

# Dynamic Personalisation for Digital Cultural Heritage Collections

Cormac Hampson<sup>1</sup>, Eoin Bailey<sup>1</sup>, Gary Munnelly<sup>1</sup>, Séamus Lawless<sup>1</sup> and Owen Conlan<sup>1</sup>

<sup>1</sup> Knowledge and Data Engineering Group, Trinity College Dublin, Ireland  
{cormac.hampson, eoin.bailey, munnellg, seamus.lawless, owen.conlan}@scss.tcd.ie

## Abstract.

The number of digital collections in the cultural heritage domain is increasing year on year. Improved quality of access to cultural collections, especially those collections which are not exhibited physically is a key objective of the digitisation process. Despite some successes in this area, many digitised collections struggle to attract users or to maintain their interest over a prolonged period. One of the key reasons for this is that users of these archives vary in expertise (from professional researchers to school children) and have different tasks and goals that they are trying to accomplish. This paper describes CULTURA, an FP7 funded project that is addressing this specific issue through its four-phase personalisation approach and accompanying suite of services. By employing such personalisation techniques, CULTURA is helping the exploration of, linking to, and collaboration around cultural heritage collections.

**Keywords:** CULTURA, Digital Humanities, Personalisation, Adaptation, Exploration; Communities

## 1 Introduction

Recent years have seen a huge growth in the digitisation of cultural heritage collections. However despite this increased availability of cultural archives, curators can still struggle to instigate and enhance engagement with these collections. Simple “one size fits all” web access is, in many cases, not appropriate in the digital humanities, due to the size and complexity of the artefacts. Furthermore, different types of users need varying levels of support, and every individual user has their own particular interests and priorities. Personalised and adaptive systems are thus important in helping users gain optimum engagement with these new digital humanities assets.

Without improved support for personalisation, digital cultural heritage collections will struggle to reach their full potential. This paper describes the CULTURA project [1][2] which is directly addressing this problem. Specifically, CULTURA employs personalisation to help support exploration of digital collections, the collaboration of users around these collections and to understand a user’s interests in external digital

archives with similar content. Section 2 outlines some background research on personalisation in digital cultural archives, and describes the aims of the CULTURA project. Section 3 introduces CULTURA's four-phase personalisation approach; section 4 outlines its user model; with section 5 highlighting the use of community adaptation. Finally section 6 discusses how CULTURA can link to external resources; with section 7 summarising the paper.

## 2 Background and the CULTURA Project

While there have been recent attempts to use Adaptive Hypermedia techniques to support the personalised retrieval, interrogation and presentation of cultural heritage content collections, these have to-date been limited. The MultimediaN N9C Eculture project<sup>1</sup> aims to provide multimedia access to distributed collections of cultural heritage objects. It is an aim of the project to support the generation of various types of personalised and context-dependent presentations of cultural material. However, the current system only provides static semantic search across entities in manually annotated content collections.

The CHIP project<sup>2</sup> aims to provide personalised presentation and navigation of the Rijksmuseum cultural resources. The Artwork Recommender supports the rating of artworks/topics to generate a user profile, which is then used to drive future artwork recommendations. The Tour Wizard is a web-based tool which uses the user profile to semi-automatically generate personalised museum tours. In the MOSAICA<sup>3</sup> project a mobile device-based demonstration is used to engage novice and intermediate users. The system does provide virtual visitors with access to structured descriptions of collections through a search interface, but little adaptivity or personalisation of the experience is used.

The QViz<sup>4</sup> project has some similarities in approach to the CULTURA project in that it makes explicit recognition of the value of users as members of communities, and as contributors to digital cultural heritage collections. The focus of the QViz system is on temporal and spatial search and retrieval of archival content. While QViz is a social semantic application, facilitating user contribution and structured representation of knowledge, it does not have a personalised or adaptive aspect.

