Personalising the Cultural Heritage Experience with CULTURA

Gary Munnelly
Knowledge and Data Engineering Group
Trinity College
Dublin, Ireland
munnelg@scss.tcd.ie

Cormac Hampson
Knowledge and Data Engineering Group
Trinity College
Dublin, Ireland
hampsonc@cs.tcd.ie

Owen Conlan
Knowledge and Data Engineering Group
Trinity College
Dublin, Ireland
Owen.Conlan@scss.tcd.ie

Eoin Bailey
Knowledge and Data Engineering Group
Trinity College
Dublin, Ireland
baileyeo@scss.tcd.ie

Seamus Lawless
Knowledge and Data Engineering Group
Trinity College
Dublin, Ireland
Seamus.Lawless@scss.tcd.ie

ABSTRACT

This paper discusses the CULTURA project and its attempts to personalise a user's experience of exploring cultural heritage collections regardless of their level of expertise. The services provided within the environment are introduced with respect to the four phase personalisation model used by CULTURA. In addition to these services CULTURA's comprehensive user model is detailed, in order to explain how a user's actions can affect the environment in which they work.

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. 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. INTRODUCTION

Cultural heritage provides a valuable resource through which future generations can be informed and learn about the origins of the modern world. In spite of being of importance, accessing the information which is inherent in artefacts is often a difficult task, particularly for the uninformed or inexperienced individual.

For a novice, searching through large quantities of information and interpreting data present a challenge. The tasks require that the researcher have some degree of knowledge regarding what they are looking for and how it will be represented in the corpus. If there are multiple differing references to an entity (e.g. 17th century Irish depositions alternatively refer to Sir Phelim O'Neill as "The O'Neill" and "The Rebel Leader"), the net effect can be that the user does not gain the full benefit of the content. The overwhelming size of the archives may even discourage the novice from researching the topic in any form.

For the expert, the challenge lies in discovering how the information is stored and what tools are available to traverse it. In general, this person knows what they are looking for and have a strong sense of how it will be documented. However they are hindered by their unfamiliarity with the services which allow them to explore the data and how these will respond to their queries.

Fundamentally, a solution to these problems should provide for the naive nature of the novice without diminishing the experience of the expert. This paper discusses the CULTURA environment, a subset of the tools it provides and how they are helping in this endeavour. It should be noted that, for the purposes of illustration, personalisation is discussed in relation to two extreme classes of user – the novice and the expert – but CULTURA is intended to cater to all levels of experience and adapt to reflect the research interests of each individual user.

Section 2 introduces the four phase model which CULTURA employs for personalisation and describes some of the tools which aid in its implementation. Section 3 describes the user model which represents the user within the system and upon which much of the personalisation is based. Section 4 discusses how these tools are being evaluated to ensure that they are providing the expected benefits for the users.

2. Personalisation in CULTURA

CULTURA is a three year, FP7 funded project [1]. Its main objective is to pioneer the development of personalised information retrieval and presentation, contextual adaptivity and social analysis in a digital humanities context. For the purposes of

personalisation, CULTURA utilises a four phase model as illustrated in Figure 1.

During the guide phase of the model, the individual travels along a sequenced path through the content of the corpus. This path can be explicitly extended with new resources by the user, or implicitly by the system which can analyse a user's level of interest. Thus the system leads the visitor through the lesson while encouraging and advising them on relevant places to explore. Ultimately it is envisioned that users will naturally transition from the guide phase to the explore phase as they become more engaged with the content.

The explore phase is where CULTRA encourages users to spend the majority of their time. Here, the user is performing personal, free-form investigations into the contents of the corpus. The environment furnishes them with the tools they require to engage and access information, but ultimately they are driven by their own desire to learn.

The reflect phase affords a user the opportunity to examine their own user model and influence how it is being constructed. Changes made to the user model directly influence the behaviour of CULTURA and how it responds to a user. Here, the system is actively engaging with the researcher, giving them the ability to correct its interpretation of their actions and adjust itself as required.

