Uncertainty in Digital Humanities track summary

Roberto Therón
Department of Computer Science and Automation
University of Salamanca
P.O. Box 37008
Spain
theron@usal.es

Eveline Wandl-Vogt
Austrian Academy of Sciences
Austrian Centre for Digital Humanities,
Sonnenfelsgasse 19
1010 Vienna
Austria
eveline.wandl-vogt@oeaw.ac.at

ABSTRACT

Digital humanities methods and techniques have been applied to a wide spectrum of disciplines with results there were not attainable otherwise. However, uncertainty is present in any DH endeavour, in various types and degrees, due to the very nature of both humanist research and the intervention of computers to enable such kind of research methodologies. On the one hand, information sources need to be converted into digital form and, on the other hand, the resulting digital objects are processed to obtain summaries, geographic and temporal distributions, networks, patterns, etc., that are presented to and interpreted by the researchers or users of the DH projects. From the digitalization to the interpretation, both the computational and cognitive steps involved are subject to uncertainty. The track "Uncertainty in Digital Humanities" of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM 2018), tries to bring together researchers in Digital Humanities and reflect on the opportunities and challenges related to the way in which uncertainty can be better understood and pave the way towards appropriate methods and tools of dealing with uncertainty that improve the support to humanists when generating new knowledge.

CCS CONCEPTS

• **Information systems** → Uncertainty • **Information systems** → Decision support systems   • **Information systems** → Data analytics • **Applied computing** → Arts and humanities

KEYWORDS

Decision-making, uncertainty, progressive visualization, data provenance, digital humanities

1 INTRODUCTION

In the recent years, with the pervasiveness of computers and a great variety of electronic devices connected to the Internet, Digital Humanities (DH) as a research field has experienced a great transformation that has permitted the completion of very ambitious projects with large impact in the society beyond the academia. This has resulted in a major economic impact in the cultural and creative industry. A number of new and powerful ICT have made possible the exploitation of a wealth of data (either digitized or digitally born) that have, on the other hand, changed enormously the practice in DH, and exposed novel challenges that must be faced in order to complete any of the said projects. 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 DH project is immensely vast (especially if we take into account the introduction of linked-data, Big Data, Artificial Intelligence and Citizen Science), so the traditional humanities tools have to be either substituted or aided with ancillary tools in the form of interactive visualisations or novel user interfaces.

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 the uncertainty pertaining to both the datasets and the models behind them.

Thus, it is a very common challenge in any DH project the fact that the combination of diverse data from varying sources not only enriches our knowledge but it also confronts us with vague, uncertain or even conflicting information [1]. Although previous efforts have been devoted in the recent past to mitigate this situation [1][2] [5][3][4], they pertain to individual projects and problems, so there is an urgent need to expand the research towards a more general conceptual framework around the nuances of dealing with uncertainty in DH.

One result of these many adjustments, adaptations and migrations is that the sources, nature and role of uncertainty in humanities research, and the options researchers have to manage it, are changing. Debates which previously could not be resolved in a satisfactory way can now be argued statistically, but, at the same time, certain rich modes of information input, from the library shelf to the potsherd, have been deprecated in the shadow of their less contextualised digital surrogates. This session will feature a range of perspectives on how humanistic researchers' relationship to uncertainty has changed in the digital age, how the risks might be managed and the opportunities exploited, and what digital research

in other disciplines might learn from the lessons of uncertainty in DH.

As a continuation of previous editions devoted to new trends in digital humanities [6] [7], the track on "Uncertainty in Digital Humanities" of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM 2018), tries to bring together researchers in Digital Humanities and reflect on the opportunities and challenges related to the way in which uncertainty can be better understood and pave the way towards appropriate methods and tools of dealing with uncertainty that improve the support to humanists when generating new knowledge.

2 LINES OF WORK

The track is intended to examine the state of the art of Digital Humanities regarding the following topics (and related topics present in other tracks):

- Concepts of uncertainty in various disciplines
- Understanding all the sources of uncertainty that can affect the DH practice
- Assessing the degree of uncertainty of data sources
- Quantifying and measurement of uncertainty in various disciplines
- Uncertainty, risks and innovation
- Uncertainty and digital transformation
- Communication of uncertainty to the user/researcher
- Uncertainty and teaching, communication of uncertainty to scholars
- Uncertainty and the media, communication of uncertainty to non-scientists
- Applications
- Software and tools for uncertainty management
- Technologies like semantics, linked data and language processing for data uncertainty
- (Progressive) Visualisation of uncertainty
- History of discussion certainty and uncertainty in science

Particularly, the track will trigger open discussion on the above topics that will be organized in four lines of work:

1. Sources of uncertainty in DH
2. Uncertainty management
3. Artificial intelligence and machine learning approaches to uncertainty
4. Frameworks and models considering uncertainty in DH

3 Accepted papers

Nine papers have been accepted in this track, with two papers in line 1 (Sources of uncertainty in DH), two papers in line 2 (Uncertainty management), three papers in line 3 (Artificial intelligence and machine learning approaches to uncertainty) and

two papers in line 4 (Frameworks and models considering uncertainty in DH).

Papers in line 1 (Sources of uncertainty in DH):

- *Analyzing Constructional Change: Linguistic Annotation and Sources of Uncertainty*. In this paper the authors present the various sources of uncertainty encountered in a project focusing on the investigation of language elaboration processes in Middle Low German.
- *Toward supporting decision-making under uncertainty in digital humanities with progressive visualization*. The authors examine in this paper the literature that have characterized the types and sources of uncertainty in other fields, with the intent of establishing a foundation upon which build novel computational tools supporting the decision-making under uncertainty processes that DH is currently facing.

