International collaboration steers future of mass digitisation

Wednesday, July 1, 2009

By 1stAngel

Mass digitisation has become one of the most prominent issues in the library world over the last 5 years.

Feeding into the EU’s i2010 vision to significantly improve access to Europe’s cultural heritage, the British Library and the University of Salford have teamed up with a group of 15 institutions from across the continent as part of the four-year IMPACT project – IMProving ACoess to Text – to remove the barriers that stand in the way of the mass digitisation of the European cultural heritage.

Led by the National Library of the Netherlands, Koninklijke Bibliotheek, the IMPACT project aims to share expertise from across Europe and establish international best practice guidelines with a view to speeding up, standardising and enhancing the quality of mass digitisation through establishing a Centre of Competence for text based digitisation. As one of the main participants, the British Library has taken the lead on one of IMPACT’s four sub-projects, establishing the operational context of the work carried out by contributors to the project.

Mass digitisation has become one of the most prominent issues in the library world over the last 5 years, with a number of experienced libraries in Europe already scanning millions of pages each year. To help establish some standardisation over the course of the project, the British Library’s team will lead work on a set of ‘Decision Support Tools’ in an effort to focus on practical implementation support, providing guidance on digitisation workflow, the capturing of material and the organisation of metadata based on the real world experiences of project partners. These measures, announced at the first IMPACT conference in April will help ensure new material can be digitised successfully and feed into existing workflows.

Aly Conteh, e Strategy & Information Systems, Programme Manager, British Library said: “It is absolutely vital institutions like the British Library, the National Library of the Netherlands and technical experts like the University of Salford work together, sharing our experiences and resolving the challenges we face in digitising historic texts. To ensure that we deliver the digital resources that are sustainable and meet the expectations of the 21st century researcher.”
With extensive experience working with the digitisation of historic material, the British Library has also been working closely with technical experts at the internationally distinguished Pattern Recognition and Image Analysis (PRImA) research group, University of Salford, exploring methods of improving Object Character Recognition (OCR) for use in the digitisation of less standardised material. OCR technology was absolutely vital for the delivery of the Library’s recent newspaper digitisation project of 19th Century UK newspapers ([http://newspapers.bl.uk/bles](http://newspapers.bl.uk/bles)), allowing the text to be fully searchable, but the current technology has its limitations.

Dr Apostolos Antonacopoulos, Director of the PRImA research group at the University of Salford commented: “This collaboration presents a unique opportunity to make a significant world-wide impact on the digitisation of historical documents, by focusing extensive research expertise to exceptional material in both breadth and volume, such as the collections in the British Library. So far libraries and archives around the world rely on service providers whose best technologies are designed primarily for modern business documents (the service providers’ largest commercial market) and cannot take fully into account the nuances of and problems posed by ageing books and newspapers.”

Through collaboration IMPACT has already established methods for overcoming issues with geometric correction, border removal and binarisation, and is looking at examples of best practice from around the world, such as the Australian Newspaper Digitisation project’s cutting edge application of collaborative user generated corrections, to increase resource discovery success for historic mass digitisation.

Further announcements will be made about the progress of IMPACT and further public events through [www.impact-project.eu/news](http://www.impact-project.eu/news).