In the suggest phase, the environment considers what it thinks the researcher is interested in and offers proposals for additional content which may be relevant. The information contained in the user model is collated and used to make conjectures regarding what the individual is trying to achieve and what may be of relevance to their current interests. These recommendations are presented to the user through various means within the environment
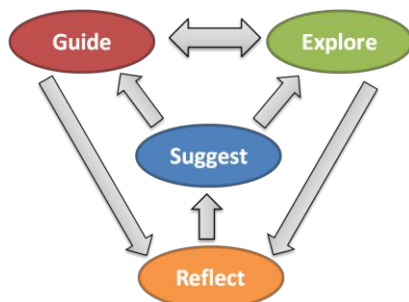
Figure 1: CULTURA's four phase personalisation model

The application of this model is facilitated by a number of services that have been incorporated into the CULTURA environment.

2.1 User Narrative

In Section 1 it was noted that both novice and expert researchers suffer from a lack of guidance when using a new system. The narrative module of the CULTURA environment is designed to provide this guidance by furnishing users with a selection of lessons which they may follow to increase their knowledge about the site and its contents [2]. As such, this module is associated with the guide phase of the personalisation model.

An individual narrative is comprised of a number of parameterised links between the various resources and services in the CULTURA environment. These links form a sequenced path which the user can traverse at their own leisure. At each stage of

the narrative, the user is presented with an artefact (or a form of visualisation of an artefact) from the corpus and a body of text which describes that stage of the lesson. This text may contain a description of the artefact, instructions for the user to follow or some other form of information.

Due to the open nature of a lesson's textual content, a narrative may additionally be used to construct a tutorial on the functionality of the site, making users aware of the various features which are available to them. This type of narrative is more likely to be of interest to the expert user than narratives which describe the artefacts themselves.

While a user is following a narrative they retain the use of other site functionality such as annotation and bookmarking of content. This information is gathered by the user model in an attempt to identify topics within the artefacts which interest them. While following the narrative, the recommender block (see Section 2.2) will continue to be updated with suggestions based on this data. Thus, even as a researcher follows a lesson plan, CULTURA encourages them to explore under their own volition.

Should a user choose to deviate from the path of a narrative (perhaps to follow an interesting subject returned by the recommender) the module allows for this. The individual's progress through a particular narrative is recorded and stored until they choose to return to the path. A button is available on each subsequent page the user visits which will return them to the narrative when they are ready to continue.

In addition to the linear path through a lesson, a narrative can also have a degree of difficulty which alters the type of content that comprises a user's path. This is to allow for the fact that certain users may not wish to cover parts of the lesson either because it does not interest them or because it does not present them with anything new. For example, a novice researcher who is studying Ireland in the 17th century may not be aware of the state of the country at that time. They will therefore potentially require a more detailed lesson in order to compensate for this lack of knowledge. An adept researcher who is familiar with 17th century Ireland, but would like to learn more does not necessarily require as much information as the novice and so may be able to take a briefer route along a similar lesson plan.

The narrative structures are implemented and stored as XML files on the server machine. The lessons themselves are designed by designated experts. The contents of the XML file dictate which resources are to appear in the lesson, how they are to be ordered and how they are to be presented (text, entity visualisation, etc).

2.2 Personalised Search

A good search engine will return a ranked list of results with artefacts being ordered according to their relevance to the query. This approach is known to be extremely effective, but it does not attempt to actively help the user achieve their goals.

When an individual performs a single search across the corpus, they are often looking for something specific. However, if they are conducting multiple searches, then it is likely that they are attempting to research a more general topic. The CULTURA personalised search is attempting to identify when a user is researching a topic, what it is they are trying to research and thus return results which are more relevant to the user's interests.

As a user executes queries, the personalised search module attempts to identify terms which are common to each of the requests. If a term is appearing with a high degree of frequency, then this is indicative that the user is trying to research a broad

topic to which this entity is relevant. When the module identifies a theme that is common to the searches, then in addition to performing the user's original query, the module can silently generate new searches by appending these relevant terms to the initial query string. The results of the multiple searches are then interwoven to produce a single ranked list which is presented to the user.