Because CULTURA is producing a generalisable solution, it must be able to add value to a wide range of digital cultural heritage collections, of which there are many. One example is the Europeana project<sup>5</sup>, which represents metadata from collections across many EU member states. While Europeana does not directly host content, it is a large repository of metadata which could be processed, alongside a specific collection's content, to seed the CULTURA environment. Many projects within the cultural heritage domain, including PATHS<sup>6</sup> and Natural Europe<sup>7</sup>, already encompass

---

<sup>1</sup> <http://e-culture.multimedien.nl/>

<sup>2</sup> <http://www.chip-project.org/>

<sup>3</sup> <http://www.mosaica-project.eu>

<sup>4</sup> <http://www.qviz.eu/>

<sup>5</sup> <http://www.europeana.eu/portal/>

<sup>6</sup> <http://www.paths-project.eu/>

rich metadata from Europeana within their environments.

Improved quality of access to cultural collections, especially those collections which are not exhibited physically, is a key objective of the CULTURA project. Moreover, CULTURA supports a wide spectrum of users, ranging from members of the general public with specific interests, to users who may have a deep engagement with the cultural artefacts, such as professional and trainee researchers. To this end, CULTURA is delivering a corpus agnostic environment, with a suite of services to provide the necessary supports and features required for such a diverse range of users.

Within the CULTURA project, two rich cultural archives are being used to showcase the features offered by the environment. The 1641 Depositions are seventeenth-century manuscripts that comprise over 8,000 witness statements, relating to the Irish rebellion of 1641; and the Imaginum Patavinae Scientiae Archivum (IPSA) collection is a digital archive of illuminated astrological and herbal manuscripts from the 14th century. The IPSA manuscripts have the rare characteristic of containing high quality and very realistic illustrations, and the archive consists of digitised images and related metadata descriptions. As such, from a technical perspective, IPSA represents a very different kind of digital humanities collection to the 1641 Depositions. Importantly, initial evaluations of the work done with both the IPSA [3] and 1641 collections [4] have yielded positive results.

### **3 Four-Phase Personalisation Approach**

The employment of adaptation techniques can help empower experienced researchers, novice researchers and the wider community to discover, interrogate, and analyse cultural heritage resources. Hence, core elements of the CULTURA architecture are its personalisation methods. The techniques employed by CULTURA have been heavily influenced by Adaptive Hypermedia (AH) and Adaptive Web systems research, which are concerned with improving the retrieval and composition of information. This improvement is achieved by creating a more context-sensitive and personalised interaction with digital content, and is often predicated on rich metadata [5].

One reason why novice users struggle to engage with large cultural collections is a lack of guidance when they initially encounter a set of resources. Likewise, more experienced researchers often lack the tools to efficiently search, share, visualise, analyse and correlate data from important cultural collections. To help counteract such issues, CULTURA employs a four-phase personalisation approach (see figure 1). Each of these phases (*guide*, *explore*, *reflect* and *suggest*) are now described in turn.

