Visual Analysis and Exploration of Ancient Texts with a User-driven Concept Search

Muhammad Faisal Cheema
PhD Dissertation

Abstract

In the last decades, the amount of digital data has grown rapidly. This also impacted the humanities field where the humanists and computer scientists collaborate to digitize historical texts for preservation purpose. Due to this effort, many historical texts that were accessible to few scholars became widely available in form of digital texts.

The focus shifted from printed to digital texts. This led to the foundations of digital humanities. As a consequence, it became necessary to develop tools that can support the searching, analysis and exploration of large historical text collections. Even though text exploration tasks have developed at a rapid pace less work has been done to support the humanities scholars in their traditional workflows and methods. A humanities scholar working with a historical text corpus usually applies keyword based search queries in order to retrieve related text passages within a historical document. These methods are insufficient for the scholar when she is looking for term that exist in various forms and phrases (called as concepts in this dissertation) in historical texts.

The dissertation focuses on combining methods from human computer interaction, visualization and visual analytics to improve the traditional workflow of humanities scholars searching historical texts. The good features of keyword based search engines are taken into account. An abstract process model for concept search is introduced that follows the classic text retrieval model embeds the visual analytics process. The dissertation introduces a novel search environment called **User-driven Concept Search** that empowers the humanities scholars to extend the traditional keyword based search using a concept based methodology where they can model their ideas of concepts and apply search on these models. The search results shown in distant and close reading views enable the scholars to perform textual analysis tasks seamlessly on the obtained search results. The implementation of visual analytics ensures that the scholar has full control over the search process at every step and she is able to iteratively improve the search results by improving the concept model. The user-driven concept search has been extended with an enhanced close reading tool, **AnnotateVis**. This tool mimics the original workflow of the scholars and supports the annotation of text fragments with various media types. For a profound text interpretation, a number of text analysis methods coupled with various distant reading visualizations are provided as annotations.