Using Collaborative Annotation to Explore and Analyse Uncertainties in Digital Humanities Datasets

Michelle Doran[*1], Jennifer Edmond[1], Alejandro Benito-Santos[*2], and Roberto Theron[*2]

[1]Trinity College Dublin – Ireland
[2]Universidad de Salamanca – Spain

Abstract

The immediate intention underlying John Unsworth's explication of scholarly primitives was to suggest some recursive functions of humanities scholarship that might be usefully embodied in tool-building in humanities computing.

The central challenge of uncertainty representation has increasingly come to the fore in Digital Humanities research. The new practices of knowledge production characterised and analysed by Digital Humanities imply uncertainty by their very nature, they are intended to alter our understanding and they almost always occur at the boundaries between different disciplines. From the creation to the consumption of digital resources, there are new stakeholders, contexts and tasks to consider. The amount of digital resources produced (or digitized), stored, explored, and analysed in any Digital Humanities project is immense and uncertainty is a key issue that could hinder substantially in the use and re-use of digital cultural heritage resources. Therefore, traditional humanities methods for managing uncertainty have to be either substituted or aided with ancillary tools and novel user interfaces in this environment. As this exposure to uncertainty can be inherently complicated, especially for users with low computational literacy, the use of appropriate visualization and HCI techniques is key to enhance the comprehensibility of the presented results and to enable a well-informed decision-making process.

This workshop presents the ongoing work of the interdisciplinary PROgressive VIsual DEcision-Making in Digital Humanities project (PROVIDEDH), a three-year project funded within the CHIST-ERA call 2016 for the topic "Visual Analytics for Decision Making under Uncertainty – VADMU." The PROVIDEDH project intends to bring uncertainty to the surface by giving Digital Humanities researchers a space to explore and assess the completeness and evolution of digital research objects, the degree of uncertainty that the models applied to the data incorporate and to share their perspectives and insights with the project's broad range of stakeholders. The project's goal is to produce example implementations of uncertainty-aware data analysis pipelines to support Digital Humanities research, putting special emphasis on advancing the state-of-the-art of HCI and visualization in this context. This goal is realisable through the kind of support and enhancement of the "Functional Primitives of Humanities Scholarship" envisioned by Unsworth, chief among which is the application of collaborative annotation of TEI texts to the explorative analysis of datasets.

In this workshop, we showcase the benefits of our platform by conducting a collaborative experiment that consists of three well-differentiated parts: Firstly, we will ask participants to

---

[*]Speaker

read and annotate example texts with uncertainty using the proposed interface. After these annotations have been stored in the system, they will be explored using a set of dedicated visualizations to gain insight on the nature of the dataset, e.g. identifying incomplete excerpts, ambiguous references and potential transcription errors, among other traits. Finally, the annotations will be used by the system to support collaborative visual decision-making in a standard entity normalization task typical of many Digital Humanities projects.

Target Audience: Those involved in a Digital Humanities project with special interest in uncertainties, data complexity, and visualization. Maximum: 20