Introduction to the 5th HistoInformatics Workshop

Melvin Wevers[1], Mohammed Hasanuzzaman[2], Gaël Dias[3], Marten Düring[4], and Adam Jatowt[5]

[1] DHLab, KNAW Humanities Cluster, Amsterdam, the Netherlands
`melvin.wevers@dh.huc.knaw.nl`
[2] ADAPT Centre, Cork Institute of Technology, Dublin, Ireland
`mohammed.hasanuzzaman@adaptcentre.ie`
[3] CNRS GREYC, University of Caen Normandy, France
`gael.dias@unicaen.fr`
[4] Luxembourg Centre for Contemporary and Digital History, Esch-sur-Alzette, Luxembourg
`marten.during@uni.lu`
[5] Graduate School of Informatics, Kyoto University, Kyoto, Japan
`adam@dl.kuis.kyoto-u.ac.jp`

Abstract

With an increasing number digitized historical archives, we also see a growing number of historians that rely on computational methods to examine such archives. This includes methods for retrieving textual and visual information from archives, studying the history of concepts, but also reflections on research practices and the construction of digitized archives. The HistoInformatics workshop series is focused on the challenges and opportunities of data-driven historical research and brings together scientists and scholars at the forefront of this emerging field, at the interface between historical disciplines on the one hand and informatics on the other. The 5th HistoInformatics workshop took place on September 12, 2019, and was a full-day workshop co-located with the 23rd International Conference on Theory and Practice of Digital Libraries (TPDL 2019) in Oslo.

1 Introduction

HistoInformatics 2019[1]—the 5th International Workshop on Computational History—was held on September 12, 2019 in conjunction with the 23rd International Conference on Theory and Practice of Digital Libraries (TPDL 2019) in Oslo. The HistoInformatics workshop series brings together researchers in the historical disciplines with those working in computer science and associated disciplines. It is designed to discuss best practices as well as possible future collaborations.

Traditionally, historical research is based on the hermeneutic investigation of preserved records and artefacts as a means to improve our understanding of the past. In recent years, the large-scale digitization of historical documents has enabled researchers to query and analyze archives using text mining, linked data, or image analysis. Historians, like other humanists, show keen interests in computational approaches to the study and processing of digitized sources. Computer scientists have developed innovative methods to query and analyze digitized materials. In bringing together historians and computer scientists, we open up a space to validate such computational methods by applying them to historical data sets to answer questions pertinent to historical research. We call this approach: HistoInformatics.

---

[1] https://sites.google.com/site/histoinformatics2019workshop/

2　Program

HistoInformatics 2019 was a full-day workshop. The program chairs accepted six papers. The accepted papers are briefly described in what follows.

The paper *Reflections on Design Principles for a Historical Archive in a Low Resource Environment* discussed strategies for building digital archives in environments characterized by a lack of technical skills, facilitates, and/or funding. The authors of the paper *LOCALE: A Rule-based Location Named-entity Recognition Method for Latin Text* introduce a method for location extraction from Latin texts. The paper *Clustering Ideological Terms in Historical Newspaper Data with Diachronic Word Embeddings* studies the emergence of *isms* in nineteenth-century Finnish newspapers as a proxy for political language formation in Finland. The authors of *Trawling for Terrorists: A Big Data Analysis of Conceptual Meanings and Contexts in Swedish Newspapers, 1780–1926* also turn to digitized newspaper archives. This paper traces the concept of terrorism over almost 150 years. The paper *Character Segmentation in Collector's Seal Images: An Attempt on Retrieval Based on Ancient Character Typeface* turns to computer vision to construct a recognition system based for document retrieval of ancient Asian books. Finally, the authors of *Plague Dot Text: Text Mining and Annotation of Outbreak Reports of the Third Plague Pandemic (1894-1952)* describe the challenges related to an interdisciplinary research project that studies reports from the third plague pandemic (1894-1952).

The workshop also included a keynote by Mikko Tolonen, Assistant Professor in Digital Humanities at the University of Helsinki, Finland, titled: *Integrated Computational Approach in Eighteenth-Century Intellectual History*. The talk focused on strategy of the Helsinki Computational Group for combining computational methods with eighteenth-century intellectual history.

3　Acknowledgements

We would like to thank the program committee members for their effort in reviewing and recommending papers. We also thank the organizers of the TPDL 2019 conference for hosting and supporting our workshop, and also the anonymous reviewers who evaluated our proposal, for their constructive criticism. Finally, we would like to express our gratitude to our keynote speaker Mikko Tolonen.