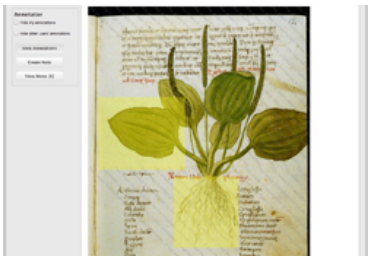
Feature Stories - Adaptive, personalised ICT to make new sense of the past



Text, illustrations, paintings and - more recently - photographs, video and audio recordings, much of them now digitised, recount many aspects of European history, from major international events to personal stories. Now, new technology is being brought to bear on these treasure troves of historical information, thanks to EU-funded researchers whose work promises to shed new light on the past.

The '1641 Depositions', held in Trinity College Dublin's library, are just one example of the many significant collections of cultural and historical heritage stored in universities, museums, archives and private collections across Europe. A rebellion by Irish Catholics in 1641 changed the course of Irish history, and also led to the creation of one of Europe's richest historical and cultural records: the 1641 Depositions, comprising 8000 witness testimonies spanning almost 20,000 pages. For decades, or in many cases centuries, researchers, students and members of the general public have scoured such collections for details about the past - a laborious and time-consuming process, fraught with pitfalls and dead-ends. Incomplete and inconsistent texts, missing words, misprints and misspellings, changes in language over time, and the sheer volume of material are just some of the challenges that need to be overcome.



One solution, being developed by a team of researchers from Austria, Bulgaria, Ireland, Israel and Italy, uses cutting-edge ICT to do much of the hard work. Supported by more than EUR 2.8 million in research funding from the European Commission, their work in the 'Cultivating understanding and research through adaptivity' (CULTURA) project is helping to quickly make sense of digitised archives, clean up inconsistencies in the language, draw links between historical events, people and objects, and make Europe's rich cultural and historical heritage more accessible to all.

'When looking at historical material a lot of information is not immediately obvious, there can be many ambiguities and inconsistencies, so what are needed are processes that can dig out that information and find those non-obvious references,' explains Dr Owen Conlan, an assistant professor in the Knowledge and Data Engineering Group at Trinity College's School of Computer Science and Statistics. 'We can then use that information to lay a path and draw connections between references that may not have been evident before.'

Dr Conlan, who is coordinating the CULTURA project, points to the example of the '1641 Depositions'. Among the many other people mentioned in the testimonies, there are repeated references to Phelim O'Neil, an Irish Catholic nobleman and rebel leader

during the uprising. But in the texts, and elsewhere, he is also known as Sir Felim O'Neill of Kinard, Phelim MacShane O'Neill or Féilim Ó Néill, or referred to simply as 'the rebel', for example:

"And he saith, that during the time he, this deponent, was so restrained and stayed amongst the rebels, he observed and well knew that the greatest part of the rebels in the county of Armagh went to besiege the Castle of Augher, where they were repulsed, and divers of the rebel O'Neils slain; in revenge whereof, the grand rebel, Sir Phelim O'Neil, knt., gave direction and warrant to one Maolmurry McDonnell, a most cruel and merciless rebel, to kill all the English and Scottish men..."

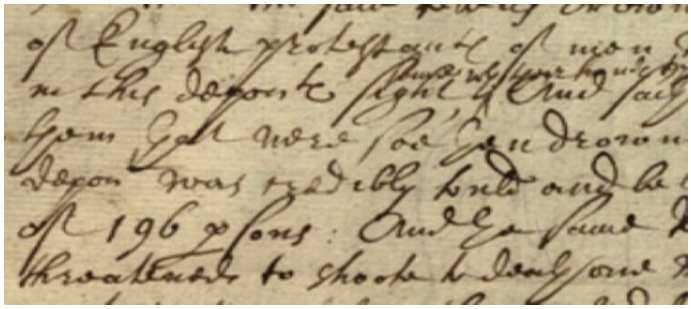
Historical social networking

To make sense of such 'noisy' historical text and begin linking references, the CULTURA team used state-of-the-art natural language processing software to 'normalise' the language and give it semantic meaning that can be understood by computers as well as humans.

'We are not altering the document and we have ensured we maintain close fidelity to the original, but our system builds another layer of information from which meaning can be extracted,' Dr Conlan says.

Powerful algorithms are employed to automatically extract entities and their relationships from the content in order to highlight the key individuals, events, dates and other entities and relationships. From there, the tools developed by the team analyse the connections between entities and relationships within the content - developing a kind of historical social network that helps place historical events and figures in context and makes them much easier to visualise and comprehend.

The approach works not only with text-based content, such as the '1641 Depositions', but also with images. In this case, metadata associated with the images, and annotated during digitisation, is used to provide semantic meaning - a process being used by the CULTURA team to analyse the Imaginum Patavinae Scientiae Archivum (IPSA) collection now held at the University of Padua in Italy. This is a digital archive of herbalists' manuscripts and illustrations, with Latin language commentaries, dating from the 14th century.



'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 processes which were prevalent during the creation of the original collection,' Dr Conlan notes. 'Using our social-network analysis, we can see, for example, who drew which illustrations, who financed them and what other illustrations they were influenced by.'

Significantly, the CULTURA system provides not just content-aware adaptivity depending on the materials being studied, but it also adapts to the needs of each user and user community. For example, a university researcher who has in-depth knowledge of a certain subject or collection of materials might use the system to look for a very specific reference. Alternatively, a member of the general public curious about a particular period of history may be looking for a much broader view.

'What we've noticed, for example, is that apprentice researchers who have used the system are going much deeper and faster with their research,' Dr Conlan notes.

Making cultural and historical heritage more accessible

The CULTURA platform can meet the needs of these and many other types of users through an innovative personalisation process that takes into account user profiles and the context in which they are searching for or accessing information. 'Widgets', integrated into the platform, make recommendations about related content that might be of interest, based in part on what was of interest to similar users. The system offers potential new paths of inquiry to follow, but ultimately leaves it up to the user to decide.

'Good personalisation is like a good storyteller. A good storyteller will arouse their audience, gauge their reactions and adjust the story as they go. But in the case of personalisation we're talking about a storyteller for just one person,' Dr Conlan says.

The system can even provide dynamic storylines around certain events, dates, places or people, generating an easy to follow narrative for any user, which adapts dynamically to the user's profile and usage history.

'Historical resources should not only be accessible to university professors and researchers, but to many different types of people, from school and university students to historical societies and interest groups and members of the general public,' Dr Conlan emphasises. 'One of the biggest challenges digital collections face is accessibility and awareness - CULTURA goes a long way towards addressing these issues.'

In addition to the '1641 Depositions' and the IPSA collection, the team has started using the CULTURA platform with a collection of historical materials related to the 1916Easter Uprising and its aftermath, another pivotal time in Irish history when Irish republicans rose up against British rule.

'The centenary of those events is coming up, so it's a very important time for Ireland. We're planning to do a lot of work with schools, especially as this material is more contemporary and more accessible,' the CULTURA coordinator says. 'In particular, we want to connect stories to real people in the documents because they're the most compelling entities, it's a way to draw users' interest into otherwise abstract events and put them into a much clearer frame of context.'

Several of the partners plan to continue supporting the platform after the end of the project with a view to expanding its use to other collections, while individual partners are looking to commercialise different parts of the technology that make up the system.

CULTURA received research funding under the European Union's Seventh Framework Programme (FP7).

Source: CORDIS website