



AxiCom-PR, Nr. AB 02/05, März 2005

Fraunhofer-Institut für Medienkommunikation IMK erschließt das Archiv der Neuen Zürcher Zeitung mit ABBYY FineReader XIX

Mit ABBYY 225 Jahre Zeitgeschichte erschließen

München, 01. März 2005 – Mit Hilfe von ABBYY-Technologie hat das **Fraunhofer-Institut für Medienkommunikation (IMK)** in Sankt Augustin die Digitalisierung des gesamten Archivs der Neuen Zürcher Zeitung (NZZ) in Angriff genommen. Das IMK setzt hierzu ABBYY FineReader XIX und das Software Development Kits ABBYY FineReader Engine ein. Zwei Millionen Seiten liegen im NZZ-Archiv in unterschiedlichsten Formaten und Schrifttypen (z.B. Antiqua und Fraktur) vor. Diese können durch den Einsatz des ABBYY FineReader XIX ab Ende September 2005 per Volltextsuche erschlossen werden. FineReader XIX ist eine Texterkennungs-Software (Optical Character Recognition, OCR), die auch alte europäische Sprachen bzw. Frakturschrift verarbeiten kann.

„Das Projekt stellt in vielen Bereichen eine Herausforderung dar. Neben dem bloßen Umfang sind dies zum Beispiel die häufig schlechte Qualität der Vorlagen und die Verwendung von Fraktur- und Antiquaschriften“, erläutert Dr. Stefan Eickeler, Projektleiter des IMK. „Für einige Funktionen mussten wir spezielle Lösungen entwickeln, für die Texterkennung setzen wir ABBYY FineReader XIX ein. Das Programm verfügt über eine hohe Erkennungsgenauigkeit, die Fähigkeit zur Verarbeitung von Frakturschrift und lässt sich über ein Software Development Kit einfach anpassen und in bestehende Anwendungen integrieren.“

Die Vorlagen bestehen aus Registerbänden und Mikrofilmen – rund 1500 Rollen 35-Millimeter-Film. Die Qualität der Mikrofilmvorlagen, die das Ausgangsmaterial für die Texterkennung bilden, ist unterschiedlich. Bei der fotografischen Erfassung gab es häufig Verzerrungen. Für die Digitalisierung werden die fotografischen Daten in Bilddateien umgewandelt. Hierzu hat das IMK eine eigene Software entwickelt, die Verzerrungen und Unschärfen weitgehend entfernt. Die Bilddateien bilden das Ausgangsmaterial für die Texterkennung durch den ABBYY FineReader XIX. ABBYY FineReader XIX verbindet alle Funktionen der bekannten OCR Software ABBYY FineReader 7.0 mit der Erkennung alter europäischer Sprachen bzw. Frakturschrift. Mit dem Programm kann der Anwender Dokumente in Frakturschrift scannen, lesen und digitalisieren, ohne das System lange zu trainieren. Die Spezialisten des IMK haben das SDK FineReader Engine und FineReader XIX in ihre Gesamtlösung integriert, die auf einem 20 Rechner umfassenden Cluster läuft. Aus den

**P
R
E
S
S
E
M
I
T
T
E
I
L
U
N
G**

Erkennungsergebnissen des FineReader erstellt die Lösung pro Seite eine XML-Datei, die Metadaten z.B. zu Absatztiteln oder typografischen Merkmalen von Wörtern enthält. Pro Seite kommen so 4 Megabyte Daten zusammen. Der Gesamtdatenbestand des digitalen Archivs wird 10 Terabyte betragen.

„Wir freuen uns sehr über die Zusammenarbeit mit dem Fraunhofer-Institut“, erläutert Jupp Stoepetie, Geschäftsführer von ABBYY Europe. „Dieses Projekt zeigt, wie weit unsere OCR-Technologie mittlerweile entwickelt ist. Digitalisierungsprojekte wie das des IMK werden durch unsere Technologie überhaupt erst technisch und finanziell möglich. Die ersten Pläne zu einer Digitalisierung ihres Archivs musste die NZZ noch vor wenigen Jahren verwerfen, weil die Kosten zu hoch waren. Durch den Einsatz des FineReader XIX lassen sich auch komplexe Abläufe automatisieren und große Kostenvorteile erzielen. Ohne die Möglichkeit der Antiqua- und Frakturerkennung wäre das Projekt sicher nicht realisierbar gewesen.“