---

<sup>7</sup> <http://www.natural-europe.eu/>

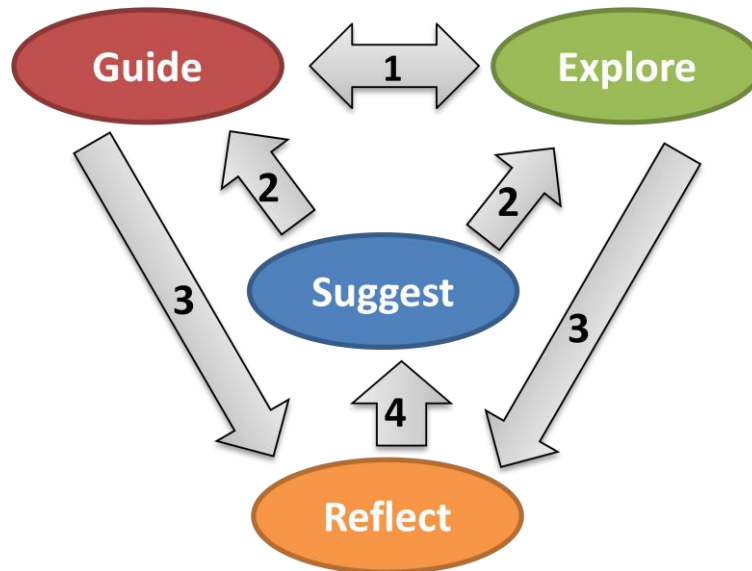


Fig. 1. CULTURA's Four-Phase Personalisation Approach

### 3.1 Guide

Users with little experience of the underlying resources typically start their investigations within the guide phase. Here CULTURA employs a “narrative” module, which enables resources within the collection to be sequenced on a specific theme e.g. the chain of derivation of illustrations in the IPSA collection, or the reliability of witness statements in the 1641 Depositions. Furthermore, how these resources are rendered to the user (text, visualisation etc.) can also be specified within the narrative metadata (encoded as XML). Guiding users through a collection is facilitated through parameterised launch of services using URLs, and is especially useful in providing users with a path through specific content (see figure 2). Importantly, the narratives can be adjusted in length, either explicitly by the user (by choosing to see more resources on the theme) or implicitly by the narrative module, which analyses the user model for changes in user interest. An authoring tool based on SABer [6], which helps non-technical people to encode domain expertise, is planned to assist the creation of these narratives.

Being on a guided path does not limit a more adventurous user, as they can use these sequenced narratives as a springboard for their own investigations (in fact many narratives explicitly encourage users to do so). Within the four-phase personalisation approach, this involves stepping from the *guide* phase to the *explore* phase (number 1 in figure 1). Importantly, by monitoring the user model metadata (perceived level of interest in particular concepts/entities etc.) as they explore the resources, the narrative path itself can be adapted. This is achieved by selecting documents for their path that most closely match the user's interests, and can result in the path being enriched with further resources and concepts.

Lesson Block

**What happened in Drogheda on the outbreak of rebellion?**

As the rebellion took hold in the North of Ireland, the colonial authorities scrambled to gather adequate forces for the defence of the colony. A key victory occurred with the successful repulsion of the rebel siege at Drogheda. What does William Fitzgerald reveal about the siege at Drogheda

3 of 18

<--PREVIOUS

NEXT-->

**Fig. 2.** Guided Lesson Plan within CULTURA

## 3.2 Explore

In the *explore* phase CULTURA offers tool assistance (e.g. data enriched maps, entity based search, social network analysis) to support exploration and browsing of the underlying resources. At any stage a user can return to where they left their path in the *guide* phase, and users with little prior knowledge of the resources often flick between the *explore* and *guide* phases several times. In contrast, professional researchers with a deep understanding of the collection typically spend the majority of their time within the *explore* phase and may never involve themselves with the *guide* phase.

The types of services that CULTURA offers during the *explore* phase include normalised search [7], entity based search [8] and geographic visualisations (see section 6). Another useful service offered is shown in figure 3, where a visualisation of the social network described within a single document of the 1641 depositions is depicted. Among the entities it depicts are the people (including their role e.g. rebel, victim, landlord etc.) and the crimes that are documented within that deposition. By clicking on an entity in this graph (represented as one of the circles in the visualisation) a user can view all other entities in the collection that are associated with that entity. This allows a user to quickly find connections between resources, which might not be so easily apparent using text searching alone. Significantly, due to the service-based architecture used by CULTURA it means that the suite of services offered within the *explore* phase can be extended iteratively over time, allowing new features to be offered to existing and future collections.

