ALEXANDRIA

Temporal Retrieval, Exploration and Analytics in Web Archives

Wolfgang Nejdl

L3S Research Center
Hannover, Germany
Web Science @ L3S

Computer Science and interdisciplinary research on all aspects of the Web

- Internet: Communication and Networks
- Information: Accessing information and knowledge on and through the Web
- Community: Supporting communities and groups on the Web, for research, education, production and entertainment
- Society: Requirements (technological, social, legal) for the Web

Selected projects

- Real-time data processing for finance predictions
- Cross-media analysis and interpretation
- Retrieval, Exploration and Analytics for Web Archives (ERC)
- ForgetIT: Concise Preservation via Managed Forgetting
- Privacy, Property and Internet Governance

CUBRIK: Searching by computers and humans
Are we loosing the past of the web?

@RSSEGYPTcom: #Egypt
@ALMasryAlYoun_E: Armed forces attacked sleeping #Copts, say Coptic leaders http://tinyurl.com/44fshv4
@Alenyout: Intel, Egypt-based SySDesign to boost 4G LTE efforts. http://tinyurl.com/4brni4n
@UnclefunkLA: comes rough, tough like an elephant tusk. Ya head rush, fly like Egyptian misk...
@Alenyout: Daily News Egypt: Egyptian Emergency Services received compromising arms from Sudan. http://tinyurl.com/4twk4jx
@RSSEGYPTcom: #Egypt #Jan25
@RSSEGYPTcom: #Egypt #Jan25 Attack on Copts
Gun running from Sudan
Spam
EgyTweets
Are we losing the past of the web?

Library of Congress
- In April 2010 LoC and Twitter signed an agreement to archive all tweets since 2006
- January 2013: It is clear that technology to allow for scholarship access to large datasets is lagging behind technology for creating and distributing such data. The Library is pursuing partnerships to allow some limited access capability in reading rooms.

German National Library
- Based on a law of June 22, 2006, the GNL should collect, enrich, catalog, archive Web publications

Internet Archive
- Archiving the Web (10 Petabyte) since 1996
- Access possible through the URL

Relevant Projects @ L3S
- Web Archiving: LiWA, ARCOMEM, ForgetIT
- Web Search: PHAROS, CUBRIK
- Web and Stream Analytics: EUMSSI, Qualimaster
- ERC Advanced Grant: ALEXANDRIA (2014 – 2018, 2.5 Mill. Euro)

Cooperations
- German National Library, British Library, Internet Archive, Rutgers University, et al
How can we provide better access to Web archives?

ALEXANDRIA Vision and Research Questions