

The CULTURA Project: CULTivating Understanding and Research through Adaptivity

Maristella Agosti and Nicola Orio

Department of Information Engineering, University of Padua
{maristella.agosti,nicola.orio}@unipd.it

Abstract. CULTURA aims at personalisation and community-aware adaptivity for Digital Humanities through the implementation of innovative adaptive services in an interactive environment. The intention is to offer genuine user empowerment and different levels of engagement with digital cultural heritage collections and communities.

1 Introduction and Motivation

The interdisciplinary field of Digital Humanities is concerned with the intersection of information communication technology, like digital libraries systems, and a wide range of humanities disciplines, including 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 history and art history), informed users (e.g. users belonging to relevant societies or interest groups, cultural or authorities) and interested members of the general public.

2 Scientific Challenges

A key challenge facing curators and providers of digital cultural heritage worldwide is to instigate, increase and enhance engagement with digital libraries that

manage cultural collections. To achieve this, a fundamental change in the way cultural collections are experienced and contributed to by communities is required.

Personalisation and adaptive contextualisation technologies such as adaptive hypermedia, adaptive web, intelligent systems, personalised information retrieval 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 libraries. 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 exploited as an input to the personalised information retrieval process. The resulting system will be one of the next generation adaptive systems that can make digital humanities resources 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 managed by the digital resources management system. 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 is promoting and integrating the following key technologies:

- Cutting edge natural language processing, to normalise ambiguities in noisy historical texts;
- Entity and relationship extraction, to identify the key individuals, events, dates and other entities and relationships within unstructured text;
- Social network analysis of the entities and relationships within cultural content, and also of the individuals and broader communities of users engaging with the content;

- Multi-model adaptivity to support dynamic reconciliation of multiple dimensions of personalization.

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 Depositions¹ and the Imaginum Patavinae Scientiae Archivum (IPSA)² 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 Contribution by Italian Research Community

The prototype of IPSA has been used as a research tool by scholars in the history of art. Efforts have been made to present it to other researchers in Italy and Europe. Building on the experience in using the actual version of the prototype, the next step in our ongoing project is to study how to extend its functions to develop it as an education and dissemination digital library. At the same time, we will elaborate on actual functions to address a number of *research questions*, which can be addressed by automatic tasks to help scholars to discover new knowledge. In this study process, IPSA can be considered as a *case study* to learn new ways of using and extracting information of interest from new categories of users. A further step will be to generalize the findings of this case study to similar digital cultural heritage collections and applications.

Using IPSA as a new starting point to develop tools for researchers in illuminated manuscripts, we began a new analysis of requirements on the *research questions* that should be addressed by a digital library of this illuminated manuscripts. The analysis was carried out on a focused group of scholars and professional users, including professors in the history of illumination, in the history of medieval art, and experts in digitised manuscripts [1].

The initial results of this ongoing study highlighted some priorities. The research questions confirmed the results of our initial analysis of requirements, introducing additional concepts to refine the existing tools.

The user group underlined that images are the main subject of scientific research on illuminated manuscripts. Although text surrounding the image is important as well, it has to be noted that in many cases the author of a manuscript

¹ URL: <http://1641.tcd.ie/about.php>

² URL: <http://www.ipsa-project.org/index.php>

copied the text from pre-existing manuscripts, while the illustrator added original drawings. These drawings can be copied, with some modifications, from previous images, or just be inspired by them.

A second group of research questions regarded the relations between the content of the digital library and external collections. Illustrators could be inspired by manuscripts that are part of other collections, but also by other art forms of the same historical period. The main research question related to this point can be expressed in two main forms, regarding either the possibility of *finding relations with other digital libraries* or the possibility of *querying the digital library using external information*. The analysis of requirements highlighted that automatic tools, which mine the content of online collections, can be a valuable tool for researchers [2]. At the same time, the scientific research on illuminated manuscripts can take advantage of any kind of documentation that can be related to the content of the manuscripts. The possibility of having this information available when studying an image is considered of great importance. 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 in the context of the CULTURA project. The actual version of IPSA already supports an annotation mechanism that is very simple and not comprehensive of the necessary primitives that are going to be identified and managed in the prosecution of the project.

Acknowledgements. CULTURA project³ is funded within the Seventh Framework Programme of the European Commission, Area “Digital Libraries and Digital Preservation” (ICT-2009.4.1), grant agreement no. 269973.

References

1. Agosti, M., Mariani Canova, G., Orio, N., Ponchia, C.: A case study for the development of methods that improve the user engagement with digital cultural heritage collections (poster). In: International Workshop on Multimedia for Cultural Heritage (MM4CH 2011). University of Modena, Italy (2011)
2. Agosti, M., Mariani Canova, G., Orio, N., Ponchia, C.: Methods of personalizing a collection of images using linking annotations. In: Agosti, M., De Luca, E.W., Lawless, S., Leveling, J. (eds.) Proceedings of the First Workshop on Personalised Multilingual Hypertext Retrieval (PMHR 2011), In Conjunction with the ACM Hypertext 2011 Conference, pp. 10–17 (2011)

³ CULTURA Project Website, URL: <http://www.cultura-strep.eu/>