CULTURA: A Metadata-Rich Environment to Support the Enhanced Interrogation of Cultural Collections

Cormac Hampson[1], Seamus Lawless[1], Eoin Bailey[1], Sivan Yogev[2], Naama Zwerdling[2], David Carmel[2], Owen Conlan[1], Alex O'Connor[1] and Vincent Wade[1].

[1]Knowledge and Data Engineering Group, Trinity College Dublin, Ireland
{cormac.hampson, seamus.lawless, eoin.bailey, owen.conlan, alex.oconnor, vincent.wade}@cs.tcd.ie
[2]IBM Research, Haifa, Israel.
{sivany, naamaz, carmel}@il.ibm.com

Abstract. The increased digitisation of cultural collections, and their availability on the World Wide Web, has made access to these valuable documents much easier than ever before. However, despite the increased availability of access to cultural archives, curators still struggle to instigate and enhance engagement with these resources. The CULTURA project is actively addressing this issue through the development of a metadata-driven personalisation environment for navigating cultural collections and instigating collaborations. The corpus agnostic CULTURA environment also supports a full spectrum of users: ranging from professional researchers seeking patterns in the data and trying to answer complex queries; to interested members of the public who need help navigating a vast collection of resources. This paper discusses the state of the art in this area and the various innovative approaches used in the CULTURA project, with a special focus on how the underlying metadata helps facilitate its semantically rich environment.

Keywords: CULTURA, Adaptation, Personalisation, Digital Humanities, Normalisation, Social Network Analysis, Entity Extraction.

1   Introduction

The interdisciplinary field of digital humanities is concerned with the intersection of computer science, knowledge management and a wide range of humanities disciplines e.g. digital libraries, cultural heritage and digital arts. Recent large-scale digitisation initiatives have made many important cultural heritage collections available online. This makes them accessible to the global research community and interested public for the first time. However, the full value of these heritage treasures is not being realised. After digitisation, these collections are typically monolithic, difficult to navigate and can contain text which is of variable quality in terms of language, spelling, punctuation, and consistency of terminology. As a result, they often fail to attract and sustain broad user engagement leading to limited communities of interest. Thus, there

still remain important challenges in the presentation of new digital humanities artefacts to the end user.

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 necessary to help users gain optimum engagement with these new digital humanities assets.

Improved quality of access to cultural collections, especially those collections which are not exhibited physically, is a key objective of the CULTURA project [1]. 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.

A central aspect of this environment is its use of rich metadata (user generated, computer generated and expert generated) coupled with natural language processing, entity extraction and social network analysis techniques, in order to support collaborative exploration, interrogation and interpretation of the underlying cultural resources. Section 2 of this paper discusses some related work in the field of digital humanities, with section 3 outlining the key challenges this research is addressing. Section 4 introduces two case studies which are central to the CULTURA project; section 5 discusses the various components and features of CULTURA's architecture; and section 6 summarises the paper and discusses the future work that remains to be undertaken.

2   Related Work

There has been substantial effort in the area of digitisation and cultural heritage preservation. Much of this work has, until recently, been focused on the creation of digital representations of cultural artefacts, and the creation of metadata and documentation associated with this. The result of this effort is that there is a vast collection of content available to digital humanists, in the form of text, images and other representations.

Textual content resources include collections arranged by theme, such as the Biodiversity Heritage Library[1], or from institutional collections, such as the Bayerische StaatsBibliothek[2] digital collections library. These collections include varying levels of metadata and some include detailed pictures which are associated with the text. It is important to note that in many cases 'textual' content actually refers to a complex cultural artefact that includes an image of the original manuscript, transcribed text associated with the content, and metadata which can describe the content, the nature of the document, and the provenance of the digital artefact. Another important type of artefact is collections of images, with detailed metadata records associated with each image.

---

[1]   http://www.bhl-europe.eu/
[2]   http://www.digital-collections.de/

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[3] 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[4] 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[5] 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 is used.

The QViz[6] 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[7], 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.

3   Challenges in the Digital Humanites

The rise of 'i', 'me' and 'my' as prefixes for various web portals (e.g. iTunes[8]) and web services (e.g. MobileMe[9]) are intended to give the impression of personal tailoring of content and service to an individual user to enhance that individual's experience. Typically however, such services tend to focus on either: a) identification and ranking of relevant content or services; b) simplistic 'personalisation' of the content

---

[3] http://e-culture.multimedian.nl/
[4] http://www.chip-project.org/
[5] http:// www.mosaica-project.eu
[6] http://www.qviz.eu/
[7] http://www.europeana.eu/portal/
[8] http://www.itunes.com/
[9] http://www.me.com/

presentation by inclusion of the user's name, recently used resources etc.; or c) simple augmentation of screen layout.

