

IMPACT widens European scope

The European research project IMPACT (Improving Access to Text) has recently entered its second phase by taking up eleven new partners from Southern and Eastern Europe into the consortium. These new partners will contribute to the project's goals of optimising OCR (Optical Character Recognition) software and language technology for historical material and sharing institutional knowledge and expertise on digitisation. They will also help to build the IMPACT Centre of Competence, which will be launched in early 2011 to provide a central service entry point for all libraries, archives and museums involved in the digitisation of text material.

IMPACT now brings together twenty-six national and regional libraries, research institutions and commercial suppliers. The project is coordinated by the National library of the Netherlands and runs from 2008-2011. In early 2010 the EC funded a proposal to extend the project's reach, adding nearly 1 M Euro to the 15,5 M Euro original budget. In the current second phase (2010-2011) an additional four national libraries, four research centres and three universities from France, Spain, Poland, Bulgaria, Slovenia and the Czech Republic have been added. The new research partners will work on building historical lexica for their languages, while the libraries will provide datasets.

After the first two years of the IMPACT project (2008-2009), initial versions of tools have been completed, including a set of tools for efficient lexicon building. An important insight was that the work on the individual languages (Dutch, German and English) contributes to the design of these language-independent tools and that adaptations may be needed to extend the applicability of the tools to new situations. Therefore, an important aim of this extension of IMPACT is to arrive at a cross-language view of the accessibility and enhancement of digitised text.

Why IMPACT?

The IMPACT objective is to significantly improve the accessibility of historical printed text. Issues that users and institutions currently face include: Material dated before 1900 is difficult to access in a digital form, because the state-of-the-art OCR software does not provide satisfactory results for old books, magazines and newspapers; Libraries, archives and other content holding institutions across Europe lack experience and know-how in the process of digitisation and the historical language barrier also forms a stumbling block. Together this causes inefficiency and slows down the process of making European cultural heritage available on the Internet.

IMPACT tools

To overcome these barriers to digitisation, IMPACT plans to innovation in the technology for text recognition and text enrichment. Two leading industry partners, IBM and ABBYY, are involved in setting up a cutting edge text recognition system based on an adaptive model which automatically tunes itself to each new book being digitised. With the online Collaborative Correction web application linked to this system, volunteers across Europe will be able to contribute their efforts to the correction process of OCR results for further improvement. In addition, IMPACT explores new approaches in image enhancement and segmentation and in the use of language technology and historical lexica in OCR processing and Information Retrieval. The IMPACT tools will become available as interoperable web services integrated with a user friendly platform.

Sharing expertise and best practice

IMPACT will also improve the process of large-scale digitisation by sharing expertise and best practice. For this a number of strategic tools such as a website, help desk, decision support tools and a training programme are developed as well as the sustainable Centre of Competence, where the requirements of content holders across Europe and the interest of research partners inside and outside the project can be matched. The eleven new partners will demonstrate and disseminate IMPACT project results and support building capacity in digitisation in their countries, which will widen the scope of IMPACT considerably across Europe.

If you want to know more about IMPACT or subscribe to the project mailing list, please visit the website at www.impact-project.eu or contact the Project Office via impact@kb.nl. You can also follow progress on the IMPACT twitter channel: <http://twitter.com/impactocr>.

NOTES TO EDITORS:

Contact persons: Hildelies Balk (General Project Manager) at hildelies.balk@kb.nl, tel +31 70 314 0138
 Lieke Ploeger (Manager Project Office) at lieke.ploeger@kb.nl, tel. +31 70 314 0958

IMPACT Consortium – 1st phase (2008-2009)

Libraries

- National Library of the Netherlands (KB)
- The British Library (BL)
- National Library of France (BnF)
- German National Library (DNB)
- Austrian National Library (ONB)
- Bavarian State Library (BSB)
- Göttingen State and University Library (UGOE)
- University of Innsbruck Library (UIBK)

Universities & Research centres

- Institute for Dutch Lexicology (INL)
- National Centre for Scientific Research – Demokritos (NCSR)
- University of Salford (USAL)
- University of Munich (CIS group)
- University of Innsbruck (InfMath group)
- University of Bath (UKOLN)

Industry partners

- IBM (Haifa Research Lab)
- ABBYY (Moscow)

IMPACT Consortium – 2nd phase (2010-2011)

Libraries

- National Library of Spain (BNE)
- National Library of Bulgaria (NLB)
- National and University Library of Slovenia (NUK)
- National Library of the Czech Republic (NKC)
- Miguel de Cervantes Digital Library (BVC)

Universities & Research centres

- CNRS / University of Nancy (ATILF)
- University of Alicante (UA)
- Poznan Supercomputing Centre (PSNC)
- University of Warsaw (UWAR)
- Bulgarian Academy of Sciences (BAS)
- Jožef Stefan Institute (JSI)
- Charles University Prague (CUP)

