Review System" they investigate the challenges hindering the development of an efficient automated system for predicting review scores and final paper decisions in the peer review process. They evaluate existing state-of-the-art models for these tasks, revealing that models face difficulty when classifying specific instances, resulting in a significant performance drop, around 23.31%, particularly with instances challenging to classify. The research underscores that current models, while promising, still encounter cases that markedly impact their accuracy, highlighting the distance from achieving a fully automated system for scoring and deciding the acceptance or rejection of academic papers.

2.5 Classification and information extraction

The first paper in this last section has the title "Coverage and similarity of bibliographic databases to find most relevant literature for systematic reviews in education". This work by Heck et al. [9] examines the critical aspect of selecting bibliographic databases for systematic literature reviews in educational research. It highlights the lack of evidence on database relevance in this interdisciplinary field. Through empirical data analysis across three recent reviews, the authors evaluate seven databases, finding that discipline-specific databases surpass international multidisciplinary sources in identifying relevant literature. The study advocates for a combination of discipline-specific international and national databases as the most effective approach, offering practical implications for researchers conducting systematic literature searches in education and related disciplines.

The next paper has to be especially mentioned as it won the Best Student Paper Award at JCDL 2022. In their paper "Sequential Sentence Classification in Research Papers using Cross-Domain Multi-Task Learning" Brack et al. [10] explore the potential of transfer learning for sentence classification. While they focus on different scientific domains, they aim to improve the structuring of research papers and academic search engines. To achieve this goal, they introduce a novel deep learning architecture and multi-task learning approach for cross-domain sentence classification, addressing challenges posed by differing text structures in full papers and abstracts and introducing methods to identify semantically related classes across different annotation schemes. Models trained across various scientific domains benefit from each other using the proposed multi-task learning architecture. Experimental results demonstrate the superiority of this approach over state-of-the-art methods for full-paper datasets and competitive performance for abstract datasets.

In "A detailed library perspective on nearly unsupervised information extraction workflows in digital libraries," Kroll et al. [11] explore the efficacy of unsupervised extraction methods in digital libraries. They address the challenges of costly and reliable extraction workflows and extend previous research by analyzing unsupervised extraction workflows in domains like encyclopedias, Pharmacy, and Political Sciences, delving into the quality and applicability of these methods. Additionally, the study scrutinizes extractions in more depth, verifies findings with a second extraction method, explores canonicalizing methods, and considers the handling of non-English texts, ultimately discussing best practices for unsupervised extraction workflows while outlining their opportunities and limitations.

All ten articles received an extensive peer review and editorial treatment for this special issue and represent the broad spectrum of work and the heterogeneity in our community. We hope you find these papers as innovative and insightful as we did. Enjoy the read.

Acknowledgements We want to thank the organizing committee of JCDL 2022 for their diligence, innovation, and flexibility in planning and executing this conference. We thank our sponsors and supporters: ACM, IEEE, SIGIR, Deutsche Forschungsgemeinschaft (DFG), GESIS—Leibniz Institute for the Social Sciences, University of Hildesheim, National Institute of Informatics, TH Köln—University of Applied Sciences, University of Waikato, and CNI. Special thanks go to Zeljko Carevic, the acting General Chair of JCDL 2022, who orchestrated the conference organization pretty well in the light of a difficult mid-to-end-pandemic context.

Funding Open Access funding enabled and organized by Projekt DEAL. Philipp Mayr received additional funding by the European Union under the Horizon Europe Grant OMINO—Overcoming Multilevel INformation Overload (Grant Number 101086321, http://ominoproject.eu [12]).

References

1. Aizawa, A., Mandl, T., Carevic, Z., Hinze, A., Mayr, P., Schaer, P., (eds.): JCDL '22: Proceedings of the 22nd ACM/IEEE Joint Conference on Digital Libraries (ACM, Cologne Germany, 2022). https://doi.org/10.1145/3529372
2. Mukumbira, S., Winschiers-Theophilus, H.: Implications of an ecospatial indigenous perspective on digital information organization and access. Int. J. Digit. Libr. (2023). https://doi.org/10.1007/s00799-023-00353-6
3. Takeshita, S., Green, T., Friedrich, N., Eckert, K., Ponzetto, S.P.: Cross-lingual extreme summarization of scholarly documents. Int. J. Digit. Libr. (2023). https://doi.org/10.1007/s00799-023-00373-2
4. Oelen, A., Stocker, M., Auer, S.: Creating and validating a scholarly knowledge graph using natural language processing and microtask crowdsourcing. Int. J. Digit. Libr. (2023). https://doi.org/10.1007/s00799-023-00360-7
5. Roy, D., Carevic, Z., Mayr, P.: Retrievability in an integrated retrieval system: an extended study. Int. J. Digit. Libr. (2023). https://doi.org/10.1007/s00799-023-00363-4
6. Hoeber, O., Storie, D.: Graduate student search strategies within academic digital libraries. Int. J. Digit. Libr. (2023). https://doi.org/10.1007/s00799-023-00378-x
7. Hu, Y., LeBlanc, Z., Diesner, J., Underwood, T., Layne-Worthey, G., Downie, J.S.: Complexities of leveraging user-generated book reviews for scholarly research: transiency, power dynamics, and cultural dependency. Int. J. Digit. Libr. (2023). https://doi.org/10.1007/s00799-023-00376-z
8. Fernandes, G.L., Vaz-de Melo, P.O.S.: Enhancing the examination of obstacles in an automated peer review system. Int. J. Digit. Libr. (2023). https://doi.org/10.1007/s00799-023-00382-1
9. Heck, T., Keller, C., Rittberger, M.: Coverage and similarity of bibliographic databases to find most relevant literature for systematic reviews in education. Int. J. Digit. Libr. (2023). https://doi.org/10.1007/s00799-023-00364-3
10. Brack, A., Entrup, E., Stamatakis, M., Buschermöhle, P., Hoppe, A., Ewerth, R.: Sequential sentence classification in research papers using cross-domain multi-task learning. Int. J. Digit. Libr. (2024). https://doi.org/10.1007/s00799-023-00392-z
11. Kroll, H., Pirklbauer, J., Plötzky, F., Balke, W.T.: A detailed library perspective on nearly unsupervised information extraction workflows in digital libraries. Int. J. Digit. Libr. (2023). https://doi.org/10.1007/s00799-023-00368-z
12. Hołyst JA, Mayr P, Thelwall M, Frommholz I, Havlin S, Sela A, Kenett YN, Helic D, Rehar A, Maček SR, Kazienko P, Kajdanowicz T, Biecek P, Szymanski BK, Sienkiewicz J (2024) Protect our environment from information overload. Nat Hum Behav 8:402–403. https://doi.org/10.1038/s41562-024-01833-8