The CULTURA Project: CULTivating Understanding and Research through Adaptivity

Maristella Agosti[1] and Séamus Lawless[2]

[1] Department of Information Engineering – University of Padua
maristella.agosti@unipd.it
[2] Knowledge and Data Engineering Group, Trinity College Dublin, Ireland
seamus.lawless@scss.tcd.ie

Abstract. CULTURA will deliver personalisation and community-aware adaptivity for Digital Humanities communities through an innovative environment which is tailored to the investigation, comprehension and enrichment of digital humanities collections. CULTURA seeks to stimulate and support the communities of interest which form around such collections.

1 Introduction and Motivation

A key challenge facing curators and providers of digital cultural heritage worldwide is to instigate, increase and enhance engagement with cultural collections. To achieve this, a fundamental change in the way cultural collections are experienced and contributed to by communities is required. The interdisciplinary field of Digital Humanities is concerned with the intersection of ICT, knowledge management and a wide range of humanities disciplines, including digital libraries, history, and art history. Despite fresh impetus experienced in the field in recent years, current research practices in the humanities still tend to be very labour intensive, solitary and characterised by research material which is often disconnected and non-digitised. This has presented a particular obstacle to novice researchers and the general public. Widespread digitisation represents an important step forward, but the requirement remains for specialist environments which offer a rich, personalised and stimulating engagement with this digitised material. The objective of the CULTURA project is to pioneer the development of personalised information retrieval and presentation, contextual adaptivity and social analysis in a digital humanities context. This is motivated by the desire to provide a fundamental change in the way digital cultural heritage is experienced, analysed and contributed to by communities of interested individuals. These communities typically comprise a diverse mixture of professional researchers, apprentice researchers (e.g. students of culture and history), informed users (e.g. users belonging to relevant societies or interest groups, cultural or authorities) and interested members of the general public.

2 Challenges

Personalisation and adaptive contextualisation technologies such as adaptive hypermedia (AH), adaptive web, intelligent systems, personalised information retrieval (IR) systems and recommendation systems have been successful in many application areas such as education, tourism, and general information sites. These technologies reconcile each user's interests, prior experience or location to provide personalised navigations of relevant digital resources (adaptive personalisation) or to suggest personalised recommendations based on similar users' behaviour and feedback (social recommendation). However, current adaptive personalisation technologies have two key weaknesses:

- They fail to take into account any broader community of which the user is a member, thus neglecting a valuable source of insight into user intention.
- They are unaware of the structure and internal dynamics of the material to which they offer access. Such "domain awareness" is an important input to the selection and sequencing of material presented by an adaptive system to the user.

It is necessary to be aware of the activity of the entire community of interest which forms around digital humanities collections. This involves tracking the trends and the points of interest in the community. This information can then be used to identify, retrieve and present material which is likely to be of interest to the individual user. The tracking of the trends will be implemented also through a categorisation of users; users will be categorised in terms of their experience levels and their demonstrated interests, so that the choices of "similar" users can be used as an input to the personalised IR process. The resulting system will be one of the next generation adaptive systems that can make digital humanities collections more appealing to the broader public, as well as supporting the activities of professional researchers. This will lead to larger and more active communities of interest focused on the collections. Such communities are key to not only sustaining interest in our heritage but in promoting deeper understanding of, and contribution to, digital humanities collections.

3 Key Technologies

CULTURA will address these challenges through the in-depth analysis of the structure and features of digital humanities collections and by providing a personalised, adaptive, community-aware research environment in which to access them:

- Cutting edge normalisation and natural language processing techniques for information extraction will be applied, because digital humanities collections are often innately noisy, contain non-standard spelling, poor punctuation and obsolete grammar and word forms. Novel linguistic models of variation, enabling normalisation techniques to remove issues of spelling, grammar and

punctuation will be applied. Once the content has been normalised, further analysis can be conducted to perform named entity and relationship extraction. This will identify the individuals, events, and dates within a collection and the relationships between them.
- Social network analysis to several content collections and to the related communities of interest which are active around each collection will be applied. By analysing the social network evidence across the collection, (and subsequently also including any associated community-created body of knowledge) CULTURA will offer a unique insight into the relative influence of the individuals and events described in the collections, and the relationships between them. This analysis can be used to influence the selection and sequencing of content in the adaptive research environment. Social network analysis will also be applied to the community of researchers interacting with the collections. This will identify the individuals who are most influential within the research community and those who share common interests, perspectives, activity records and preferred navigation pathways with other users.
- A novel combination of AH and IR techniques to provide a personalised research environment to each user is foreseen. The resulting environment will deliver adaptive narrations, personalised IR and visualisations of both the cultural heritage collections and of the communities of researchers to which the individuals belong. CULTURA will model and use individual context, cognitive awareness, experience level and research relationships to inform a personalised interface to the cultural collections.