To effectively empower communities of researchers with personalised mechanisms which support the collaborative exploration, interrogation and interpretation of complex digital cultural artefacts, it requires the adaptivity provided in CULTURA to be more integrated and intelligent than in the portals described above. Such next generation adaptivity, as espoused by CULTURA, must support the dynamic composition and presentation of digital cultural heritage resources. However, just automated adaptivity is not enough. Ensuring that the user is in control of the personalisation process is essential. Such user-centred control is enhanced through: correlating usage patterns with self-expressed user goals; pre-defined strategies (e.g. research strategies, investigation strategies, discovery strategies, explanatory strategies etc.); and the provision of appropriate tools for users to explore and navigate large cultural heritage information spaces.

A common challenge in the humanities is that historical language hinders the accessibility of historical text documents. One solution to this problem is the use of a computational historical lexicon, supplemented by computational tools and linguistic models of variation. However, because of the absence of language standards, multiple orthographic variations of a given word or expression can be found in a collection of material, even in the same document. Hence, issues arising from the need to contend with noisy inputs, the impact noise can have on downstream applications, and the demands that noisy information places on document analysis, are addressed by CULTURA.

Social Network Analysis (SNA) can be used to analyse the people and relationships contained within humanities content collections. However, the effective application of SNA techniques to content which has major inconsistencies in the naming and identification of entities, poses a significant challenge which must be overcome. Typically digital cultural heritage collections contain complex relationships between entities which must be identified and extracted from the artefacts. This is an area CULTURA directly addresses by augmenting the existing metadata with new attributes. A second challenge for SNA is to leverage the user communities, activities, contributions and profiles to discover the rich influence network that interlinks users of these digital humanities content collections. The application of SNA to both the artefacts and the community that surrounds those artefacts, as incorporated in the CULTURA environment, is novel in the digital humanities. Such community-aware adaptivity creates an integrated, engaging experience for users of all types within the CULTURA environment.

4   Case Studies

In order to validate the CULTURA environment, two major artefacts have been selected - the *IPSA Illuminated Manuscript Collection*[10], held in the University of Pa-

---

[10] http://www.ipsa-project.org/

dua, Italy and the *1641 Depositions[11]*, held in Trinity College Dublin, Ireland. These resources, and the communities of users who work with them, are central to the design, development and evaluation of the CULTURA environment. Each are now discussed in turn.

4.1   The 1641 Depositions

The 1641 Depositions are seventeenth-century manuscripts that comprise about 8,000 witness statements, examinations and associated materials, in which Protestant men and women of all classes and from all over Ireland told of their experiences following the outbreak of the rebellion by the Catholic Irish in October 1641. This body of material is unparalleled anywhere else in early modern Europe and provides a unique source of information for the causes and events surrounding the 1641 rebellion and for the social, economic, cultural, religious, and political history of seventeenth-century Ireland, England and Scotland.

The 1641 Depositions have been digitised and transcribed and are being used to validate the techniques implemented in CULTURA. From a technological perspective, the 1641 Depositions represent a textually-rich digital humanities collection, which is characterised by noisy text, inconsistent sentence structure, grammar and spelling. The English language manuscripts contain rich metadata and descriptions of individuals, locations, events, social structures and contrasting / conflicting narratives. These artefacts have active communities of interest because of their wider social and historical implications that transcend geographical and chronological boundaries and continue to shape opinions and values to this day. The 1641 Depositions represent an ideal example of a digital humanities collection, which has deep resonance with social and cultural issues encountered throughout Europe.

4.2   The Imaginum Patavinae Scientiae Archivum

The Imaginum Patavinae Scientiae Archivum (IPSA) collection is a digital archive of illuminated medieval astrological and herbal manuscript codices dating from the 14th century with Latin, Paduan and Italian language commentaries. Herbals are manuscripts which contain hand-drawn depictions of plants, such as trees, bushes or shrubs, and their parts, such as flowers or leaves. The IPSA collection contains manuscripts written and illustrated by the Paduan School, and successive manuscripts produced in Europe under its influence. Such manuscripts have the rare characteristic of containing high quality and very realistic illustrations. IPSA is a combination of digitised images of the manuscripts and related metadata descriptions.