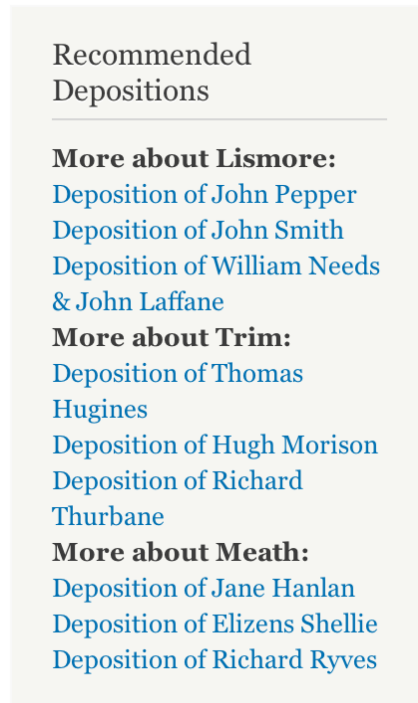


**Fig. 3.** Social Network Analysis in the "Wheel"

### 3.3 Suggest

Whether, within the *guide* or *explore* phase, a user will be given personalised suggestions for related content (see figure 4) or tools to view resources in. This process occurs in the *suggest* phase (which works in parallel with the *guide* and *explore* phases), with hints pushed to users for their review (number 2 in figure 1). These hints are influenced by the content the user is currently viewing, as well as the data stored in their user model e.g. search terms, entities commonly viewed, annotations created etc.

Figure 4 shows one example of the recommended content shown to users who browse the 1641 Depositions using CULTURA. When a user views a deposition, entities (people, places etc.) are extracted from the text, and complementary depositions that also mention these entities are located within the collection. For instance, in figure 4, blue text links to depositions related to the entities "Trim", "Meath" and "Lismore" are displayed. By showing the entity names beside the recommended links, it makes clear to the user why these links are being rendered to them. The recommender box is displayed beside the deposition text and enables users to quickly locate new resources that are relevant. Furthermore, by clicking on one of these links, the user model is updated with the corresponding entity, as it indicates a user interest in said entity. Other services that are triggered in the explore phase include the search module, where the results presented to users are influenced by terms stored in a user's model.

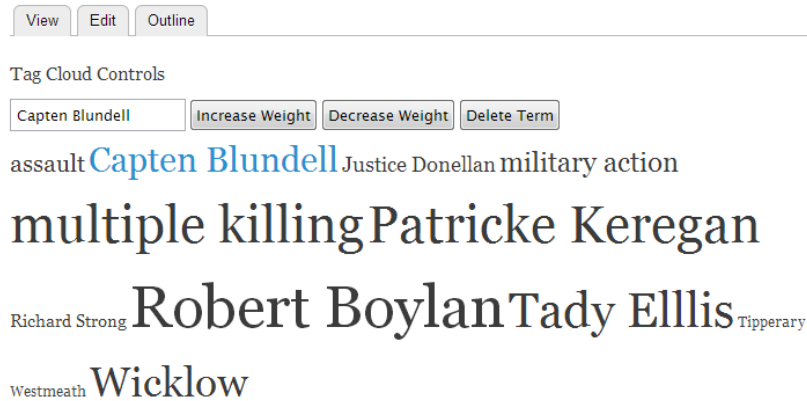


**Fig. 4.** An example of the recommended content displayed to users

### 3.4 Reflect

At any stage within the *guide* or *explore* phases, a user may enter the *reflect* phase (number 3 in figure 1), which involves viewing a subset of their user model (rendered as a tag cloud) and seeing what terms are influencing the recommendations they are receiving. Moreover, they can easily delete terms they deem irrelevant or increase/decrease the relative size of terms depending on how influential they think they should be. New terms will also be allowed to be manually added. Figure 5 shows the current user interface for interacting with a user model. Such scrutability [9] of user models is vital for making the underlying personalisation processes more transparent, and gives users a greater feeling of control over the adaptation features offered by CULTURA. Importantly, any changes made by users during the *reflect* phase directly impacts on the *suggest* phase and the recommendations that eventually filter down into the *guide* and *explore* phases (number 4 in figure 1).

