The project is named after the Library of Alexandria, in Egypt, which was one of the largest and most significant libraries of the ancient world. The following paragraphs give a short summary of the ALEXANDRIA project.

Significant parts of our cultural heritage are produced on the Web, yet only insufficient opportunities exist for accessing and exploring the past of the Web. 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.

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 longrunning 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.

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.

Kontakt: Prof. Wolfgang Nejdl, nejdl@l3s.de

ERC Advanced Grant “Alexandria” at L3S