Introduction to the focused issue on the 2017 ACM/IEEE-CS Joint Conference on Digital Libraries JCDL 2017

Catherine C. Marshall¹ · Ian Milligan² · Adam Jatowt³

Published online: 18 May 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

This special issue of International Journal on Digital Libraries (IJDL) brings together a selection of notable papers nominated for the two best paper awards—the Vannevar Bush Best Paper Award and the Best Student Paper Award—at the 2017 ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2017). JCDL is an annual international interdisciplinary conference that brings together researchers, developers, and practitioners in the field of digital libraries to share their progress on associated technical, practical, and social issues.

The seventeenth Joint Conference on Digital Libraries was held in Toronto, Ontario, Canada from June 19–23, 2017. Because the field of digital libraries has undergone dramatic changes as collections grow in size, diversity, and online accessibility, this year's theme highlights working at scale: developing novel analytical tools and methodologies for making sense of large, heterogeneous collections, for deriving diverse kinds of knowledge, and for linking across different collections and research disciplines. As our field matures, we are striving to improve outreach efforts, engage diverse communities, and provide a broad range of users with effective and flexible access to materials, along with the ability to make new observations and discoveries. A search on the hashtag #jcdl2017 will retrieve real-time commentary and photos from the conference.

As best papers, the articles in this special issue of IJDL have been through a particularly rigorous review process. In their initial form as conference papers, the submissions were peer-reviewed by three experts in the authors' specialty, discussed at a two-day in-person program committee meeting,

and given a summary of this initial feedback by a metareviewer (a senior member of the community) to help the authors make substantive changes to improve their papers. To transform these already fine examples of work in our field into journal articles, we invited the nominees for the Vannevar Bush Award and the Best Student Paper Award to resubmit their work with at least 30% new content, including progress the authors have made on their research topic since the conference. These enhanced submissions were reviewed and went through an extensive revision process to address the additional feedback IJDL reviewers gave them.

The seven papers that emerged from this lengthy editorial process are well-worth reading, even if you read the originals in the conference proceedings or saw them presented at JCDL 2017 in Toronto. The papers published in this special issue of IJDL include:

Our first paper, “On the Effectiveness of the Scientific Peer-review System: A Case Study of the Journal of High Energy Physics,” by Sandipan Sikdar, Paras Tehria, Matteo Marsili, Niloy Ganguly, and Animesh Mukherjee, explores the pressing question of what value the peer-review system adds to scholarly publishing. The authors do so by using a corpus of *Journal of High Energy Physics* articles and reviews (over 29,000 papers with 70,000 reviews) between 1997 and 2015. By looking at a series of features from this large corpus, the authors are able to provide several findings around the impact of peer review on both the papers but also the long-term citations of a paper (for example, the fewer review rounds a paper is subjected to, the more citations they tend to receive). Ultimately, the paper supports the peer-review system through quantitative analysis and provides data which will support editors and authors alike as new models and approaches to peer review are considered.

The second paper, “Towards Meaningful Notions of Similarity in NLP Embedding Models” by Ábel Elekes, Adrian Englhardt, Martin Schäler, and Klemens Böhm, explores the notion of term similarity in digital document collections based on a series of experiments on 1 Billion word dataset. The motivation for this work stems from the fact

✉ Ian Milligan
i2millig@uwaterloo.ca

Catherine C. Marshall
nomiddleinitial@gmail.com

Adam Jatowt
adam@dl.kuis.kyoto-u.ac.jp

¹ Texas A&M University, College Station, Texas, USA

² University of Waterloo, Waterloo, Canada

³ Kyoto University, Kyoto, Japan

that the notion of similarity and the meaning of similarity values remains ambiguous despite the recent success of word embedding models like Word2Vec or GloVe. The authors systematically analyze similarity distributions of word embeddings based on the different embedding model algorithms and parameters, and then they introduce an approach to decide the similarity values and thresholds that are meaningful for a given embedding model. They find that similarity values highly depend on the algorithms and parameters, and so, the same value can represent different grades of similarity in different models.

