Enhancing Temporal Summarization using Web Archives

Master thesis

The L3S Research Center, under supervision of Prof. Dr. Wolfgang Nejdl offers a Master thesis as part of the ERC Advanced Grant Project ALEXANDRIA. The main objective of the temporal summarization task is to investigate the summarization of events and entities over time, where entities (people, location, or organization) and events will be explicitly modeled as first class citizens in analytic processes. This requires an entity- and event-aware model and algorithms taking into account coverage, diversity, importance and overlap/redundancy of information, time-ware contextualization for long-term interpretability, as well as adequately dealing with evolution and with information over time.

The goal of this Master/Diploma thesis is to develop (1) innovative models and methods to unlock the insights of big, fast, open data, through systematic analyses, and data mining techniques, and (2) evaluate the studying models and methods, which include creating test collections and metrics for understanding relevance, user intent, and system performance. In particular, the work will be related to various topics: NLP processing (e.g., entity/event extraction), opinion mining, document summarization, and machine learning.

The tasks for the student will be:

I. Process web documents, e.g., web archives, social media, or a Wikipedia articles.

II. Implement novel models and algorithms for supporting dynamic information processes including (1) model entity/event aspects in documents, (2) handle the temporal dynamics of documents, and (3) provide a robust, meaningful summarization.

III. Conduct usability study to evaluate the proposed methods.

You should have:

- Strong motivation and willingness to learn
- Programming skills in Java, databases, and web applications

Are you interested or have questions? Please contact us!

Dr. Nattiya Kanhabua <kanhabua@L3S.de>
Forschungszentrum L3S, Appelstr. 9a, 30167 Hannover