## User Model



**Fig. 5.** User Model Tag Cloud

## 4 User Model

By espousing the four-phase personalisation approach, CULTURA dynamically adapts to users, and renders useful suggestions to them at appropriate times. Moreover, this process provides mechanisms that are appropriate for a range of users with different levels of ability or different interests in the underlying resources. A central component of this process is CULTURA's user model, which is stored in MongoDB<sup>8</sup>, a scalable high-performance NoSQL database. One of the main benefits of MongoDB is that it allows the database schema to evolve over time, which is very useful in a service orientated environment such as CULTURA, where new suites of tools are being introduced gradually.

A model of each user is built silently as a user interacts with the system, however, as described in section 3.4, each user will be given the opportunity to scrutinise and change their model. All actions a user performs are recorded in order to build up detailed information on each user. This includes viewing, bookmarking or annotating a deposition, as well as the searches performed, entities commonly viewed and the visualisations they rendered. Further user information is to be incorporated into the model such as the relative expertise of a user, and the communities they are a part of. By analysing the entities that a user has annotated it will also provide better evidence of a user's interest.

Important elements of CULTURA's user model to be developed in the next phase of the project include allowing its partition into different parts (users can choose to reset their user model when tackling different tasks or projects), as well as to have

---

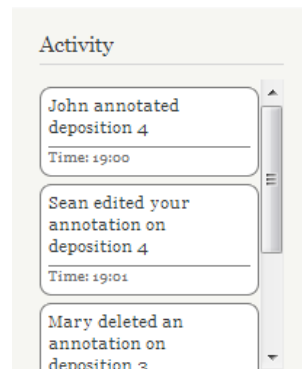
<sup>8</sup> <http://www.mongodb.org/>



both short-term and long-term models of interests. By storing such rich details in the user model, it will enable a better understanding of the user and ultimately better recommendations to be made to them.

## 5 Community Adaptation

Complementing the user model information stored in MongoDB are the comprehensive logs that are stored in Drupal's MySQL database. These provide further rich metadata (such as time-stamped logins and page views) that can be exploited by the system for personalisation and adaptation purposes. This supplementary information is being combined with the user model and can be displayed to the user so that they have a record of the actions they have performed in the system. There will also be the option to share this data with your community and to view what others have been doing in the system (see figure 6). Such features help encourage community awareness within the CULTURA environment.



**Fig. 6.** Prototype Activity Feed in CULTURA

A central component of CULTURA is its annotation service CAT [10][11], which enables the annotation of both text and images within the CULTURA environment. Further to allowing a user to comment on document text and images, the annotations created using CAT allow a user to link their annotations to other resources internal and external to the CULTURA environment. While CAT is beneficial for researchers and educators, it is also being used as an important source of user metadata for CULTURA. For instance, annotations provide an insight into which entities are of interest to a user. If a user is frequently annotating a document, it is likely that this document is of interest to them. Furthermore, by analysing the text being annotated (as well as the text of the annotation itself) using entity extraction; it is possible to discern specific items of interest to the user. Other data that the annotation tool allows to be collected is the interaction of users within a group, and how often a person views annotations by a specific contributor. This data can improve the recommendations given by CULTURA, as well as help foster collaboration within the environment. However, the development of adaptive user communities is a challenging problem. For instance, the different user types require different degrees of sharing and

