

Web Archiving and Digital Libraries 2015 (WADL 2015) Overview

Edward A. Fox
Digital Library Research Laboratory
Virginia Tech
Blacksburg, VA 24061 USA
fox@vt.edu

Zhiwu Xie
University Libraries
Virginia Tech
Blacksburg, VA 24061 USA
zhiwuxie@vt.edu

Martin Klein
UCLA Library
Los Angeles, CA 90095 USA
martinklein@library.ucla.edu

ABSTRACT

Our understanding of the past will, to a large extent, depend on our success with Web archiving. WADL 2015 brought together international leaders from industry, government, and academia, who are tackling this important challenge. This special issue includes summaries of twelve presentations on 24 June 2015. It is hoped that these works will stimulate other digital library (DL) and related investigations and efforts that will help ensure that broader and better archiving takes place, that more tools (beyond the excellent ones mentioned in these papers, that generally are freely available) become available, and that wider support develops to expand Web archiving.

1. WORKSHOP LOGISTICS

In conjunction with the 2015 ACM/IEEE Joint Conference on Digital Libraries, JCDL 2015, held in Knoxville, TN, USA (<http://www.jcdl2015.org>), WADL 2015 was held at the end of that conference. On June 24, attendees and remote participants spent a day exploring Web archiving. There were five long presentations and seven short presentations; the latter connected with a poster and demo session. To round out the day, there was an introduction, statements by participants, breaks, and a closing discussion. The three authors of this work served as co-chairs.

2. OBJECTIVES

In addition to producing our website (<http://www.dlib.vt.edu/WADL2015/>) and this archival publication, and planning for a related journal special issue (see call at <http://fox.cs.vt.edu/DL/Web-Archiving-Focused-Ed-Fox-final-2015.pdf>), this workshop strove:

- to continue to build the community of people integrating web archiving & DLs;
- to help attendees learn about useful methods, systems, and software in this area;
- to help chart future research and improved practice in this area;
- to promote synergistic efforts including collaborative projects and proposals; and
- to produce an archival publication (this issue) that will help advance technology and practice.

3. DESCRIPTION

WADL 2015 explored 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. In particular, submissions were solicited on topics of interest such as:

Archiving (events)	Big data
Classification	Community building
Crawling (focused)	Curation, quality control
Databases / collections	Discovery
Extraction & analysis	Filling gaps
Globalization, languages	Linking archives
Metadata	Mobile devices
Network science	Preservation
Resource description	Social sciences
Standards, protocols	Systems, tools
Tweet connections	

4. PAPERS FOLLOWING

This special issue includes, in order, the following works:

1. Robert S. Comer and Andrea J. Copeland. Methods for Capture of Social Media Content for Preservation in Memory Organizations
2. Mohamed M. G. Farag and Edward A. Fox. Building and archiving event web collections: A focused crawler approach
3. Zhiwu Xie, Prashant Chandrasekar and Edward Fox. A UWS Case for 200-Style Memento Negotiations
4. Luis Meneses, Sampath Jayarathna, Richard Furuta and Frank Shipman. Grading Degradation in an Institutionally Managed Repository
5. Sawood Alam, Michael L. Nelson, Herbert Van de Sompel, Lyudmila L. Balakireva, Harihar Shankar and David S. H. Rosenthal. Profiling Web Archives For Efficient Memento Query Routing
6. Smiljana Antonijevic and Ellysa Cahoy. Connecting the IR with the User: Personal Archiving via Zotero
7. Gerhard Gossen, Elena Demidova and Thomas Risse. The iCrawl System for Focused and Integrated Web Archive Crawling
8. Ian Milligan. Finding Community in the Ruins of GeoCities: Distantly Reading a Web Archive
9. Zhiwu Xie, Herbert Van de Sompel, Jinyang Liu, Johann van Reenen and Ramiro Jordan. Web Archiving Inconsistency: A Research Agenda

10. Tomas Foltyn and Martin Lhotak. The Czech Digital Library - Fedora Commons based solution for aggregation, reuse, dissemination and archiving of digital documents
11. Tarek Kanan, Sagnik Ray Chowdhury, C. Lee Giles, Prashant Chandrasekar, and Edward A. Fox. Digital Library and Archiving for Qatar
12. Todd Suomela. Analytics for Monitoring Usage and Users of Archive-It Collections

5. RELATED EVENTS

There have been a number of closely related events preceding this workshop, such as:

- Working with Internet Archives for Research (WIRE 2014) NSF workshop, 17-18 June 2014, Cambridge, MA – see <http://wp.comminfo.rutgers.edu/nsfia/>
- Web Archiving and Digital Libraries (WADL'13), 25-26 July, at JCDL 2013, see <http://www.ctrnet.net/sites/default/files/JCDL2013WorkshopWebArchiving20130603.pdf> and report in SIGIR Forum <http://sigir.org/files/forum/2013D/p128.pdf>
- Web Archive Globalization Workshop, WAG 2011 – see <http://cs.harding.edu/wag2011/>, with 4 organizers plus 5 presenters and about 20 participants, held in Ottawa after JCDL 2011 (June 16-17)
- Ongoing work by attendees in this area, growth in collaborative activity involving the Internet Archive, and specific community building successes like the Web Archive Cooperative – see <http://infolab.stanford.edu/wac/>
- Annual meetings of the International Internet Preservation Consortium (IIPC), partner meetings of the Internet Archive

(Archive-It), and ten workshops held with ECDL/TPDL: International Web Archiving Workshop (IWAW), 2001-2010

6. ACKNOWLEDGEMENTS

For their help with planning and reviewing, we thank the organizing committee:

- Jefferson Bailey, Internet Archive, jefferson@archive.org
- Prashant Chandrasekar, Virginia Tech, peecee@vt.edu
- Mohamed Magdy Farig, Virginia Tech, mmagdy@vt.edu
- Vinay Goel, Internet Archive, vinay@archive.org
- Frank McCown, Harding University, fmccown@harding.edu
- Michael Nelson, Old Dominion Univ., mln@cs.odu.edu
- Andreas Rauber, TU Vienna, rauber@ifs.tuwien.ac.at
- Matthew Weber, Rutgers, matthew.weber@rutgers.edu

This special issue arose in part because of work at numerous institutions supported by multiple sponsors. The first author of this paper was supported by the US National Science Foundation under Grant No. IIS-1319578 and by NPRP grant # 4-029-1-007 from the Qatar National Research Fund (a member of Qatar Foundation). The first and the second authors were also supported by the Incentive Awards of the Mellon Web Archiving Grant. The statements made herein are solely the responsibility of the authors.