

Guest editors' introduction to the special issue on the digital libraries conference 2014

Martin Klein¹ · Andreas Rauber²

Published online: 12 August 2015
© Springer-Verlag Berlin Heidelberg 2015

For the first time in the long-standing history of digital library conferences, the Joint Conference on Digital Libraries (JCDL) and the Conference on Theory and Practice of Digital Libraries (TPDL) were joined into one—the 2014 conference on Digital Libraries (DL). DL 2014¹ was held at City University in London from September 9th through September 11th. Both conferences are premiere international venues on the broadly interpreted domain of digital libraries and hence DL 2014 attracted researchers, educators, industry leaders, and students working in realm of digital libraries and associated organizational, practical, social, as well as technical issues. Its program covered a broad spectrum of topical areas such as user research, system architectures, collection policies, and specialist domains such as digital humanities, preservation, and scholarship.

This International Journal on Digital Libraries (IJDL) special issue brings together extended and enhanced versions of outstanding DL 2014 publications. The authors of the six papers included here were invited to submit an extended version of their conference paper which was reviewed by three to four reviewers each. The reviewers were asked to judge not only the quality of the work but also the degree to which

it was substantially enhanced with new material compared to the conference version.

The papers in this special issue address a wide range of topics: an analysis of preservation strategies, a discussion of the role of digital libraries in knowledge infrastructures, author disambiguation approaches, web archive coverage, large-scale processing environments, and an investigation of the quality of archived web pages.

The paper “When Should I Make Preservation Copies of Myself?” by Charles Cartledge and Michael L. Nelson investigates how different replication policies (ranging from least aggressive to most aggressive) affect the level of preservation achieved by autonomic processes used by web objects. The work shows that a moderately aggressive replication policy makes the best use of distributed host resources by not causing spikes in CPU resources nor spikes in network activity while meeting preservation goals. The conference version of this work won the *Vannevar Bush Best Paper Award* at DL 2014.

The contribution titled “Knowledge Infrastructures in Science: Data, Diversity, and Digital Libraries” by Christine Borgman, Peter T. Darch, Ashley E. Sands, Irene V. Pasquetto, Milena S. Golshan, Jillian C. Wallis, and Sharon Traweck discusses today’s role of digital libraries in knowledge infrastructures for science by presenting evidence from the authors’ long-term studies of various “big science” and “little science” research sites. The authors found, for example, that big sites invested in digital libraries for data management as part of their initial research design, whereas smaller sites made smaller investments at later stages.

“On the Combination of Domain-Specific Heuristics for Author Name Disambiguation—The Nearest Cluster Method” by Marcos Andre Goncalves, Alan Filipe Santana, Alberto Laender, and Anderson Ferreira introduces a set of carefully designed heuristics and similarity functions to help

¹ <http://www.jcdl.org/archived-conf-sites/dl2014/www.city.ac.uk/digital-libraries-2014.html>

✉ Martin Klein
martinklein@library.ucla.edu

Andreas Rauber
rauber@ifs.tuwien.ac.at

¹ Research Library, University of California Los Angeles, Los Angeles, CA 90095, USA

² Institute of Software Technology and Interactive Systems, Vienna University of Technology, Favoritenstrae 9-11/188, 1040 Vienna, Austria

solve the problem of author name disambiguation. The work shows that this method can outperform state-of-the-art supervised methods in terms of effectiveness while being orders of magnitude faster in comparison.

The paper “Lost but Not Forgotten: Finding Pages on the Unarchived Web” by Hugo C. Huurdeman, Jaap Kamps, Thaer Samar, Arjen P. de Vries, Anat Ben-David, and Richard A. Rogers argues that web archives will always be incomplete due to a variety of restrictions such as crawling depth, crawling frequency, and selection policies. The work proposes an approach to uncover unarchived web pages and websites and to reconstruct different types of descriptions for these resources based on their links and anchor texts to ultimately provide a significant increase in the coverage of web archives.

“Bridging the Gap Between Real World Repositories and Scalable Preservation Environments” is the title of the paper by Bolette Ammitzbøll Jurik, Asger Askov Blekinge, Rune Bruun Ferneke-Nielsen, and Per Møldrup-Dalum. It describes a solution to integrate large-scale processing environments, such as Hadoop, with traditional repository systems, such as Fedora Commons 3. The introduced system is based on software that was developed as part of the SCAPE project.

The contribution by Justin F. Brunelle, Mat Kelly, Hany SalahEldeen, Michele C. Weigle, and Michael L. Nelson titled “Not All Mementos Are Created Equal: Measuring The Impact Of Missing Resources” looks at missing portions of archived web pages and their importance values. Not all embedded resources are equally important as their impact on the web page may vary. This work proposes a method to measure the relative value of embedded resources and assign a damage rating to archived pages as a way to evaluate archival success. The conference version of this contribution won the *Best Student Paper Award* at DL 2014.

We would like to thank all authors of this special issue for their valuable contributions and express our gratitude to all reviewers for their invaluable feedback. We believe this issue is of utmost interest to existing as well as potentially new IJDL readers as it covers a broad spectrum of topics and describes exciting new developments in the area of digital libraries and beyond.