From a technical perspective, IPSA represents a very different kind of digital humanities collection to the 1641 depositions collection. The IPSA collection is primarily image based, with substantive metadata available. This metadata not only provides descriptive passages, but is also historically valuable as it captures the scientific processes which were prevalent during the creation of the original collection. However,

---

[11] http://1641.tcd.ie/

the IPSA metadata is user generated which can lead to inconsistencies in terminology, spelling and grammar. The metadata contains descriptions of entities, individuals, activities and locations in multiple languages. The contrast in knowledge domain and structure of the IPSA and 1641 content collections demonstrate the broad applicability of the CULTURA methodology. Moreover, it highlights how the techniques delivered in CULTURA are not specific to an individual domain or collection but can be of benefit a wide range of digital humanities collections.

5   The CULTURA Architecture

CULTURA consists of multiple distinct services all accessed via the CULTURA portal. The services available in CULTURA are shown in figure 1 and include personalised search tools, faceted search tools, annotators, social and influencer network exploration tools, and recommenders. One of the key challenges for the CULTURA architecture is to reconcile the various models (user model, content metadata, extracted entities etc.) at runtime, in order to seamlessly provide the end user with the most appropriate content and services.
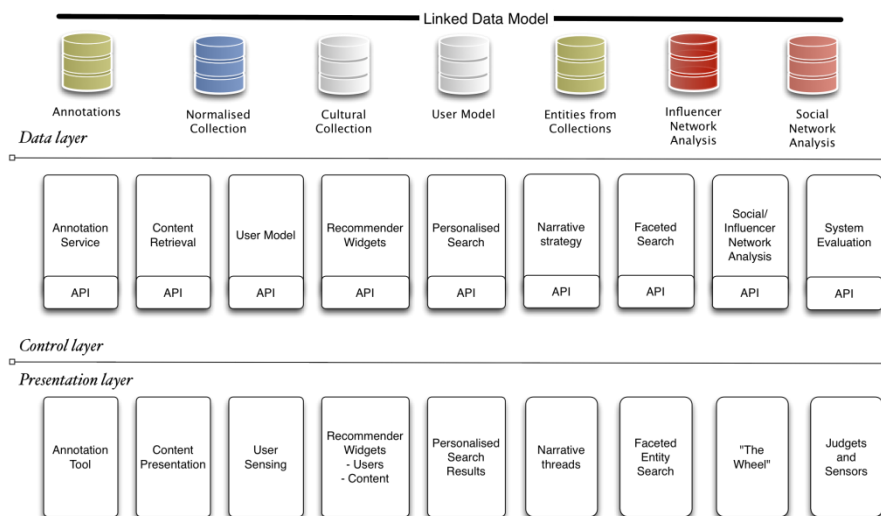

Fig. 1. CULTURA Services

A service is triggered by a user's interactions with the CULTURA portal, with requests sent from the presentation layer to the service via its API. For example, a person using CULTURA to search over the 1641 depositions triggers a series of events. First the search terms are identified and logged in the user model, which alters the areas of interest for the user. The personalised search service then normalises the search terms and identifies related entities via the linked data model. Multiple searches can now be executed using the original search terms, the normalised terms,

and the related entities. Results from the searches are then merged and presented to the user. The merge process prioritises results that are related to areas that the user has an interest in, as determined by their user model.
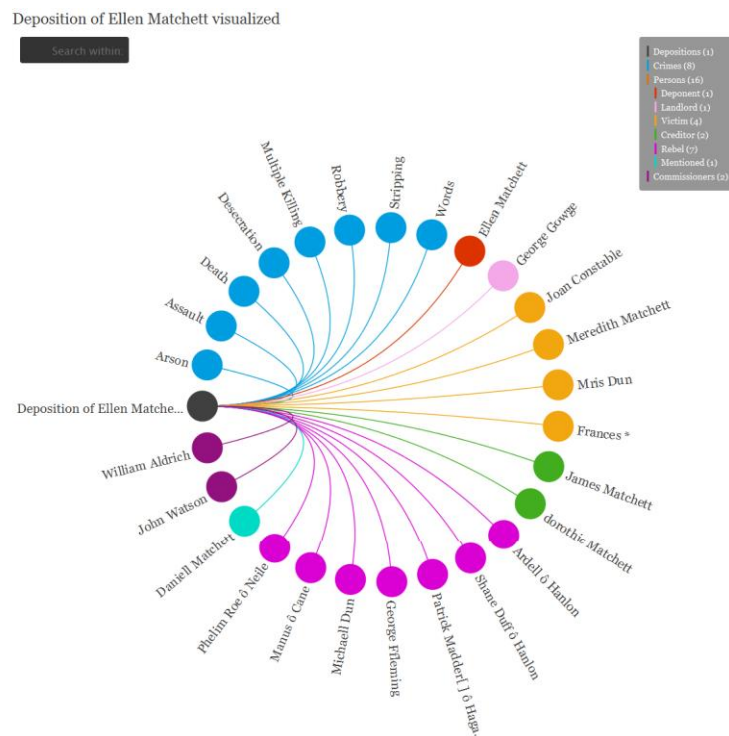

