

## Supporting Evolving Search Tasks in Digital Environments via Fuzzy String Matching

Sanna Kumpulainen and Heikki Keskustalo

University of Tampere, Finland

### ABSTRACT

We present a newly started research project funded by the Academy of Finland (2017-2021). We aim at networking and collaboration with history researchers working with digital resources in Finland. This presentation describes the aims and the implementation of the project. Last, we will sketch what kind of collaborations we are looking for.

In addition to historians themselves, the digital methods in history research have drawn the attention of researchers in information studies. Our fields of study include information retrieval and information behavior. This is an international research effort aiming at supporting digital historians research work tasks. The context of the research is historian's research work and we will particularly focus on (a) research processes and needs descriptions of the historians' work tasks, and (b) the study the information retrieval in historical document collections, and how to support these with fuzzy matching methods.

Traditionally, digital access methods in historical research have been studied from the viewpoint of the document collections and data. Instead of this data-centric view, we adopt *information interaction centered view* towards accessing digital collection. Information interaction is understood as behavioral and cognitive activities related to task planning, searching and selecting information items, working with information items, and synthesizing and reporting [1].

Information interaction encompasses searching in the larger task context. This means that the actions are framed by a motivating task that triggers searching. We aim at better understanding the rationale behind actions and by this enable combining task-based approach into tool development and evaluation.

### Case: Historians' work in digital environment

In using the task-based approach in information interaction, we will first start by empirical research in task-based information interaction. We will collect accounts of information needs and human activities framed by the research work. We will also focus on the information tools and how they support searching the historical documents. The descriptions of the use cases and human information interaction processes will inform the following constructive part.

The constructive part involves creating novel support for the interactions in the digital information environments. By creating such tools, the project will contribute to the digital history field. There is a need for better understanding regarding the historians' digital information environment (ecology). So far, the TBII framework has not been used in the digital historians' work contexts and we expect to create some novel insights. The information environment encompasses heterogeneous data and information resources. The current systems and search engines may have limited view on how the data and information may be used. The researchers may bump into barriers during their information interaction processes because the current systems may constrain the research work instead of supporting the human creativity.

## **Implementation**

The research project has several tasks. Firstly, we will start with the empirical task. This includes the close study of work tasks (including work processes and work goals) and information interactions using existing tools. We will use interviews and observational methods to create understanding on the current state. We need to understand how heterogeneous data sources are used during interactions in particular digital information environments. By doing this we will understand what are the needs and research goals of the researchers using the digital collections, and what kinds of hurdles or barriers they encounter during their research processes. Empirical part is followed by a constructive task that includes designing methods and information tools informed by the empirical phase. Since the project has just started, we are working on the first tasks. We will begin with a research pilot to learn about the information environment, history research work, and the suitability of the selected data collection methods.

We are seeking for collaboration regarding both the pilot phase, and the whole project. We hope that this presentation helps in building networks with the following aims in mind:

- 1) finding participants for data collection (digital historians): our goal is to understand and learn about the information interaction processes and particular state of the art technologies actually used in history research. Especially, we are looking for researchers working with digitized historical newspaper collections (e.g. the newspaper collection of National Library of Finland).
- 2) finding subject experts as potential collaborators to provide valuable information about experiential views on the topic. Especially, comments of the type “we need this / we do not need this” are highly valuable.
- 3) learning about related research projects, topics, and synergies across them.

The outcomes of the research project include several aspects. Firstly, we learn from the historians' work and interactions in digital information environments. Secondly, we will create suitable support tools tailored to help with the particular tasks and interactions. Finally, the novel support may enhance or change the work processes and goals themselves.

## **Acknowledgements**

This research is supported by the Academy of Finland under grant 310278.

## **References**

- [1] K. Järvelin, P. Vakkari, P. Arvola, F. Baskaya, A. Järvelin, J. Kekäläinen, H. Keskustalo, S. Kumpulainen, M. Saastamoinen, R. Savolainen, E. Sormunen (2015) Task-Based Information Interaction Evaluation: The Viewpoint of Program Theory. *ACM Transactions of Information Systems*, Vol. 33, No. 1, Article 3.