

Universität Innsbruck: ABBYY Fine Reader Scripting Edition als Universal-Tool

Universität Innsbruck Abteilung für Digitalisierung und elektronische Archivierung - DEA

Die Universität Innsbruck unterhält eine eigene Abteilung für Digitalisierung und elektronische Archivierung. Die Abteilung unter Leitung Dr. Günter Mühlbergers ist auch als Dienstleister aktiv und hat bereits zahlreiche Projekte für externe Organisationen realisiert. Die DEA bietet die Digitalisierung von Büchern, Zeitschriften, Karteikarten, Manuskripten und anderen Materialien. So wurde bereits jedes Buch der Universitätsbibliothek Innsbruck - sofern es nicht mehr regulär im Buchhandel vertrieben wird – als digitales e-Book zur Verfügung gestellt. Darüber hinaus bietet die DEA die komplette Konvertierung von Zettelkatalogen: Scannen, Texterkennung, Indexierung mittels Oracle Datenbank und Hosting des Katalogs auf den Rechnern des Zentralen Informatik Instituts der Universität Innsbruck.



**Universität Innsbruck
 Digitalisierungsprojekte**
 Dr. Günter Mühlberger und
 Dr. Kurt Habitzel
 Innrain 52
 A-6020 Innsbruck
 Tel +43/+512/507/9053
 Fax +43/+512/5072607

Die Universität Innsbruck unterhält eine eigene Abteilung für Digitalisierung und elektronische Archivierung. Die Abteilung unter Leitung Dr. Günter Mühlbergers ist auch als Dienstleister aktiv und hat bereits zahlreiche Projekte für externe Organisationen realisiert. Für die Implementierung von OCR-Prozessen setzen die Spezialisten der Universität Innsbruck auf ABBYY-Technologie. Die Zusammenarbeit besteht seit vier Jahren und begann durch eine Partnerschaft im EU Projekt METADATA ENGINE. Das Projekt wurde von der Universität Innsbruck geleitet und im Rahmen des Projekts wurde auch die Basisentwicklung von FineReader XIX durchgeführt.
 (Link: <http://meta-e.aib.uni-linz.ac.at/>).

Zur Zeit setzt die Abteilung für Digitalisierung und elektronische Archivierung ABBYY FineReader Scripting Edition in den Versionen 6.0 und 7.0 ein, deren Funktionen sie in den unterschiedlichsten Bereichen nutzt. Die SDK Version 7.1 wird derzeit gerade getestet. Drei Projekte sind in diesem Zusammenhang besonders hervorzuheben:

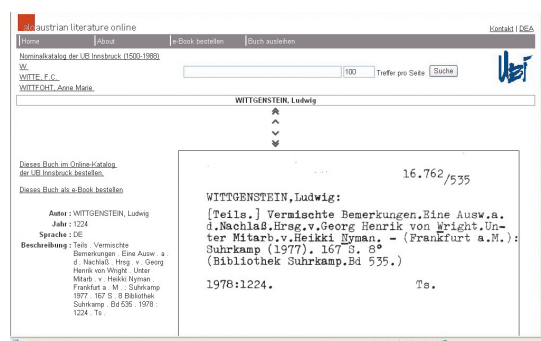
1. Die Digitalisierung des kompletten Karteikartenkatalogs der Universitätsbibliothek mit 1,1 Millionen Karteikarten im Format A6.
2. Die Bereitstellung alter Werke in Frakturschrift als eBook mit der Möglichkeit zur Volltextsuche.
3. Die Digitalisierung eines großen Zeitungsausschnittsarchivs mit rund einer 1 Millionen Artikel. Das digitale Archiv wird sowohl im Volltext als auch mittels Bildsuche erschlossen.

1. Digitalisierung des Katalogs

Mit der Digitalisierung des Zettelkatalogs, der von 1500 bis 1988 reicht und mehr als 1,1 Millionen Karteikarten umfasst, hatte sich die Abteilung für Digitalisierung und elektronische Archivierung etwas Besonderes vorgenommen. Für die Online-Nutzung sollte der Katalog die Funktionsweise der traditionellen Karteikästen möglichst exakt nachbilden, d.h. der gesamte Inhalt des Katalogs sollte zwar per Volltextsuche zu erschließen sein, aber die ursprüngliche Struktur sollte erhalten bleiben: Anwender sollten wie gewohnt im Katalog blättern und zudem die Ergebnisse der Digitalisierung am Original kontrollieren bzw. ergänzen können.