Deposition of Ellen Matchett visualized

Fig. 2. Social Network Analysis with the "Wheel"

Additional tools (implemented as distinct services) can also be enabled depending on the results shown to the user. The selection of these tools is driven by modelling the correlation between a tool's affordances and the type of content to be displayed. A user's preferences in tool use, as identified by their user model, will also influence the choice of tool. When a user selects a result, the user model is again updated to reflect the preference. Additionally the system evaluation service is invoked, to inform which result the user has selected. This information can aid in evaluating the usefulness of search results, e.g. if the user rarely selects one of the top three results the ranking of results is deemed poor for that user.

The CULTURA portal utilises Drupal[12] as it provides numerous services that, while essential to CULTURA, are not core research elements, such as user authentication and system-wide logging. Drupal also has an extensible architecture that allows new modules to be developed in order to extend or replace functionality. Hence, all services developed by CULTURA are implemented as Drupal modules, and when

---

[12] http://www.drupal.org

accessed by users, the responses from these services are displayed in appropriate form e.g. social network analysis of people mentioned within the 1641 depositions can be displayed as a "wheel" as seen in figure 2.

5.1 Personalisation and Adaption Techniques in CULTURA

A core element of the CULTURA architecture is its use of personalisation and adaption techniques. These techniques 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 [4]. CULTURA is advancing adaptivity in four main areas:

- Adaptive Information Retrieval and Presentation
- Entity-Driven Adaptivity
- Social or Community-Based Adaptivity
- Support for Micro-Adaptation Strategies and Storytelling

In order to address the limitations of traditional keyword-based approaches to the retrieval of digital content, AH and adaptive web systems have the ability to enhance user queries with contextual and user-specific information, leading to more relevant materials being presented to the user [5]. CULTURA improves upon the current 'ranked list' approach to the presentation of retrieved resources, by delivering adaptively composed responses to enable deeper interrogation of the cultural heritage collections. CULTURA also extends AH approaches to adaptive presentation and navigation techniques, in order to facilitate the user in navigating across and combining resources. This is achieved by using techniques such as:

- dynamic hyperlinking across key sections of retrieved documents
- adaptively composing and presenting responses which combine key elements of the retrieved documents
- providing guidance for the analysis of the retrieved documents
- providing facilities to add and search digital annotations [6].

The application of these techniques helps empower experienced researchers, novice researchers and the wider community to discover, interrogate, and analyse the cultural heritage resources. Figure 3 shows one example of the recommended content shown to users who browse the 1641 depositions using CULTURA. In this example, green text links to specific depositions are listed under four headings (Place, Occupation, Person Type and Nature/Crime). These links are generated for each heading by comparing the most prominent terms in a user's browsing history (displayed under "Influencing Terms") with the metadata of all depositions, and rendering the most relevant.

Personalised Content Recommendations

The recommendations on this page will change in response to your browsing pattern.

Place	Occupation	Person Type	Nature/Crime
Influencing Terms	**Influencing Terms**	**Influencing Terms**	**Influencing Terms**
Armagh	Captain	Deponent	killing
Levileglish	Clerk	Victim	multiple killing
Agralohoe	Colonel	Rebel	robbery
Armaghe	Farmer	Mentioned	captivity
Lurgan	Husbandman	Witness	words
onelaw	Knight	Landlord	arson
Turkharry	Wife	Proxy	stripping
Lissenesky	Esquire	Debtor	military action
Segoe	Gentleman	Apostate	death
Ballymore	Lord	Creditor	assault
Recommendations	**Recommendations**	**Recommendations**	**Recommendations**
Titlepage	Evidence against Michaell Doyne	Deposition of Lawrence Whitmore	Examination of Brian Clarke
List of losses	Examination of Thomas Clark	Deposition of Ralphe Twyford	Examination of William McIlduffe
Deposition of Gilbert Pemerton ex parte Thomas and Elizabeth Powell	Examination of William Sym	Certificate of John Whitman	Information of William Beere
Deposition of Lawrence Whitmore	Examination of Valentine Blacker	Deposition of William Duffeild ex parte George Williams	Examination of Toole Mc Rory Mc Cann
Deposition of John Wisdome	Deposition of Frauncis Sacheverell	Deposition of Thomas Turke	Examination of Raiph Fester
Deposition of Thomas Turke	Deposition of William Duffeild	Examination of Grace Croves	Examination of William Bentley