In “Bias-aware News Analysis using Matrix-based News Aggregation,” Felix Hamborg, Norman Meuschke, and Bela Gipp approach the challenging problem of uncovering media bias in news articles. As current news aggregators cannot effectively detect and warn readers about media bias, the authors propose an approach in the form of matrix-based aggregation to help users gain a broad and diverse news understanding by presenting various perspectives on the same news topic. The proposal organizes and visualizes news content in the cells of a matrix over two dimensions according to its key attributes in order to reveal diverse and controversial information. They also present NewsBird which is an extensible news aggregator for international news and is equipped with the proposed matrix-based news aggregation. User studies indicate high effectiveness of this news aggregator in helping users to stay informed about news biases and about inherent differences in news reporting. The authors release the code of NewsBird as an open-source system.

The paper, “Building and Querying Semantic Layers for Web Archives,” tackles one of the major problems facing the scholarly use of web archives—the difficulty of extracting meaningful information from these extremely large repositories of information. Accordingly, Pavlos Falafios, Helge Holzmann, Vaibhav Kasturia, and Wolfgang Nejdl outline a process for building a semantic profiles (or “layers”) that can describe metadata and other concepts (such as named entities) which can subsequently be published as linked data. The paper presents three semantic layers: one built on top of the New York Times Annotated Corpus, one for the Internet Archive’s Occupy Movement 2011/2012 Archive-It collection, and one for a social media archive of over 1.3 million Tweets. Their evaluation convincingly demonstrates that these semantic layers can help find information where keyword-based models have failed.

The paper, “Fusion Architectures for Automatic Subject Indexing under Concept Drift: Analysis and Empirical Results on Short Texts” by Martin Toepfer and Christin Seifert, is devoted to the study of concept drift in automatic subject indexing of digital libraries. The authors first discuss and exemplify the existence of the concept drift in the output

of machine learning-based methods for automatic assignment of descriptor terms to documents in digital libraries. They analyze next architectures for automatic indexing to elucidate strengths and weaknesses and justify the need for fusion approaches. Fusion approaches are then found to outperform the non-fusion ones based on experimental analysis on titles and author keywords in the domain of economics. The authors state that the comparison findings should help researchers and practitioners in digital libraries to choose appropriate methods for automatic subject indexing. The authors also propose descriptor-invariant fusion of associative and lexical indexing approaches.

The paper, “A Framework for Modelling and Visualizing the US Constitutional Convention of 1787,” by Nicholas Cole, Alfie Abdul-Rahman, and Grace Mallon, describes Quill, a platform for presenting time-based records that reflect the negotiations and discussions that go into producing a large-scale document (or set of documents). The research builds on a long history of computer support for argumentation and decision-making, especially as these processes are presented in a pedagogical setting. Non-technical domain experts are given tools for structuring events and the associated changes to texts, and students are provided with ways of visualizing and understanding how important documents emerge over time according to a formal model of debate. In the example described in the paper, debate is conducted as a parliamentary process, and the resulting document is the US Constitution. The article also discusses the authors’ experiences with the platform’s use from various perspectives (e.g., that of an editor encoding debate records and that of student interrogating the results in an educational setting). This paper is an important milestone in argumentation research, and an excellent example of how Digital Humanities research can be presented to a Digital Libraries audience.

Our final paper, “Towards Comprehensive Event Collections,” by Federico Nanni, Simone Paolo Ponzetto, and Laura Dietz, explores the creation of event collections from much larger web archival collections. Researchers wanting to study specific events are confronted by the vastness of web archives. Even if they are able to get all of the data, it can elude their processing and storage abilities, and manual curation is time and resource intensive. To get around this very real constraint, the authors propose a method for automatically building event-centric sub-collections which not only finds documents directly relevant to the event but also documents that explore the consequences of causes of it as well. Their process draws on both the documents themselves and a knowledge base to find similarities between the two and find relevant documents.

The seven articles that have gone through the editorial process for this special issue of IJDL represent a broad spec-

trum of work in our community. If readers of this special issue would like to recommend papers from this special issue to their peers, or otherwise comment on them, we recommend the use of the hashtag #IJDL SI17. We hope you find these papers as innovative and insightful as we did.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.