

# Digital Preservation

Lunch lecture

Intro: Prof. Dr. Michael A. Herzog

18.09.2015

13:30 - 14:30

House 7, R 0.06

Campus Herrenkrug

## Web Archiving: A Brief Introduction

The Web is present in a broad range of fields of our lives including education, research, governance, news, entertainment, communication, publishing, shopping, social interactions, and many more. We collectively create and consume the Web more than ever before and are truly living in the Web era. However, the Web has a volatile nature. Very often web services disappear and content gets modified, deleted, or moved to other places. We need to preserve our trail on the Web in order for the historians of the future generations to know about us accurately. Apart from the long term preservation of the history there are other advantages and use cases of the archived Web such as personal satisfaction, storytelling, legal evidences, and many more.

In this talk we will briefly discuss the following:

- Archiving and Web archiving
- Purpose and importance of the Web archiving
- Scope of the web archiving (personal or institutional)
- Issues and challenges
- Tools and techniques
- Memento: Time Travel for the Web
- Archive X-Ray: High level summary of the archives' holdings
- Research opportunities in Web archiving
- Our Web Science and Digital Libraries Research Group

### about Sawood Alam

Sawood received his B.Tech. degree in Computer Science from Jamia Millia Islamia, New Delhi, India in 2008 and his M.Sc. in Computer Science from Old Dominion University, Norfolk, Virginia in 2013. His Master's Thesis title was "HTTP Mailbox - Asynchronous Restful Communication". Sawood is currently working on his Ph.D. thesis titled, "Archive X-Ray - Web Archive Profiling for Efficient Memento Aggregation". Apart from his academic research in Web Science and Web Archiving field, he is also interested in solving technical challenges of Urdu and other Right-to-Left complex script languages.



A full list of his publications, presentations, research, and development efforts can be found at <http://www.cs.odu.edu/~salam/>.

## Using Web Archives to Enrich the Live Web Experience Through Storytelling

The web has become an integral part of our lives, shaping how we get news, shop, and communicate. When critical events occur, social media and news websites cover the stories as they break and continually revise them as the story evolves. Unfortunately, much of the content around these stories is vulnerable and prone to loss. Thus, web archives have become a significant repository of our recent history and cultural heritage. Meanwhile, "storytelling" is becoming a popular technique in social media for selecting representative tweets, videos, web pages, etc. and arranging them in chronological order to support a particular narrative or "story".

I will explain how to (semi-)automatically integrate archived collections with social media to create stories that summarize the holdings in these archived collections through identifying a set of resources that best represent the topic of the collection, then place this set of resources in an interface that users are already familiar with, such as Storify (a storytelling service that allows users to create their narratives). We can sample from the collections and provide one or more summaries or abstracts of the entire collection.

### about Yasmin AlNoamany

Yasmin is a researcher and Ph.D. Candidate at Old Dominion University in USA majoring in Computer Science with a focus on Information retrieval and Web Preservation. She received her Bachelor's and Master's Degrees in Computer Science from Mansoura University in Egypt. Yasmin had two internships at the Internet Archive, one of the largest and oldest public web archives in the world, in San Francisco as a software engineer in 2014/2015. She has multiple awards on the academic field, such as "the best teaching assistant award" from the college of science and "the best student paper award" in the International Conference on Theory and Practice of Digital Libraries (TPDL) in Sept. 2013. Her research interests include digital preservation, web mining, information retrieval, and digital humanities.



More information: <http://www.cs.odu.edu/~yasmin/>