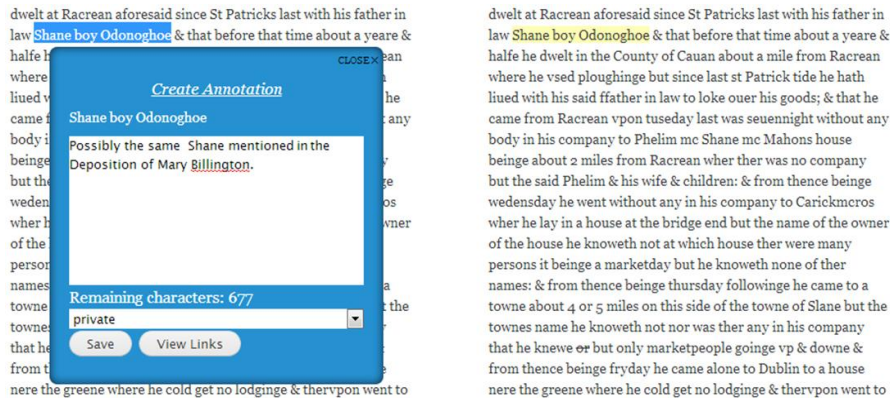


Fig. 7. Screenshot of CAT within CULTURA

grouping, e.g. a professional researcher generally does not want to share their annotations prior to publication of research whereas an interested member of the public often will have no such issues.

With the increased richness of user models and logs, coupled with increased usage of the system, it means that it will also be possible to establish recommenders for similar users in both the 1641 and IPSA collections. This will greatly benefit those wishing to find other users to collaborate with, or relevant communities to join. Importantly, by analysing the members of specific communities, it will be possible to determine common characteristics shared in their individual user modules. This gives a better understanding of the community as a whole. Furthermore, with increased community use of CULTURA, it means that Influence Network Analysis (a form of social network analysis that identifies the most influential nodes within a graph) can be performed over these communities. The metrics used to rank an entity's influence in its community can also be used to adjust the adaptations received. Enhanced community-based adaptations, which account for items such as a group's type (students, professionals, a mixture etc.), size, diversity, activity levels, and length of time in existence, are also a priority.

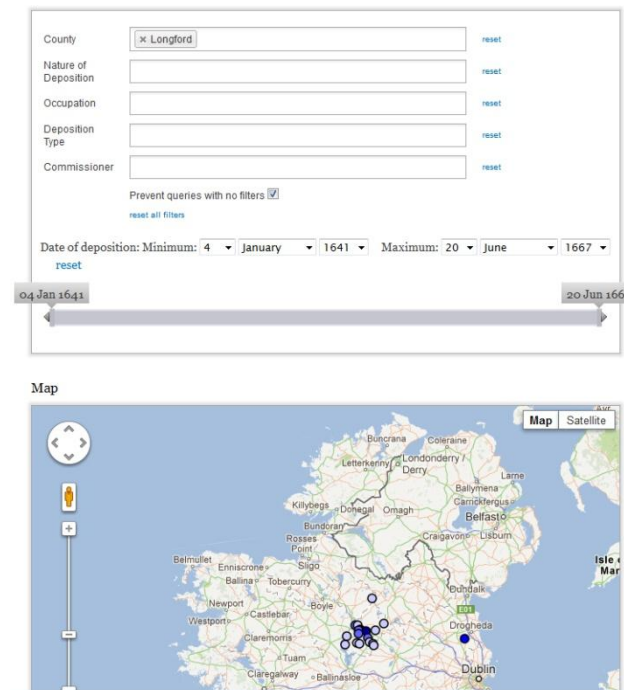
## 6 Links to External Resources

If a data collection contains entities that are referenced on the web of data, it is possible to map between these entities so that external information on that entity can be incorporated into the CULTURA environment. For example, the 1641 depositions contain many place names which are linked to the geonames<sup>9</sup> database. An example of what can be achieved with such an approach is shown in figure 8, where depositions from the 1641 collection are mapped to a specific geographical area and visualised on an interactive display. By creating links between CULTURA collections and

<sup>9</sup> <http://www.geonames.org/>

the wider web, it helps to increase traffic to archives and provide a richer landscape for digital humanities research. This approach of enriching CULTURA with data from open corpus content will be further examined in future.

If the external collections use dereferenceable URIs for entities (as Europeana<sup>10</sup> does), a complementary process can occur, where data stored in CULTURA's user model is used to help deliver more relevant recommendations for resources within these external archives. Furthermore, when new collections are added to CULTURA itself, and they contain similar content and entities as existing archives, a user's model can help to determine relevant resources from the new collection, as well as assisting against the *cold start* problem. Thus, incorporating personalisation into CULTURA's interactions with external collections has great potential, and can greatly help in improving a user's experience with new cultural heritage archives.