Fig. 3. An example of the personalised content displayed to users within the CULTURA portal.

5.2 Normalisation of Cultural Collections

Performing document analysis techniques (i.e. information extraction) on historical texts, which contain non-standard spelling, historical grammar and many old word forms, is a non-trivial challenge requiring normalisation of word spelling and entity extraction. The primary purpose of the normalisation process is to produce documents without historical variations on letter level. This normalised text enables better identification of entities, e.g. people, places, events, dates, as well as facilitating improved search across the collection by taking account of spelling variants of a search term. The statistical model built to automatically normalise the historical text utilised manually normalised documents. These manually normalised documents were randomly selected and accounted for approximately 6% of documents from the 1641 depositions. The Translation Model was developed on the top of the previously developed OCR correction methodology [7].

5.3 CULTURA Meta-Model and Automatic Entity Detection

Data modelling is one of the crucial aspects of designing a data-centric system. In the context of digital humanities the data modelling challenge has two specific characteristics. The first requirement is to allow the incorporation of new concepts which augment the original data during the research process e.g. detection of new type of enti-

ties such as "murder" events within the 1641 depositions, which aren't explicitly encoded a priori. The other requirement is to support a layer of services that allow a range of user actions. These actions could include the manual manipulation of existing data, the user referencing of specific data elements or the interactions between sets of users.

In both cases, the system schema must evolve over time to reflect the work of researchers and others. Therefore, the data management part of the system must support easy on-the-fly modifications of the underlying schema. This requirement rules out relational databases, since modifying the columns of a populated relational database table is a costly task. However, conceptual modelling based on the Entity-Relationship model, which is commonly used in the process of relational databases schema definition, is an effective methodology for capturing data requirements. Hence, this meta-model has been chosen by CULTURA to define the digital cultural archives using its environment e.g. the IPSA collection and the 1641 depositions.

Recently, a growing number of systems which allow schema evolution have emerged, labelled under the generic term "NoSQL". There are three prominent types of NoSQL systems: Key-Value Stores, Document Stores and Graph Databases, which greatly differ in their meta-model definitions. In Key-Value Stores only pairs of key and value are allowed, thus no structure can be defined. In Document Stores the basic element is a document which contains a set of fields. In contrast, with Graph Databases, the data is managed in the form of nodes with properties, and edges (that can be labelled) connecting the nodes. In terms of these three NoSQL types, they have variable suitability for handling Entity-Relationship data. Since Key-Value Stores contain no structure, they are clearly not suitable for implementing an Entity-Relationship model over them. In contrast, Document Stores do provide a good basis for handling entities, by mapping the entities to documents and their attributes to fields. On the other hand, Graph Stores are suited to handling entities as graph nodes and relationships as edges.

Given the features of the various NoSQL types, a combination of a Document Store (to handle entity attributes) with a Graph Store (to handle relationships), appears to be an appropriate solution for digital humanities data management. However, there are a number of drawbacks to such a solution. The first issue is the weak support in Graph Databases for non-binary relationships. It is expected that digital humanities data will include complex relationships with possibly multiple entities and relationship attributes. Thus, the graph support of current Graph Databases is not sufficient. Another insufficiency is in Document Stores support of textual search, which requires the addition of a search engine to the system. Another challenge is the need to coordinate the data between the Document Store and the Graph Database during both data ingestion and query processing.

In light of these problems, a new solution has recently been proposed, which uses the open source search library Lucene[13] to index entity-relationship data, and allows expressive search capabilities ranging from simple keyword search to complex structured queries [8]. The 1641 depositions data has already been modelled and processed

---

[13] http://lucene.apache.org/core/

as part of an entity-oriented search component [9], providing a powerful exploratory search system. This entity-oriented component provides CULTURA with a powerful tool for incremental research, where results of prior analysis are accessible for search and exploration. Furthermore, it enables the improvement of existing analysis processes, and supports new insights to the collection to be discovered. The entity-oriented approach also allows researchers and other interested users to be added as entities in the system, and link them to entities identified in the data in a non-trivial manner. Such connections between the system users and the data can afterward be used to automatically define the social network of the community interested in the digital archive, and later this social network can be used to provide social-based services to the community.