Papers in line 2 (Uncertainty management):

- *Dotmocracy and Planning Poker for Uncertainty Management in Collaborative Research*. In this paper the author presents a study as an exploratory approach describing and analysing the use of dotmocracy (from participatory design) and planning poker (from Agile frameworks) for decision-taking and uncertainty management in the early definition of collaborative research processes.
- *Managing Uncertainty in the Humanities: Digital and Analogue Approaches*. This paper provides a high-level view of both the sources and status of uncertainty in humanities research and the attributes a digital system would ideally have.

Papers in line 3 (Artificial intelligence and machine learning approaches to uncertainty):

- *Applying Commercial Computer Vision Tools to Cope with Uncertainties in a Citizen-driven Archive*. This paper highlights both technical and social difficulties in reducing uncertainty in a citizen-driven archive (Topotheque), exploring ways of opening the archive for better access and provide better quality of data querying detected by the need of citizens.
- *Artificial intelligence and the automatic classification of historical photographs*. In this paper the authors attempt to answer two main questions: (1) What is the uncertainty in current archival documentation and (2) can algorithmic concepts help to quantify the level of uncertainty?. As a first result, a neural network was trained and applied to a small set of data.

- *Applying topic modeling techniques to degraded texts. Spanish historical press during the Transición (1977-1982)*. In this work, topic modeling was applied to two Spanish newspapers, ABC (converted to text via OCR) and El País (converted to text manually), for a specific period: Transition from Francoism to democracy: 1977-1982, with very similar results.

Papers in line 4 (Frameworks and models considering uncertainty in DH):

- *Representing Imprecise and Uncertain Knowledge in Digital Humanities: A Theoretical Framework and ConML Implementation with a Real Case Study*. In this paper a theoretical framework is proposed to define, characterise and express information imprecision and uncertainty in digital humanities, and describe a semi-formal implementation on top of the ConML conceptual modelling language. In addition, the framework has been applied to a real project in digital humanities.
- *Uncertainty in the spatial metadata of historical photographs – A geomatic and photogrammetric driven argumentation*. In this paper authors discuss both the uncertainty of the exact geometric conditions under which a photograph was taken and the uncertainty of the projection result from the lack of meta data standards to represent outer orientation of a photograph. Consequently, a new concept to store and handle geo-information as metadata for historical photographs is presented.

6 CONCLUSIONS

As can be seen from the works participating in the track "Uncertainty in Digital Humanities" of the TEEM 2018 conference, the DH field and its tools and ongoing developments are proving more and more useful for those involved in the decision-making processes that hold a certain degree of uncertainty.

While we're still in a period of transition from purely humanistic approaches to data exploration and knowledge extraction in digital devices, the relation between humanists and their research processes is changing and new opportunities are rising in different disciplines. As the core topic of the track, uncertainty is present in almost every DH dataset researchers must deal with, and its shape and handling are managed differently depending on the project.

Some ongoing projects, such as those presented in the discussed papers, have identified various sources of uncertainty) and propose the establishment and development of tools to handle those uncertain research objects. Other authors tackled the challenge of how to manage the uncertainty present in their work, from collaborative perspectives and the idea of building a set of common attributes to be present in an ideal digital system. In terms of opportunities, potential approaches to handling uncertainties are

discussed from the use of artificial intelligence or citizen-driven techniques to the opening of archives, the definition and use of algorithms to quantify uncertainties and train machines to handle them, or the creation of models to find common topics in document datasets. Finally, a framework to define the presence of uncertainty in the data is presented along its application to a real DH project, which can set the basis to the generalization of this kind of frameworks for projects in DH domain.

Overall, the track's papers show multiple angles already being explored on how to tackle uncertainties in different disciplines dealing with different backgrounds on DH. While a sound understanding of uncertainty in DH it is still in its early days, the proposal of new ideas and the development of systems and tools that ease the comprehension of datasets and the potential uncertainties present in them is promising and we anticipate a greater interest from humanists to approach their researches from a deeper relationship between ICT and the humanities that goes beyond the mere use of computers to store and process cultural data.

REFERENCES

[1] Binder, F. et al. 2014. Uncertain about Uncertainty: Different ways of processing fuzziness in digital humanities data. *Proceedings of the Digital Humanities*. 15, (2014).

[2] Jänicke, S. et al. 2015. On close and distant reading in digital humanities: A survey and future challenges. *Eurographics Conference on Visualization (EuroVis)-STARs. The Eurographics Association* (2015).

[3] Jänicke, S. et al. 2017. Visual text analysis in digital humanities. *Computer Graphics Forum* (2017), 226–250.

[4] Li, H. 2018. *Social Network Extraction and Exploration of Historic Correspondences*.

[5] Theron, R. and Fontanillo, L. 2015. Diachronic-information visualization in historical dictionaries. *Information Visualization*. 14, 2 (2015), 111–136.

[6] Therón, R. and Wandl-Vogt, E. 2016. New Trends in Digital Humanities. *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality* (New York, NY, USA, 2016), 945–947.

[7] Therón, R. and Wandl-Vogt, E. 2017. Overview of the "New Trends in Digital Humanities" Track. *Proceedings of the 5th International Conference on Technological Ecosystems for Enhancing Multiculturality* (New York, NY, USA, 2017), 84:1–84:2.