
PROgressive VISual DEcision-Making in Digital Humanities: Towards Uncertainty Annotation in TEI

Michał Kozak^{*1}, Jennifer Edmond², and Roberto Theron³

¹Poznan Supercomputing and Networking Center – Poland

²Trinity College Dublin – Ireland

³Universidad de Salamanca – Spain

Abstract

*** Prepared for submission of both a paper and a poster for the Marketplace ***

Authors: Alejandro Benito (abenito@usal.es), Michelle Doran (doranm1@tcd.ie), Jennifer Edmond (jennifer.edmond@tcd.ie), Michał Kozak (mkozak@man.poznan.pl), Cezary Mazurek (mazurek@man.poznan.pl), Alejandro Rodríguez (jancho@usal.es), Roberto Theron (theron@usal.es).

In recent years, Digital Humanities (DH) as a research field has experienced a great transformation that has permitted the undertaking of academic projects of great scope and impact, while allowing their immediate exposure to society. At the same time, a number of new and powerful information and communication technologies (ICTs) have made possible the exploitation of a wealth of data (either digitised or born digital) that have changed enormously the practice in DH. From the creation to the consumption of digital resources, there are new stakeholders, contexts, and tasks to consider. The volume of digital resources produced (or digitised), stored, explored, and analysed in any DH project is immense. Therefore, traditional humanities tools for managing uncertainty have to be either substituted or aided with ancillary tools in the form of interactive visualisations or novel user interfaces in this environment. Furthermore, during the whole lifecycle of any DH project - from the data preparation to the actual analysis or exploration phase - many decisions have to be made in order to yield the desired results that depend on managing the uncertainty pertaining to both the datasets and the models behind them and which may in turn introduce their own uncertainty into the research process.

This presentation will introduce the ongoing work of the interdisciplinary PROgressive VISual DEcision-Making in Digital Humanities (PROVIDEDH) project, a three-year project funded within the CHIST-ERA call 2016 for the topic "Visual Analytics for Decision Making under Uncertainty – VADMU." The project aims to give DH scholars a space to explore and assess the completeness, evolution, and interconnectedness of digital research objects, the degree of uncertainty that the models applied to the data incorporate, tolerate or introduce, and to share their perspectives and insights with the project's broad range of stakeholders.

The project itself can be broken down into the following research objectives:

To establish a taxonomy of sources of uncertainty that may appear during the lifecycle of a DH project.

*Speaker

To develop a set of metrics that convey the degree of uncertainty that research objects, data sets, and collections introduce as well as the different computational models applied to them.

To propose a framework that makes use of the uncertainty metrics, so any given representation of the data can be assessed according to its degree of uncertainty.

To propose a Progressive Visual Analytics solution that ensures that users are able to trace changes in data and its inherent uncertainty as well as in the way it is perceived.

To develop a web-based multimodal collaborative platform for the progressive visual analysis of different DH collections, both for scholars and citizen scientists.

To trigger the formation of a "community of practice" that humanists can build on to reinforce each other's efforts to achieve metrics that are both practical and of high quality.

The objective of the presentation is to share information and resources relating to the use of the PROVIDEDH collaborative platform. In addition to enabling progressive visual analyses of large research datasets, users of the platform will be able to perform a close reading of annotated TEI P5 files with a focus on uncertainty. Encoders of text often find it useful to indicate that some aspects of the text are problematic or uncertain, and to indicate who is responsible for various aspects of the markup. A standard way to annotate DH research datasets is through the use of TEI tags. Although the TEI provides various methods to indicate that some aspects of the encoded text are problematic or uncertain, it is not a common practice and uncertainty remains a challenge that any researcher using the dataset has to face. The PROVIDEDH project intends to bring uncertainty to the surface by associating the existing TEI elements with the developed taxonomy of sources of uncertainty and by implementing an environment for progressively visualising uncertainty in humanities data.

The talk will focus on visualisations and the user-friendly interface of annotating electronic text implemented in the close reading module of our collaborative platform. Solutions for many possible cases will be presented, like annotating the same or intersecting texts by many people with different profiles, or annotating already tagged entities by many people with different perspectives of those entities. Furthermore, potential directions of tracking and processing these annotations will be outlined.

As an ancillary objective of the presentation, it is hoped that it will encourage the attendees of the annual event to join the project's interdisciplinary and international "community of practice". The presentation will be accompanied by a poster for the Marketplace.

Keywords: Progressive Visual Analytics, Uncertainty, Taxonomy of Uncertainty, TEI, Humanities Research Data, Annotation.