5.4   Network Analysis of Cultural Collections

Normalised historical texts can then be analysed using Influencer Network Analysis (INA), a form of Social Network Analysis (SNA), which is used to identify the social structure described in the historical texts and the influential people involved in the incidents portrayed. Influential entities not only include the individuals involved within the collection, but also the issues, topics and opinions that are detailed. Entities within the historical texts are either discovered in the text through a combined manual and automatic process, or categorised as descriptive metadata that is additional to the historical text. Relationships between entities are also identified and can augment the existing metadata from the collection. These entities and relationships are graphed and shown to users allowing a user to explore the relationships between entities across multiple historical texts (see figure 2).

Network analysis also enables a tracking of entities both temporally and geographically. Due to the scale and complexity of the information held in these content collections, this has rarely been attempted by humanities researchers. Furthermore, apart from performing network analysis on the historical texts themselves, CULTURA also analyses how the community of users engaged with the collection interact. Thus it makes it easier for users to discover relevant experts, for new communities to be created based on similar activities, and for community collaboration and contribution to occur.

6   Summary and Future Work

This paper discussed the CULTURA project and outlined key challenges that it is addressing within the digital humanities field. Two contrasting cultural collections (the 1641 depositions and the IPSA collection) that have been incorporated into CULTURA were described, along with details of the service oriented architecture underpinning the environment. Three specific features of the CULTURA environment were discussed in further detail (personalisation and adaption techniques, automatic entity detection, and network analysis), with an emphasis on the importance that metadata plays in facilitating such functionality.

Further evaluation studies within the CULTURA project will take place involving both the IPSA collection and the 1641 Depositions. Moreover, all the key stakeholders in this domain (from professional researchers to members of the public) will be involved. The outcomes of these studies will help refine the implementation and underlying methodology, especially in how the various models and metadata interact. Finally, the new technologies which have been developed within CULTURA (text normalisation service, SNA and INA services, entity detection service etc.) will be fully integrated into the architecture. This will result in an end-to-end system, which encompasses all stages from the initial normalisation of cultural heritage collections, to the deployment of these resources within an online personalised portal.

References

1. 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. In Press. October 27th-28th, (2012).
2. Agichtein, E., Brill, E., Dumais, S, 2006, "Improving Web Search Ranking by Incorporating User Behaviour Information", Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, Seattle, Washington, USA, pp19-26 , (2006)
3. Ankolekar, A. Krötzsch, M. Tran, T. Vrandečić, D, "The two cultures: Mashing up Web 2.0 and the Semantic Web", in Web Semantics: Science, Services and Agents on the World Wide Web, Volume 6, Issue 1, Elsevier Science Publishing (2008),
4. Brusilovsky, P., Kobsa, A. & Nejdl, W. (eds.). "The Adaptive Web: Methods and Strategies of Web Personalization". In The Adaptive Web, Lecture Notes in Computer Science, LNCS, vol. 4321, P. Brusilovsky, A. Kobsa, W. Nejdl (eds.), Springer-Verlag. (2007).
5. Jones, G. & Wade, V. "Integrated Content Presentation for Multilingual and Multimedia Information Access". In the Proceedings of the Workshop on New Directions in Multilingual Information Access at the 29th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 49-52, Seattle, U.S.A.. August, (2006)
6. Agosti, M., Ferro, N. "A Formal Model of Annotations of Digital Content". ACM Transactions on Information Systems (TOIS), 26 (1), 3:1-3:57 (2008).
7. Mihov, S., Mitankin, P., Gotscharek, A., Reffle, U., Schulz, K.U., and Ringlstetter C., Tuning the Selection of Correction Candidates for Garbled Tokens using Error Dictionaries, Proceedings of FSTAS 2007, pp. 25-30, (2007).
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. ACM, New York, USA, 83-92 (2012).
9. Carmel, D., Zwerdling, N., Yogev, S., Entity oriented search and exploration for cultural heritage collections: the EU cultura project. In Proceedings of the 21st international conference companion on World Wide Web. ACM, New York, USA, 227-230 (2012).