Profil des Fraunhofer IMK

Das Fraunhofer-Institut für Medienkommunikation IMK ist Innovations- und Entwicklungspartner für Wirtschaft, Kultur, Bildung und den öffentlichen Bereich auf dem Gebiet der digitalen Medientechnologien. Das Institut berücksichtigt die gesamte Bandbreite der digitalen Medien – von der Gestaltung über die Produktion bis zur Distribution der Inhalte – um dem Anwender ihre optimale Nutzung zu ermöglichen. Dabei spielt die intuitive Interaktion mit den Medien eine essentielle Rolle. Seit seiner Entstehung hat sich das IMK immer wieder durch Pionierleistungen zur innovativen Nutzung digitaler Medien hervorgetan – so im Bereich des Multimedia Broadcast, der Webanwendungen, der Virtual und Mixed Reality und der Verbindung von Kunst und Technologie. Die fachlichen Schwerpunkte des IMK liegen heute in den Feldern Media Information Systems, Digital Broadcast, Virtual Engineering und e-Culture; seine Kernkompetenzen sind Automatische Medienanalyse, TV-Technologien, Virtual Environments-Technologien und Design Interaktiver Systeme. Weitere Kompetenzschwerpunkte sind Web-Technologien und Netztechnologien.

In seinen internationalen, multidisziplinären Teams beschäftigt das IMK zurzeit rund 110 Mitarbeiterinnen und Mitarbeiter, darunter Wissenschaftlerinnen und Wissenschaftler der Fachrichtungen Informatik, Ingenieurwissenschaften, Physik und Mathematik sowie Medienkünstler, Designer, Architekten, Wirtschafts- und Geisteswissenschaftler aus fast zehn Ländern. Das Jahresbudget des Instituts beträgt rund 9 Mio. € Seine Erträge erwirtschaftet das IMK durch Forschungs- und Entwicklungsaufträge aus der Wirtschaft, durch die Vermarktung von Eigenentwicklungen (u.a. über Lizenzmodelle), durch die Teilnahme an öffentlichen Förderprogrammen sowie durch Beratungsaufträge und Durchführung von Kompetenzseminaren.

Das IMK ist Mitglied der IuK-Gruppe der Fraunhofer-Gesellschaft, einem Verbund der 17 Fraunhofer-Institute, die Forschung und Entwicklung auf dem Gebiet der Informations- und Kommunikationstechnik (IuK) betreiben.

ABBYY Europe GmbH

ABBYY Europe ist ein Mitglied der ABBYY Gruppe und veröffentlicht und vertreibt ABBYY-Produkte in Westeuropa. ABBYY (ABBYY Software House) entwickelt linguistische Software und Software für künstliche Intelligenz, darunter eine komplette Produktlinie für OCR und Texterkennungsprogramme und Lösungen. Das Produkt-Portfolio von ABBYY bietet: FineReader OCR Systeme – eine Produktfamilie für Endkunden und SDKs zur Erkennung von gedruckten Texten, Tabellen und Formularen; FormReader – ein ICR Programm zur Erkennung und Verarbeitung von Formularen, die per Hand ausgefüllt wurden; ABBYY Retrieval & Morphology Engine – Tools für Entwickler zur Integration von Volltextsuche und linguistischen Funktionen in externe Applikationen. Zu den Lizenznehmern der Erkennungs-Technologien von ABBYY zählen Siemens Nixdorf, Samsung Electronics, C-Technologies, Sumitomo Electric Systems, Banctec, BenQ, Documentum, EPSON, Freedom Scientific, Hewlett-Packard, Kurzweil, Kofax, Microtek, NewSoft, Notable Solutions, Panasonic, Stellant, Toshiba, Verity, Neurascript, Saperion und SWT. Die Zentrale der ABBYY Gruppe befindet sich in Moskau, Büros gibt es in der Ukraine (ABBYY Ukraine), den USA (ABBYY USA, Fremont, CA), Großbritannien (ABBYY UK, Bishops Stortford, England) und Deutschland (ABBYY Europe, München). Weitere Informationen über ABBYY finden sich im Web unter www.abbyy.com

oder direkt bei ABBYY Europe GmbH, Anglerstraße 6, 80339 München, E-Mail:
sales@abbyeu.com.

Weitere Informationen:

ABBYY Europe GmbH
Gino Boscia
Anglerstraße 6
D - 80339 München
Tel.: 089/511 159-17
Fax: 089/511 159-59
E-Mail: press@abbyeu.com

Pressekontakt:

AxiCom GmbH
Marcus Birke
Junkersstraße 1
D - 82178 Puchheim
Tel.: 089/800908-26
Fax: 089/800908-10
E-Mail: marcus.birke@axicom.de
Web: www.axicom.de

Kontakt Fraunhofer-Institut für Medienkommunikation IMK:

Dr. Stefan Eickeler
Schloss Birlinghoven
D - 53754 Sankt Augustin
Tel: 02241/14-1969
Fax: 02241/14-2597
E-Mail: stefan.eickeler@imk.fraunhofer.de
Web: www.imk.fraunhofer.de