

WADL 2016: Third International Workshop on Web Archiving and Digital Libraries

Edward A. Fox
Virginia Tech
Dept. of Computer Science
Blacksburg, VA 24061 USA
+1-540-231-5113
fox@vt.edu

Zhiwu Xie
Virginia Tech
University Libraries
Blacksburg, VA 24061 USA
+1-540-231-4453
zhiwuxie@vt.edu

Martin Klein
U. California Los Angeles
Research Library
Los Angeles, CA 90095 USA
+1-310-206-9781
martinklein@library.ucla.edu

ABSTRACT

This workshop will explore integration of Web archiving and digital libraries, so the complete life cycle involved is covered: creation/authoring, uploading/publishing in the Web (2.0), (focused) crawling, indexing, exploration (searching, browsing), archiving (of events), etc. It will include particular coverage of current topics of interest, like: big data, mobile web archiving, and systems (e.g., Memento, SiteStory, Hadoop processing).

CCS Concepts

•Information systems → Digital libraries and archives; World Wide Web;

Keywords

Web archiving; Internet Archive

1. INTRODUCTION

Our understanding of the past will, to a large extent, depend on our success with Web archiving. WADL 2016 (see <http://fox.cs.vt.edu/wadl2016.html>) will bring together international leaders from industry, government, and academia, who are tackling this important challenge. They will explore the integration of Web archiving and digital libraries, over the complete life cycle: creation/authoring, uploading, publishing in the Web, crawling/collecting, compressing, formatting, storing, preserving, analyzing, indexing, supporting access, etc. Table 1 gives a partial list of some of the topics discussed.

The objectives of this workshop are to:

- continue to build the community of people integrating Web archiving with digital libraries;
- help attendees learn about useful methods, systems, and software in this area;

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

JCDL '16 June 19-23, 2016, Newark, NJ, USA

© 2016 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-4229-2/16/06.

DOI: <http://dx.doi.org/10.1145/2910896.2926735>

Table 1: Partial List of Topics

Archiving (events)	Big data	Classification
Community bldg	Crawling	Curation, Q/C
DBs / collections	Discovery	Extraction
Filling gaps	Globalization	Linking
Metadata	Mobile	Network science
Preservation	RDF	Social sciences
Standards, protocols	Systems, tools	Tweets

- help chart future research and practice in this area, so more and higher quality Web archiving occurs;
- produce an archival publication that will help advance technology and practice; and
- promote synergistic efforts including collaborative projects and proposals.

2. RELATED WORK

The most recent related workshop, WADL 2015, was held in conjunction with JCDL 2015. It led to a special issue of the IEEE TCDD Bulletin with 13 papers [4]. An earlier workshop, WIRE, focused on research leading to or making use of archives that preserve Internet content [8]. The first workshop on Web Archiving and Digital Libraries, WADL 2013, led to a summary [2] after a group responded to the call for meeting [3] as part of the JCDL 2013 workshop program. An earlier similar workshop at a prior JCDL took place in Ottawa in 2011 [6], partly as a result of the emergence of a cooperative to explore Web archiving [5]. Broader in scope but related are the annual General Assembly meetings of the International Internet Preservation Consortium (IIPC) [7].

3. LOGISTICS

3.1 Audience and Attendees

There is a growing community interested in this topic. Given that the prior WIRE workshop on a similar theme [8] had a significant number of participants from Rutgers University and that Columbia University played a lead role in a closely related Mellon funded initiative [1], we expect to have 15-30 attendees, including a solid representation of students. We will advertise, solicit submissions, have them reviewed by the program committee, and then organize an

interesting program. Selections are guided by an international program committee of 13 people, in addition to the three co-chairs. We anticipate the program to include aspects from multiple disciplines such as Computer Science, Library and Information Science, Web Science, Social Sciences, etc.

3.2 Format and Duration

A full-day workshop is split across two days, with invited speakers, selected papers, posters, demonstrations, and panels. A WebEx teleconference during the workshop can enable those unable to attend to be involved.

As in 2015, we expect to have a special issue of IEEE TCDD Bulletin; that led to a call for a special issue of IJDL, so if the timing is right, WADL 2016 contributions may go there too.

