



Exploration and Contextualization through Interaction and Concepts

M. Bron

Summary

Today's technology enables the continuous production, recording, and storage of all types of digital information. The abundance of information that is therefore available provides researchers with the opportunity to ask new questions but also requires new research methods and appropriate tools.

Fields of study differ in the way they have adapted to deal with this flood of data. The natural sciences such as earth sciences and computational biology have readily adopted computationally intensive methods, as they study signals recorded by radars, sensors, or produced by simulations. This type of data lends itself well for analyses through data mining and visualization techniques. In contrast, the traditional objects of study in the humanities have always been analogue records such as books, letters, and photographs. These objects are studied using analytical, critical, and interpretative approaches instead of computational ones. As the introduction of new technology and information sources is changing the way humanities researchers work and the questions they seek to answer, a new challenge arises for the development of tools and algorithms that support new practices as well as traditional ones using new types of information.

Particular challenges for humanities researchers raised by the abundance of available material are to gain insight in which materials to consider for a study through exploration and once chosen to obtain a holistic view of the research topic through contextualization. This thesis investigates two dimensions along which tools to support humanities researchers in dealing with the flood of information may be improved: by providing richer means of interaction with information systems and developing algorithms that allow discovery of information through relations between concepts. One of the findings, along the interaction dimension, is that for a particular group of humanities researchers the ability to make comparisons between alternative search results leads to further exploration of the material available. Additional results, along the concepts dimension, include algorithms that support identifying relations between concepts based on structured and unstructured data. Tools incorporating these algorithms allow humanities researchers to identify additional concepts related to the ones already identified as relevant to their research topic. The results in this thesis show how both richer interactions and more effective related concept finding algorithms may be used to improve tools to support the research practices of humanities researchers. The insights from the work in this thesis may be used to inform the design and evaluation of future tools to continually support new needs and developments in the humanities.