For example, consider a scenario where a user conducts a search for the Irish rebel Phelim O'Neill across a collection of depositions. The user model indicates that the researcher has been gathering information related to Louth prior to this search. Based on this, it would appear that the person's interests are more deeply rooted in County Louth than anywhere else in the country. The personalised search will thus return a list of all depositions which reference Pheilim O'Neill, but it will also silently conduct a search for depositions which reference both Pheilim O'Neill and County Louth. The results of this silent search will be assessed and interwoven with those of the initial query before the results are shown to the user.

In this manner the personalised search module attempts to actively and dynamically assist the user in their search for information rather than mechanically responding to static queries.

2.3 Recommender Module

As mentioned previously, CULTURA attempts to encourage users to actively explore and investigate the data that it curates. The recommender module is one of the primary means by which this is achieved. This service presents the user with lists of various entities which appear to be relevant to the topics which they are exhibiting an interest in. This entices the user into following new threads of information through the repository. In addition to encouraging active explorers, it also has a role in easing researchers from the guide phase to the explore phase by recommending deviations from the lesson plan.

In a similar manner to the personalised search, the recommender module attempts to identify entities and themes in which the user is expressing an interest. It can derive these terms from the user model which is constructed as the user interfaces with the system. The recommender then attempts to locate entities which are related to these identified terms. This is achieved through the use entity relationship extraction which is performed in CULTURA using IBM's LanguageWare [3].

As all the entities which exist in the corpus are known, CULTURA attempts to identify and link entities which are related. The recommender module uses these links to locate new artefacts which may be of interest to the user. The recommender module's suggestions are presented as a list in a statically located block alongside the content (Figure 2). In addition to presenting the user with the recommender module's suggestions, the recommender block also indicates why a particular entity is being suggested to them. This is particularly important due to the means by which the user interfaces with their user model (see Section 3).

As an example of how the recommender module may be of assistance, consider a user who has been reading many documents which describe the town of Trim in the 17th century. The recommender notices that Trim is a common term in many of the artefacts that the user is viewing. It examines the network of entities which are linked to Trim and finds that Hugh Morison gave a deposition which contained information about this town. It therefore suggests that the user read the deposition of Hugh Morison in order to get more information about this place of interest (see Figure 2).

Recommended Depositions

More about Lismore:
Deposition of John Pepper
Deposition of John Smith
Deposition of William Needs & John Laffane
More about Trim:
Deposition of Thomas Hugines
Deposition of Hugh Morison
Deposition of Richard Thurbane
More about Meath:
Deposition of Jane Hanlan
Deposition of Elizens Shellie
Deposition of Richard Ryves

Figure 2: The recommender block showing suggestions

3. User Model

Much of the functionality described in Section 2 is facilitated by CULTURA's comprehensive user model. The user model reflects the system's interpretation of the researcher and their goals.

The model is constructed dynamically as the user interacts with the environment and maintained for the lifetime of the user's account. Information such as pages the user visits, content annotated, artefacts bookmarked, searches performed, visualisations viewed and commonly observed entities are logged. Out of this data the user model extracts the common entities which were present in the user's searches. Recurring instances of a particular place, person etc. are considered indicative of an interest in that particular artefact. The more often an entity is observed, the greater the user's curiosity is believed to be. From this, conjectures can be made with regards to the user's long term interests, short term interests and level of expertise with respect to the site's content.

Due to the fallible nature of computers, it is possible that the model which the system generates will not be an accurate representation of the user. Facilities are therefore provided which allow an individual to tweak their model so that it is more in line with their actual intentions. This forms the basis of the reflect phase.

At present, modification is performed by means of an interactive tag cloud (Figure 3) where the model's interpretation of a researcher's interests is presented as a list of weighted terms within the cloud. Users can select entities which interest them and increase their weights so that they have a larger bearing on the recommendations that they receive. Alternatively, the user can reduce the weight of a term so that it has a negligible effect on their experience.