Aufgrund begrenzter Budgets kamen nur zwei Vorgehensweisen in Frage: Die automatische Digitalisierung per OCR-Software oder manuelles Abtippen in einem Billiglohnland. Zwei Argumente gaben letztlich den Ausschlag für den Einsatz der ABBYY OCR-Software: Die hohe Präzision und Zuverlässigkeit des Systems und die Kosten. „Wir gehen davon aus, dass wir durch den Einsatz der ABBYY-Software Budget in der Größenordnung von 1:10 gegenüber einer Auslagerung in ein Billiglohnland gespart haben“, schätzt Mühlberger und ergänzt: „Wir haben das Projekt im März 2004 abgeschlossen und die Ergebnisse geben uns Recht. Wir haben die Suchresultate in unserem neuen Online-Katalog in Stichproben untersucht und festgestellt, dass lediglich 15% der Karteikarten nicht auf Anhieb zu finden sind. Mit etwas Spielerei bei der Formulierung der Suchanfrage und durch die Möglichkeit, im Katalog wie in einem herkömmlichen Karteikartenkatalog zu blättern, erzielt man aber sehr gute Ergebnisse.“

Der Online-Katalog ist unter www2.uibk.ac.at/ub/dea allgemein zugänglich und liefert die OCR-Ergebnisse Seite an Seite neben einer gescannten Karteikarte:



Mit den Cursor-Feldern oberhalb der Karteikarte kann man sowohl wie in einem physischen Kasten blättern als auch in den Suchergebnissen vor und zurück springen.

Zur Zeit verzeichnet der Katalog täglich ca. 1500 Abfragen, was eine deutliche Verbesserung gegenüber den früheren Werten bedeutet. Die Universität und ihre Studierenden können somit einen weitaus umfangreicheren Gebrauch der vorhandenen Ressourcen machen. Die Altbestände wurden vor dem Versinken in den Tiefschlaf gerettet.

Auch in anderer Hinsicht profitiert die Universität Innsbruck von dem Projekt: Aufgrund des großen Erfolges bietet die Abteilung für Digitalisierung und elektronische Archivierung den Aufbau eines elektronischen Archivs externen Anwendern als Dienstleistung an. Derzeit wird mit Interessenten aus Österreich und Deutschland über weitere Projekte verhandelt.

2. Alte Werke in Frakturschrift als eBook

Neuland betreten ABBYY und die Universität Innsbruck bei der Erstellung elektronischer Bücher (eBooks) aus Vorlagen in Frakturschrift. Bis vor kurzem existierte keine Lösung für die automatische Erkennung von Frakturschrift. Der Bedarf an solcher Software war allerdings groß. Deshalb entwickelte ABBYY im Rahmen des METADATA ENGINE Projekts zusammen mit zahlreichen Universitäten und Bibliotheken eine OCR-Standardanwendung für Frakturschrift und brachte 2004 ein entsprechendes Zusatzmodul für die ABBYY FineReader Scripting Edition heraus, das in der Abteilung für Digitalisierung und elektronische Archivierung der Universität Innsbruck seinen Dienst verrichtet. Mittlerweile ist die Fähigkeit der Frakturschrifterkennung auch in Form des ABBYY FineReader XIX als separates Produkt erhältlich. Ein Erfolg, der maßgeblich auch der engen Zusammenarbeit zwischen ABBYY und der Abteilung für Digitalisierung und elektronische Archivierung unter der Leitung Mühlbergers zu verdanken ist. Die Software verarbeitet Texte in Deutsch, Englisch, Französisch, Italienisch und Spanisch. FineReader XIX erkennt Frakturschrift, Schwabacher und viele Textschriften (Gotische Schriften) und romanische Schriftzeichen, die nicht mehr verwendet werden, wie z.B. das lange „s“.

Die Universitätsbibliothek Innsbruck nutzt die Frakturerkennung derzeit vor allem in ihrem e-Book-On-Demand Service. Nutzer können die Bibliothek online beauftragen, ein eBook von einem Werk aus dem Katalog zu erstellen. Sie erhalten dann eine elektronische Version des Buches, die nicht nur seine grafische Reproduktion bietet, sondern auch die Möglichkeit zur Volltextsuche eröffnet. Die PDF-Datei bietet an der Oberfläche ein Faksimile des Originals und die Volltextsuche arbeitet im Hintergrund mit den Erkennungsergebnissen des FineReader. Selbst alte Texte lassen sich so mit modernen Recherchemethoden erschließen: Forscher sind in der Lage, Referenzstellen leichter aufzufinden und neue Bezüge herzustellen. Zudem können sie sich auf ihren Rechnern ihre ganz persönliche digitale Bibliothek anlegen. Ganz abgesehen von diesen Vorteilen bietet der elektronische Versand der eBooks auch die Möglichkeit, Bücher weltweit verfügbar zu machen, die per Fernleihe aufgrund ihres Alters nicht mehr zur Verfügung stehen. „Unsere Nutzer sind begeistert von den Möglichkeiten der eBooks“, freut sich Mühlberger. „Viele Universitäten können eBooks nicht einmal für aktuelle Werke anbieten. Dank ABBYY sind wir in der Lage, alle Werke unserer Bibliothek – und das sind mehr als 1,5 Millionen – in elektronischer Form zur Verfügung zu stellen, sofern sie nicht mehr im Handel erhältlich sind. Das ist meines Wissens zur Zeit weltweit einzigartig.“