4. BIOGRAPHICAL AND CONTACT INFORMATION

4.1 Fox

Edward Fox holds a Ph.D. and M.S. in Computer Science from Cornell, and a B.S. from M.I.T. Since 1983 he has been at Virginia Tech, where he serves as Professor. He directs VT's Digital Library Research Laboratory and the Networked Digital Library of Theses and Dissertations. He was a member of the Board of CRA (the Computer Research Association). He was chair of the IEEE Technical Committee on Digital Libraries, and earlier was chair of ACM SIGIR. He was chair of the steering committee for JCDL, and is on the international advisory committee for ICADL. He has been (co-)Principal Investigator on 122 research grants/contracts. He taught over 81 tutorials and has given 66 keynote/distinguished/ international invited talks. He has (co-)authored 18 books, 118 journal/magazine articles, 49 books chapters, 207 refereed conference/workshop papers, 71 posters, and about 160 other publications/reports, plus over 320 additional talks. Fox is editor for IR and DL for ACM Books. He was Co-Editor-in-Chief for ACM JERIC, and is on the boards of IJDL, JEMH, JIIS, J. UCS, Multimedia Tools and Applications, and PeerJ CS.

Contact information

Dept. of Computer Science, 114 McBryde Hall, M/C 0106
Virginia Tech, Blacksburg, VA 24061 USA
Tel: +1-540-231-5113 [direct], -6931[dept.]
Mobile: +1-540-553-1856, Fax: +1-540-231-6075 [CS]
Email: fox@vt.edu, Website: <http://fox.cs.vt.edu>

4.2 Xie

Zhiwu Xie is an associate professor at Virginia Tech Libraries and leads its technology development team. He leads the development of transactional archiving based UWS, the Goodwin Hall Living Lab data management system, IMLS ETDplus, and VTechData, among others. He is deeply involved in Fedora/Hydra, APTrust, PREMIS, ResourceSync, and AltMetrics Data Quality. His research has received support from Mellon, IBM, Amazon, and NSF XSEDE. He was a co-chair of WADL 2015. His website is at: <http://scholar.lib.vt.edu/staff/zxie/>.

4.3 Klein

Martin Klein holds a Ph.D. in Computer Science from Old Dominion University. He currently is a scientist in the Research Library at the University of California Los Angeles. He was program chair of DL 2014, poster chair of JCDL 2015, and is currently co-conference chair of iPres 2016. In addition, he is a board member of the Web Archiving Collaboration at Columbia University and guest editor of the International Journal on Digital Libraries as well as the Bulletin of IEEE Technical Committee on Digital Libraries. He also was co-chair of WADL 2015. Martin Klein is the lead editor of the ANSI/NISO Specification Z39.99 and has published numerous journal/magazine articles and conference/workshop papers. More information can be found at: <http://www.library.ucla.edu/staff/martin-klein>

5. ACKNOWLEDGMENTS

Our thanks go to NSF for support through IIS 0916733 and 1319578, to Columbia and the Mellon Foundation for supporting "Archiving Transactions Toward Uninterruptible Web Service," and to QNRF for support through NPRP 4-029-1-007. The opinions expressed in this document are solely our own.

6. REFERENCES

- [1] Columbia U. Libraries. Web Resources Collection Program, 2015. https://library.columbia.edu/bts/web_resources_collection.html.
- [2] E. Fox and M. Farag. Report on the Workshop on Web Archiving and Digital Libraries (WADL 2013), WADL Workshop Report. *ACM SIGIR Forum*, 47(2):128–133, 2013. <http://sigir.org/files/forum/2013D/p128.pdf>.
- [3] E. A. Fox. Web Archiving and Digital Libraries (WADL 2013). Virginia Tech CTRnet announcement, 2013. <http://www.ctrnet.net/sites/default/files/JCDL2013WorkshopWebArchiving20130603.pdf>.
- [4] E. A. Fox, Z. Xie, and M. Klein. Introduction to the Web Archiving and Digital Libraries 2015 Workshop issue: Web Archiving and Digital Libraries 2015 (WADL 2015) Overview. *Bulletin of IEEE Technical Committee on Digital Libraries*, 11(2), October 2015. <http://www.ieee-tcdl.org/Bulletin/v11n2/papers/intro.pdf>.
- [5] H. Garcia-Molina, F. McCown, M. Nelson, and A. Paepcke. Web Archive Cooperative making web archives useful today, supported by the NSF (1009916), Stanford University, 2011. <http://infolab.stanford.edu/wac/>.
- [6] H. Garcia-Molina, F. McCown, M. Nelson, and A. Paepcke. Web Archive Globalization Workshop. In conjunction with JCDL 2011, Ottawa, Canada, June 16-17, 2011. <http://cs.harding.edu/wag2011/>.
- [7] IIPC. International Internet Preservation Consortium homepage, 2015. <http://netpreserve.org/>.
- [8] M. Weber, D. Lazer, K. Carpenter-Negulescu, and A. Kosterich. Working with Internet Archives for Research (WIRE 2014 Workshop, Cambridge, MA, June 17-18), 2014. <http://wp.comminfo.rutgers.edu/nsfia/>.