

Web Archiving and Digital Libraries (WADL)

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

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). It will include particular coverage of current topics of interest: big data, mobile web archiving, and systems (e.g., Memento, SiteStory, Uninterruptible Web Service).

Categories and Subject Descriptors

H.3.5 [Information Storage and Retrieval]: Online Information Services – *Web-based services*. H.3.6 [Information Storage and Retrieval]: Library Automation – *Large text archives*. H.3.7 [Information Storage and Retrieval]: Digital Libraries – Collection, Standards, Systems issues.

General Terms

Management, Standardization.

Keywords

Web archiving; Internet Archive.

1. INTRODUCTION

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;
- help chart future research and practice in this area, so more and higher quality Web archiving occurs; and
- promote synergistic efforts including collaborative projects and proposals.

2. RELATED WORK

The most recent related workshop, WIRE, focused on research leading to or making use of archives that preserve Internet content [1]. The first workshop on Web Archiving and Digital Libraries, WADL, 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 [4], partly as a result of the emergence of a cooperative to explore Web archiving [5]. Broader in scope are the international activities for Internet preservation [6], including a 2015 meeting.

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. For all other uses, contact the authors. Copyright is held by the authors. *JCDL'15*, June 21-25, 2015, Knoxville, TN, USA. ACM 978-1-4503-3594-2/15/06. <http://dx.doi.org/10.1145/2756406.2756934>

3. TOPICS

This workshop will cover all topics of interest, including but not limited to:

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

4. ACKNOWLEDGMENTS

Our thanks go to NSF for support through IIS 1319578 and to QNRF for support through NPRP 4-029-1-007. The opinions expressed in this document are solely our own. Thanks also go to the JCDL organizing committee, and to the WADL organizing committee: Jefferson Bailey, Prashant Chandrasekar, Mohamed Magdy Farig, Vinay Goel, and Martin Weber.

5. REFERENCES

- [1] Weber, M., Lazer, D., Carpenter-Negulescu, K., and Kosterich, A. 2014. *Working with Internet Archives for Research* (WIRE 2014 Workshop, Cambridge, MA, June 17-18, 2014). <http://wp.comminfo.rutgers.edu/nsfia/>
- [2] Fox, E.A. and Farag, M.M. 2013. *Report on the Workshop on Web Archiving and Digital Libraries* (WADL 2013), WADL Workshop Report, ACM SIGIR Forum, 47(2): 128-133. <http://sigir.org/files/forum/2013D/p128.pdf>
- [3] Fox, E.A. 2013. *Web Archiving and Digital Libraries* (WADL 2013). Virginia Tech CTRnet announcement, <http://www.ctrnet.net/sites/default/files/JCDL2013WorkshopWebArchiving20130603.pdf>
- [4] Garcia-Molina, H., McCown, F., Nelson, M., and Paepcke, A. 2011. Web Archive Globalization Workshop. In conjunction with JCDL 2011, Ottawa, Canada, June 16-17. <http://cs.harding.edu/wag2011/>
- [5] Garcia-Molina, H., McCown, F., Nelson, M., and Paepcke, A. 2011. Web Archive Cooperative Making Web Archives Useful Today, Supported by the NSF (1009916), Stanford University, <http://infolab.stanford.edu/wac/>
- [6] IIPC. International Internet Preservation Consortium. 2015. Homepage <http://netpreserve.org/>