CULTURA: the main achievements

The CULTURA Virtual Research Environment]

The Virtual Research Environment (VRE) for the Digital Humanities is the central component of the CULTURA project and it supports users with different levels of experience to use a variety of tools to interact with a number of cultural heritage collections. A number of different cultural collections have been incorporated into the CULTURA Virtual Research Environment, which provides a suite of useful services, integrated into a unified portal. These services are central to enhancing a user's engagement with an archive and include faceted search tools, annotators, social network visualisation tools, and personalised recommenders. The CULTURA portal is built on top of the Drupal content management system, as it provides numerous services that, while essential to CULTURA, are not core research elements, such as user authentication and system-wide logging.


The Cultural Heritage Collections]

The concepts developed within the CULTURA project have been proved through the use of two reference cultural heritage collections: The 1641 Depositions and IPSA. These two collections are particularly relevant for scholars in the humanities, respectively for historians and historians of art, and have been chosen as case studies because they cover different and complementary aspects in the research on cultural heritage. The 1641 Depositions are composed of textual documents that required the development of tools for text processing, while IPSA is mainly a collection of digital images and metadata that required the development of tools for multimedia delivery. Yet the CULTURA approach showed to be effective for both collections, with a number of tools (annotations, visualisations, narratives, personalisation) that proved to be independent on the content and thus easily extensible to additional collections. This was proved when a third collection, 1916 Rising - a set of witness statements collected by

the Irish Bureau of Military History - was integrated into CULTURA.


Personalisation]

Cultural heritage collections often contain a large amount of resources and support a wide community of users that have varying levels of expertise. It is vital that the various types of users who interact with collections within the CULTURA portal are supported in locating content relevant to their current interests and tasks. Hence, the application of personalisation techniques by CULTURA helps empower experienced researchers, novice researchers and the wider community to discover, interrogate, and analyse cultural heritage resources.
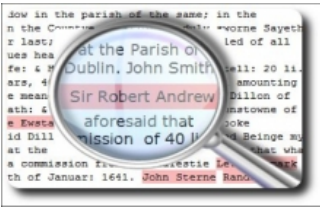

Entity Relationship Extraction]

The Entity Relationship Recognition module is a linguistic solution, developed by the IBM team, for extracting key entities from cultural resources and understanding the relationships among them. The process is done through a UIMA pipeline for annotating textual content, using IBM LanguageWare Workbench to create custom dictionaries and parsing rules. The module was applied on the 1641 Depositions to extract persons, locations, dates, and events, as well as the relationships among them and their associated attributes.
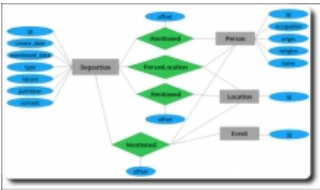

Entity Oriented Search]

EoS features Entity Relationship Data Discovery model, a data discovery extension to the classical ER conceptual model and a new logical Document Category Sets model used to represent entities and relationships within an enhanced document model. Based on this data modelling, the IBM team developed a novel approach for exploratory search over rich entity-relationship data that utilizes a unique combination of expressive yet intuitive query language, faceted search, and graph navigation.
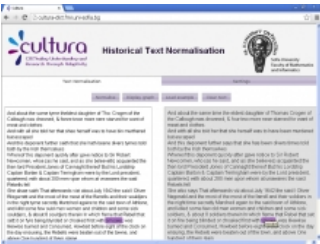
Text Normalization]

In the framework of the CULTURA project, it has been developed and implemented a new general and language independent approach to the noisy text correction problem. This issue has been addressed by designing a novel machine learning technique called REBELS and combining the correction candidates provided by REBELS with respect to their rank and some standard language features and ignores the principle complications.
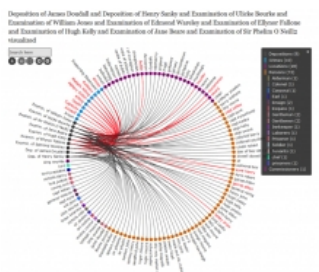
Network Visualisations in the Drupal Module]

The network visualisations in the Drupal module in the CULTURA portal visualise the entities, discovered in digital humanities content - be it the 1641 depositions, the IPSA collection or the 1916 statements collection. The networks take two forms - the Wheel and the Octopus. Both show the same data, but are organised in a different way. The Wheel orders all entities in circular form by grouping them by type and sorting them alphabetically. The Octopus follows an algorithm to reveal the structure of the network and interconnections and is very useful when the networks of multiple items (depositions, statements or works) are visualised at the same time. These interactive network visualisations help users navigate massive amounts of content in a fast and efficient manner. When reviewing the entity network of a single 1641 deposition, the user can easily access the network of a location, crime or a person within this deposition, and see other contexts where these were featured.
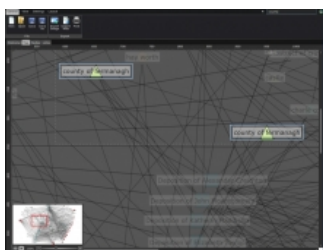

Desktop Premapper and Web Premapper]

The PreMapper is a custom-built tool for applying social network analysis to digital humanities resources. The desktop version of the software was originally planned as the only tool to edit entities and relations, add new entities, arrange custom maps and export results. The web PreMapper was developed in the second half of 2013 following the second stage user trials of the desktop version. It supports the main functionalities of the desktop application: creating new entities, selecting, editing and merging entities, adding or editing entity relations, filtering.

FAST Annotation Service]

The Flexible Annotation Semantic Tool (FAST) service covers many of the uses and applications of annotations ranging from metadata to full content; its flexible and modular architecture makes it suitable for annotating general Web resources as well as digital objects managed by different digital library systems; the annotation themselves can be complex multimedia compound objects, with varying degree of visibility which ranges from private to shared and public annotations and different access rights.
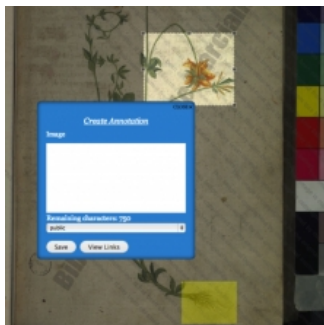
Content Annotation Tool]

CAT is a web annotation tool developed with the goal of being able to annotate multiple types of documents and assist collaboration in the field of digital humanities. At present, CAT allows for the annotation of both text and images. The current granularity for annotation of text is at the level of the letter. For image annotations, the granularity is at the level of the pixel. This allows for extremely precise document annotation, which is very relevant to the Digital Humanities domain due to the variety of different assets that prevail.
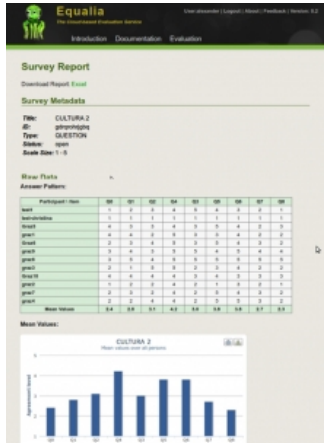
Equalia]

Evaluation is an important task in the context of digital libraries, because it reveals relevant information about the quality of the technology for all stakeholders and decision makers. The main objective of Equalia is to support the systematic and sound evaluation of digital libraries systems in line with the key phases of evaluation. For this reason, its approach is based on an evaluation model, multi-modal data collection, and automatic reports.

For further information visit http://www.cultura-strep.eu/it/outcomes.