The user model is an invaluable resource for the provision of a unique, personal experience for the individual researcher. If well-structured, the model allows the system to offer meaningful information during the suggest phase. Furthermore, by allowing

the user to modify their representation as they require, accurate personalisation can be assured.

User Model

View | Edit | Outline

Tag Cloud Controls

Capten Blundell | Increase Weight | Decrease Weight | Delete Term

assault Capten Blundell Justice Donellan military action

multiple killing Patricke Keregan

Richard Strong Robert Boylan Tady Elllis Tipperary

Westmeath Wicklow

Figure 3: Tag cloud for interacting with the user model

4. Deployment

At present the personalisation services described in this document are being trialled as part of the CULTURA environment. In its current form, CULTURA aims to provide adaptive and personalized access to two historical collections – the 1641 depositions [4] and IPSA collection [5].

The 1641 depositions are a collection of hand written manuscripts, mostly legal in nature, which provide a unique insight into the cultural and political state of Ireland in the 17th century. The content of these documents is entirely textual. However, due to the non-standardised state of the English language in that century, the data is noisy with many inconsistencies in spelling and referencing of entities. The collection challenges the personalisation tools due to the inconsistent nature of the language it contains.

The IPSA (Imaginum Patavinae Scientiae Archivum) manuscripts are a collection of illustrated documents from Italy which describe the various properties of herbs and plants dating from as far back as the 14th century. Each record is comprised of hand drawn images of various floras, but contains little or no text. For investigators, the interest in a particular artefact is influenced by the content of the image, rather than textual data. Thus these manuscripts present a different challenge for the personalisation tools to the 1641 depositions.

The personalisation services have been trialled on several occasions during the development of the system and continue to be adapted to reflect the feedback from the users [6]. Volunteers who tested the system ranged widely in levels of expertise from the complete novice to relative expert. Among the groups involved were a group of secondary school students from Lancaster and a number of researchers from Padua, Italy.

As development of the tools continues, more user trials will be conducted to determine the usefulness and accuracy of the services.

The results of trials are being gathered, managed and examined using CULTURA's own evaluation engine – Equalia [7].

5. Conclusion

The personalisation tools which have been developed as part of the CULTURA project show great promise for improving user interaction with cultural collections. Although many of the tools are still in development, early user trials indicate that researchers of all levels of expertise are deriving benefits from their provision.

For the remainder of the project, these tools and services will continue to be enhanced in order to maximise the advantages experienced by the users. As mentioned in Section 1, ultimately the goal is to provide a system which is both accessible and equally useful to researchers at all levels of expertise.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] The CULTURA Project: Supporting Next Generation Interaction with Digital Cultural Heritage Collections. C. Hampson, M. Agosti, N. Orio, E. Bailey, S. Lawless, O. Conlan, V. Wade Ding, W. and Marchionini, G. 1997. *A Study on Video Browsing Strategies*. Technical Report. University of Maryland at College Park.

[2] The Narrative Approach to Personalisation, Owen Conlan, Athanasios Staikopoulos, Cormac Hampson, Seamus Lawless & Ian O'Keeffe, New Review of Hypermedia and Multimedia [In Press]

[3] LanguageWare, IBM. Retrieved 06 2013, from http://www-01.ibm.com/software/globalization/topics/languageware/

[4] 1641 Depositions, Trinity College Dublin. Retrieved 06 2013, from http://1641.tcd.ie/index.php

[5] IPSA – Imaginum Patavinae Scientiae Archivum, Universita Degli Studi Di Padova. Retrieved 06 2013, from http://www.ipsa-project.org/

[6] "CULTURA: Supporting Enhanced Exploration of Cultural Archives through Personalisation", Bailey, E., Lawless, S., O'Connor, A., Sweetnam, S., Conlan, O., Hampson, C. and Wade, V. In the Proceedings of the 2nd International Conference on Humanities, Society and Culture, ICHSC 2012, Hong Kong, China, 2012.

[7] An Evaluation System for Digital Libraries, Alexander Nussbaumer, Eva-Catherine Hillemann, Christina M. Steiner, Dietrich Albert. TPDL 2012: 414-419