Body:

Motivation

Easy access to historical Web information becomes more and more important, as significant parts of our cultural heritage are produced and consumed online. Traditional institutions keeping our cultural heritage need to be complemented with facilities for preservation and public access of online cultural assets. The ALEXANDRIA project aims to develop models, tools and techniques necessary to archive and index relevant parts of the Web, and to retrieve and explore this information in a meaningful way. While the easy accessibility to the current Web is a good baseline, optimal access to Web archives requires new models and algorithms for retrieval, exploration, and analytics which go far beyond what is needed to access the current state of the Web. This includes taking into account the unique temporal dimension of Web archives, structured semantic information already available on the Web, as well as social media and network information.

Challenges & Highlight

Within ALEXANDRIA, we will significantly advance semantic and time-based indexing for Web archives using human-compiled knowledge available on the Web, to efficiently index, retrieve and explore information about entities and events from the past. In doing so, we will focus on the concurrent evolution of this knowledge and the Web content to be indexed, and take into account diversity and incompleteness of this knowledge. We will further investigate mixed crowd- and machine-based Web analytics to support long-running and collaborative retrieval and analysis processes on Web archives. Usage of implicit human feedback will be essential to provide better indexing through insights during the analysis process and to better focus harvesting of content.

Potential applications & future issues

The ALEXANDRIA Testbed will provide an important context for research, exploration and evaluation of the concepts, methods and algorithms developed in this project, and will provide both relevant collections and algorithms that enable further research on and practical application of our research results to existing archives like the Internet Archive, the Internet Memory Foundation and Web archives maintained by European national libraries.

Related projects at L3S

- Timetool - Large-Scale Temporal Search in MapReduce
- iCrawl - The Integrated Focused Crawling ToolBox
- WikiTimes - Bringing Order to Events in Wikipedia
- WikipEvent - Challenges and Opportunities for Temporal Information Retrieval and Evolution in Wikipedia
- Hadoop@L3S

The Alexandria project is funded by the European Research Council (ERC Advanced Grant No.: 339233).

Project abstract:

The ALEXANDRIA project (ERC Nr. 339233) aims to develop models, tools and techniques necessary to explore and analyze Web archives in a meaningful way. ALEXANDRIA will significantly advance semantic and time-based indexing for Web archives using human-compiled knowledge available on the Web, to efficiently index, retrieve and explore information about entities and events from the past. The ALEXANDRIA Testbed will provide relevant collections and algorithms that enable further research on and practical application of research results to existing archives.

Project duration:
01.03.2014 - 29.02.2019

Publications:
tag/alexandria

Bibsonomy key:
200204d02c371a6dc9d1789194ac9444

Bibsonomy tags:
alexandria,sys:relevantfor:l3s

Bibsonomy url:
Bibsonomy user: mirsha
Bibsonomy show project publications: 1
Bibsonomy use tabs to list publications: 1
Members: anand fetahu gottschalk holzmann joao marenzi nejdl singh souza
Project manager: Dr.-Ing. Thomas Risse, Dr. Avishek Anand
Project research areas:
Web Search
Project type: EU FP7 ERC Advanced Grant
URL: http://www.alexandria-project.eu
Research Area: Intelligent Access to Information
Status of the Project: