The CULTURA Project: Supporting Next Generation Interaction with Digital Cultural Heritage Collections

Gary Munnelly[1], Cormac Hampson[1], Séamus Lawless[1], Maristella Agosti[2], Owen Conlan[1]
[1] Trinity College Dublin, Ireland
[2] University of Padua, Italy
munnellg@tcd.ie,cormac.hampson@scss.tcd.ie,seamus.lawless@scss.tcd.ie,
agosti@dei.unipd.it,owen.conlan@scss.tcd.ie

ABSTRACT

This demonstration will present CULTURA [1], a dynamic, customizable web portal which provides a suite of tools designed to empower and assist a variety of users in their exploration of a number of cultural heritage collections.

CULTURA is a three year, FP7-funded project, whose main objective is to pioneer the development of personalized information retrieval and presentation, contextual adaptivity, and social analytics, all in a digital humanities context. A wide array of tools and services are employed to aid and inform the user in their exploration of digital collections.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – Information filtering, query formulation, relevance feedback, retrieval models, search process, selection process. H.3.5 [**Information Storage and Retrieval**]: Online Information Services - Web-based services. K.3.1 [**Computers and Education**]: Computer Uses in Education – Collaborative learning, computer-assisted instruction (CAI), computer-managed instruction (CMI).

General Terms

Algorithms, Management, Measurement, Design, Reliability, Experimentation, Human Factors.

Keywords

Personalisation, adaptive environments, cultural heritage, digital collections, user modelling, assisted learning.

1. CULTURA PORTAL AND DEMO

CULTURA currently supports three digital cultural collections: the 1641 Depositions, a manuscript collection of 17th century witness statements from Trinity College Dublin; the Imaginum Patavinae Scientiae Archivum (IPSA), a medieval collection of illuminated manuscripts of herbal and astrological codices from the University of Padua; and Bureau of Military History, an early 20th century collection of witness statements from the Irish Military Archives. Due to their contrasting modalities, these collections present very different challenges.

CULTURA has been developed using a service-oriented architecture. There are both pre-processing and runtime services which act upon the collections. These services include:

- Text Normalisation
- Entity Extraction
- Entity Oriented Search
- Social Network Analysis and Visualisation
- Annotation
- Personalisation.

Different combinations of these services are appropriate and required for different types of content, and the CULTURA architecture allows the toolset provided by the portal to be tailored to suit each particular collection.

This demonstration will emphasise how the end-user experience delivered is facilitated by both data layer and presentation layer services. Before the data of a collection is exposed to the user, it undergoes preprocessing to extract meaningful information which can be used by both the system and potentially the end user.

In an Early Modern English textual collection, such as the 1641 Depositions, the raw data undergoes a process of normalization, which is intended to introduce consistency into the document language by resolving some of the variant spelling into a more modern form.

Textual collections can then be passed in both original and normalized form to the entity extraction service. Entity extraction is used to identify the named entities within the texts including people, places, dates, etc. These entities provide an important insight into the nature of the text and can be used to guide a user towards resources which are relevant to their research as well as link documents which are thematically similar, such as those which mention the same individuals, places or events.

Social network analysis (SNA) can be conducted on the output of the entity extraction process, or on the metadata of an

SIGIR'13, *ENRICH Workshop* August 1, 2013, Dublin, Ireland. ACM

image-based collection such as IPSA, in order to help identify and visualise the important individuals involved, and the social networks that exist within a collection.

A user model is maintained for each individual who interacts with the CULTURA environment. Information about a user's browsing history, inferred interests and exhibited level of expertise is persistently stored, thus allowing the environment to personalized the user's experience across several sessions. For example, for a user who is exhibiting a particular interest in content relating to County Wicklow, both by bookmarking documents which relate to it and annotating bodies of text which contain references to it, CULTURA will attempt to establish what aspects of Wicklow are of interest to that user by correlating their user model with the entities extracted from those documents. The relationships determined by entity extraction process can then be used to produce lists of alternative sources which may interest the user.

Simple user interface controls in the CULTURA environment allow users to interact directly with the cultural heritage collections through annotating, sharing, bookmarking and searching. Services can also be called to visualize the social networks contained within a single resource or across the entire collection. Through methods such as this a user can explore the vast range of related entities which chain documents together.

User's seeking a more guided, tutorial based experience of a corpus can avail of "narratives" [2] which guide the user through a collection, explaining the content along the way and providing insight into the nature of the sources. A collection of narrative threads, designed by experts in the domain of the source material, can be offered to the user.

2. ACKNOWLEDGMENTS

The work reported has been partially supported by the CULTURA project, 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. We acknowledge the contribution of all project partners involved in CULTURA (http://www.cultura-strep.eu/).

3. REFERENCES

[1] C. Hampson, M. Agosti, N. Orio, E. Bailey, S. Lawless, O. Conlan and V. Wade (2012). The CULTURA Project: Supporting Next Generation Interaction with Digital Cultural Heritage Collections. In: M. Ioannides, D. Fritsch, J. Leissner, R. Davies, F. Remondino, R. Caffo (Eds), *Progress in Cultural Heritage Preservation, 4th International Conference, EuroMed 2012*, Limassol, Cyprus, LNCS Vol. 7616, Springer, Berlin Heidelberg, 2012, pp. 668-675.

[2] O. Conlan, A. Staikopoulos, C. Hampson, S. Lawless and I. O'Keeffe. The Narrative Approach to Personalisation. *New Review of Hypermedia and Multimedia* [In Press].