ABBYY®

CUSTOMER STORY

ABBYY® FineReader XIX

Opening up 225 years of history with ABBYY

The Fraunhofer Institute for Media Communication (IMK) opens up the archive of the Neue Zürcher Zeitung newspaper with ABBYY FineReader XIX

Thanks to ABBYY technology, the Fraunhofer Institute for Media Communication (IMK) in Sankt Augustin can tackle the digitisation of the entire archive of the Neue Zürcher Zeitung newspaper (NZZ). The IMK is using ABBYY FineReader XIX and the ABBYY FineReader Engine software development kit. The NZZ archive contains two million pages of text in various formats and fonts (e.g. roman and gothic typeface). Using ABBYY FineReader XIX these were opened up to full-text searching. FineReader XIX is an optical character recognition (OCR) software that can also process old European languages and gothic print.

"The project is a challenge on many levels. Besides the actual scale, there is also the often poor quality of the documents and the use of gothic and roman typefaces, for instance" explains Dr. Stefan Eickeler, IMK Project Manager. "We had to develop special solutions for some functions, and for text recognition we used ABBYY FineReader XIX. The program has a high recognition accuracy, the ability to process gothic print and, thanks to a software development kit, can be easily adapted and integrated into existing applications."

The documents are on index volumes and microfilms - around 1,500 rolls of 35-mm film. The quality of the microfilm documents, which are the starting point for the text recognition, varies. The photographic capturing often led to distortions. The photographic data was then transformed into image files for digitalisation. IMK developed its own software for this process which virtually eliminated the distortions and blurring. The image files form the material that ABBYY FineReader XIX uses for its text recognition.

ABBYY FineReader XIX combines all the functions of the well-known ABBYY FineReader 7.0 plus recognition of old European languages and gothic print. Using this program, the user can scan, read and digitise documents in gothic print without having to train the system. The IMK specialists have integrated the FineReader Engine SDK and FineReader XIX into their overall solution, which operates on a cluster of 20 computers. The solution then takes FineReader's recognition results and produces an XML file for each page which contains metadata on paragraph titles or other typographic features of words, for example. Each page requires around 4 megabytes of data. The entire digital archive data inventory will occupy 10 terabytes.

"We're very pleased to be working with the Fraunhofer institute," says Jupp Stoepetie, ABBYY Europe's CEO "This project shows just how far our OCR technology has evolved. Digitisation projects such as that of the IMK are only technically and financially possible thanks to our technology. The first plans to digitalise the NZZ archive were rejected a few years ago on cost grounds. But by using FineReader XIX, the entire process can be automated and made more cost effective. Were it not for the option of recognising roman and gothic print, this project would surely not have been possible."

About Fraunhofer IMK

The Fraunhofer institute for Media Communication (IMK) is an innovation and development partner for businesses, culture, education and the public sector in the field of digital media technology. The IMK is a member of the IuK group of Fraunhofer companies, a group of 17 Fraunhofer institutes that undertake research and development in the areas of information and communication technology (IuK).

www.fraunhofer.de

About Neue Zürcher Zeitung

The Neue Zürcher Zeitung can look back on more than 225 years of tradition. The first issue appeared on 12 January 1780. Today, the NZZ AG is a modern media company, publishes quality titles and is counted among the leading Swiss companies.

www.nzz.ch

About ABBYY

ABBYY is a leading developer of document recognition, document conversion, data capture and linguistics technologies.

*ABBYY's products include: **FineReader** and **PDF Transformer** – end-user applications for document conversion; **Recognition Server** – a server-based OCR and PDF conversion solution; **FlexiCapture** – data capture programs for processing forms, semi-structured and unstructured documents; **FineReader Engine SDKs** that provide a full spectrum of ABBYY's recognition technologies; and **Lingvo** – a line of dictionary software.*

More information about ABBYY at
www.ABBYY.com

www.ABBYY.com

ABBYY Europe GmbH
Elsenheimerstr. 49
80687 Munich, Germany
Tel: +49 89 51 11 59 - 0
Fax: +49 89 51 11 59 - 59
sales_eu@abbyy.com
www.ABBYY.com

ABBYY UK Ltd.
Abbey House, Grenville Place
Bracknell RG12 1BP, UK
Phone: +44 1344 392 610
Fax: +44 1344 392 611
sales_UK@abbyy.com