**Fig. 8.** Visualisation of deposition locations within CULTURA

## 7 Summary

This paper has described how personalisation is being used in the CULTURA project to help users of different skills and level of interest, to engage more meaningfully

---

<sup>10</sup> <http://pro.europeana.eu/linked-open-data>

with cultural heritage collections. CULTURA's four-phase personalisation approach was introduced along with details of how user models and logs are deployed within its environment. CULTURA's focus on community adaptation and links with external data collections were also outlined. Detailed evaluations, which include all relevant stakeholders from the domain are ongoing, and will feedback into the refinement of the personalisation technologies and methodologies employed by CULTURA

**Acknowledgments.** The work reported has been funded by the Seventh Framework Programme of the European Commission, Area "Digital Libraries and Digital Preservation" (ICT-2009.4.1), grant agreement no. 269973.

## References

1. Hampson, C., Lawless, S., Bailey, E., Yogev, S., Zwerdling, N. and Carmel, D. 'CULTURA: A Metadata-Rich Environment to Support the Enhanced Interrogation of Cultural Collections'. In the Proceedings of the 6th Metadata and Semantics Research Conference, Cádiz, Spain, pp. 227-238, Springer, Heidelberg (2012)
2. Hampson, C., Agosti, M., Orio, N., Bailey, E., Lawless, S., Conlan, O., and Wade, V. 'The CULTURA project: supporting next generation interaction with digital cultural heritage collections', In the Proceedings of the 4th International Euromed Conference, Limassol, Cyprus, pp. 668-675, Springer, Heidelberg (2012)
3. Ponchia, C., 2012. Engaging the User: Elaboration and Execution of Trials with a Database of Illuminated Images. Taking Care of Legacy and Cultural Heritage Documents – Proc. of the 8th Italian Research Conference on Digital Libraries, [In Press] (2013).
4. Bailey, E., Lawless, S., O'Connor, A., Sweetnam, S., Conlan, O., Hampson, C. and Wade, V. "CULTURA: Supporting Enhanced Exploration of Cultural Archives through Personalisation". In the Proceedings of the 2nd International Conference on Humanities, Society and Culture, ICHSC 2012, Hong Kong, China. October 27th-28th (2012)
5. Brusilovsky, P., Kobsa, A., and Nejdl, W. (Eds.) 'The adaptive web: methods and strategies of web personalization' Lecture Notes in Computer Science, LNCS, vol. 4321 (2007)
6. Hampson, C. and Conlan, O., Facilitating Casual Users in Exploring Linked Data through Domain Expertise, DEXA 2011, Toulouse, France, pp.319-333 (2011)
7. Lawless, S., Hampson, C., Mitankin, P., and Gerdjikov, S., 'Normalisation in Historical Text Collections', In Proceedings of Digital Humanities 2013, Lincoln, Nebraska, USA [In Press] (2013)
8. Yogev, S., Roitman, H., Carmel, D. and Zwerdling, N. 'Towards expressive exploratory search over entity-relationship data'. In Proceedings of the 21st international conference companion on World Wide Web, New York, USA, pp.83-92 (2012)
9. Kay, J. "Scrutable adaptation: Because we can and must." In Adaptive hypermedia and adaptive web-based systems, pp. 11-19. Springer Berlin Heidelberg, (2006).
10. Ferro, N., Munnelly, G., Hampson, C. and Conlan, O., Fostering Interaction with Cultural Heritage Material via Annotations: The FAST-CAT Way, In Proceedings of the 9th Italian Research Conference on Digital Libraries, Rome, Italy [In Press] (2013)
11. Munnelly, G., Hampson, C., Ferro, N., and Conlan, O., 'The FAST-CAT: Empowering Cultural Heritage Annotations', In Proceedings of Digital Humanities 2013, Lincoln, Nebraska, USA [In Press] (2013)