3. Volltext- und Bildsuche im Zeitungsarchiv

Als dritte Anwendung der ABBYY OCR-Technologie baut die Abteilung für Digitalisierung und elektronische Archivierung der Universität Innsbruck zur Zeit ein digitales Archiv mit einer kombinierten Volltext- und Bildsuche für das Zeitungsarchiv der Universitätsbibliothek auf. Insgesamt sollen bis zur offiziellen Verfügbarkeit des Dienstes ca. eine Millionen Zeitungsartikel eingepflegt werden. Diese Artikel dürfen aus urheberrechtlichen Gründen nicht komplett bereitgestellt werden, sind dann aber sowohl über die Volltextsuche als auch über die Bildsuche verfügbar. Anwender können Bilder dann über Suchbegriffe finden; das Prinzip ist dabei dasselbe wie bei der Suchmaschine Google. Zu jedem gefundenen Ausschnitt bzw. Bild wird zusätzlich die Titelregion des Originalabdrucks angezeigt, inklusive eventuell vorhandener handschriftlicher Anmerkungen z. B. Ausgabe der Zeitung und zum Datum. Die erste Phase des Projektes läuft äußerst viel versprechend und der Dienst soll bereits im Frühjahr 2005 online sein. Auch ein solches digitales Archiv, das zudem online weltweit zur Verfügung steht, ist ein Novum in der Erschließung wichtiger Zeitzeugnisse und Dokumente.

In Anbetracht der vielfältigen Einsatzmöglichkeiten der ABBYY FineReader Scripting Edition und der erfolgreich realisierten umfangreichen Projekte seit Beginn der intensiven Zusammenarbeit 2003 wundert es nicht, dass Mühlberger ein positives Fazit zieht und weitere Einsatzbereiche erschließen möchte: „Zu Beginn haben wir uns für eine OCR-Anwendung entschlossen, um Kosten zu sparen. Überzeugt hat uns ABBYY dann mit leichter Integration, Anpassungsfähigkeit und hoher Stabilität. Der Support war exzellent und die Erkennungsquote auch bei schwierigen Vorlagen überzeugt. Außerdem bietet nur ABBYY die Frakturerkennung. Zu Beginn hätten wir uns nicht träumen lassen, wie umfangreich unser Einsatz der Software einmal sein würde, und dass wir damit als externer Dienstleister zusätzliches Geschäft würden generieren können. Da auch das Feedback unserer Anwender sehr positiv ist, werden wir weitere Bestände elektronisch erschließen und online verfügbar machen. Für uns hat sich die ABBYY FineReader Scripting Edition zu einem Universal-Tool der Digitalisierung entwickelt.“

Mehr Information über ABBYY finden Sie unter www.ABBYY.com



Westeuropa:
ABBYY Europe GmbH
Anglerstr. 6, 80339 München
Deutschland
Tel.: +49-(0)89-511159-0
Fax: +49-(0)89-511159-59
E-mail: info@abbyeu.com

Großbritannien:
ABBYY Europe GmbH UK Office
19 Church Street, Bishops
Stortford, CM23 2LY, England
Tel.: +44-(0)1279-656539
Fax: +44-(0)1279-465574
E-mail: sales@abbyy.co.uk

Nord-Amerika:
ABBYY USA
47221 Fremont Boulevard,
Fremont, CA 94538, USA
Tel.: +1-510-2266717
Fax: +1-510-2266069
E-mail: sales@abbyyusa.com

Osteuropa:
ABBYY Ukraine
P.O. Box 23, 02002 Kiev,
Ukraine
Tel.: +380-44-4909999
Fax: +380-44-4909461
E-mail: sales@abbyy.com.ua

Alle anderen Länder:
ABBYY Software House
P.O. Box 54, Moskau, 129301
Russland
Tel.: +7-095-783-3700
Fax: +7-095-783-2663
E-mail: sales@abbyy.com