Through the environment offered by CULTURA each researcher will not only be able to explore a body of noisy digital cultural data in a personalised and adaptive manner, but also make their own contributions to the body of knowledge around the collection. The environment will significantly advance the current research practices in humanities and offer unprecedented opportunities to empower a broader community of researchers to access and work with cultural artefacts.

CULTURA will provide rigorous evaluation and validation of its adaptive services using high impact, contrasting, multicultural digital cultural heritage collections and diverse user communities and individuals. The 1641 Depositions[3] and the Imaginum Patavinae Scientiae Archivum (IPSA)[4] are the collections and communities to be used. The use cases, defined in collaboration with real users, will clearly illustrate how the adaptive environment will offer genuine user empowerment and unprecedented levels of engagement with these collections and communities.

4 Achieving CULTURA's Aims

To reach its aims, the project team operates on three planes: the user, the community and the content. It tracks the actions of the individual user and builds

[3] URL: `http://1641.tcd.ie/about.php`
[4] URL: `http://www.ipsa-project.org/index.php`

models of that users expertise, interests and requirements. Based on this model, it provides an adaptive personalised research environment which assists the user to navigate large and complex cultural corpuses of digital data, offering and highlighting elements and information which match the actions and choices of the user. It builds an understanding of the research community which is active in a given humanities environment. It identifies the key researchers, the influential thought leaders and the most insightful. Groups of users with similar expertise and/or interest, and uses the activities of the group to inform its adaptive responses to the individual are going to be characterised. Models of the digital humanities content artefact, identifying the items which are most popular, most important and which attract the most annotation are going to be defined. Social network analysis to the content, identifying entities, relationships, linkages and influence in a model of what is important and impactful within the corpus itself are applied. This knowledge of the content is applied in making informed and intelligent decisions regarding what content is retrieved for an individual researcher and the most meaningful manner in which that content should be presented.

5 Evaluation

In order to demonstrate the potential and value of adaptive systems, such as the CULTURA research environment, the evaluation and validation of the effects of adaptivity is crucial. Evaluating a system that is adaptive is, however, a difficult task [1]. The core characteristic of such systems, their adaptive behaviour, means that individual experiences and interactions with the system are produced for each user. Though a general awareness of this challenge has been developed, research on evaluation techniques and frameworks for adaptive and personalised systems is still in its infancy.

Current evaluation approaches build up on a past long tradition in information retrieval evaluation, among the many relevant reports two examples are [2,3], but current approaches still lack in taking into account and assess some relevant aspects, among those the situational and context-specific aspects of adaptation in the user interface when attempting to determine its value and benefit. In order to achieve this, there is a need for the adoption of new approaches which combine different techniques, identify appropriate metrics and specify evaluation designs which appropriately assess adaptivity. Recent research in this field have started to investigate methodologically sound ways of comparison-based studies, the use and triangulation of multiple evaluation methods, and the separate consideration of different evaluation layers, but further efforts are necessary.

6 Summary

CULTURA uses leading natural language processing technology to convert “noisy” historic text (e.g. with inconsistent spelling, abbreviations, and punctuation) into consistent language which can be analysed automatically. It applies patented

influence analysis to identify the entities and relationships which are most important, within complex data artefacts and also within the communities of researchers studying these artefacts. It builds new types of personalised, adaptive environments which support the user as he participates in distributed communities of researchers, focusing on large collections of "noisy" digital humanities material.

Acknowledgements

The work has been partially supported by the CULTURA project[5], as part of the Seventh Framework Programme of the European Commission, Area "Digital Libraries and Digital Preservation" (ICT-2009.4.1), grant agreement no. 269973.

References

1. Mulwa, C., Lawless, S., Sharp, S., Wade, V.: The Evaluation of Adaptive and User-Adaptive Systems: A Review. International Journal of Knowledge and Web Intelligence, IJWKI **In Press** (2011)
2. Voorhees, E., Harman, D., eds.: TREC: Experiment and Evaluation in Information Retrieval. MIT Press, Cambridge, Massachusetts (2005)
3. Ingwersen, P., Järvelin, K.: The Turn. Springer, The Netherlands (2005)

[5] CULTURA Project Website, URL: `http://www.cultura-strep.eu/`