

ABBYY® FineReader



ABBYY hilft, den Zugang zu historischen Aufzeichnungen zu verbessern und das Geschichtsbewusstsein zu schärfen

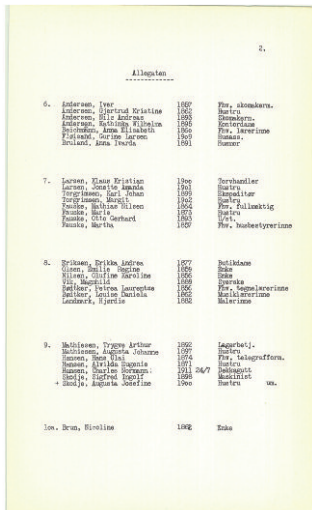
FineReader reduziert den Arbeitsaufwand und beschleunigt Projekt einer norwegischen Organisation

DIS-Hordaland in Bergen ist eine Abteilung der Genealogiegesellschaft Norwegens (DIS-Norge) und darauf spezialisiert, Ahnenforschern weltweit Datenbanken zur Verfügung zu stellen, schwerpunktmäßig mit Informationen über Menschen, die in Hordaland gelebt haben. Ihr Ziel ist es nicht nur, professionellen und privaten Genealogen effiziente Datenbanken bereitzustellen, sondern auch das Interesse der Norweger an der Stammbaumforschung zu wecken.

Die Herausforderung

Der freie Zugriff auf Datenbanken mit persönlichen Daten hat in Norwegen eine lange Tradition. Um diese Datenbanken mit Informationen zu füllen, haben ehrenamtliche DIS-Norge-Mitarbeiter jahrelang die Daten aus Gemeindeaufzeichnungen, Volksbefragungen und anderen öffentlichen Quellen manuell eingegeben. DIS-Hordaland wurde vor Kurzem beauftragt, eine Datenbank über alle Menschen aus Bergen zu erstellen, die zum Zeitpunkt der Regionalwahlen 1934 21 Jahre oder älter waren. Die Datenbank sollte Informationen – wie Name, Adresse und Geburtsdatum – von allen Bürgern enthalten, die dieses Kriterium erfüllten.

Da die meisten Dokumente mit den relevanten Daten in Maschinschrift oder in gedruckter Version vorlagen, stellte DIS-Hordaland fest, dass durch den Einsatz einer Texterkennungssoftware (OCR) zur Umwandlung der Dokumente in editierbaren Text die ehrenamtlichen Mitarbeiter viel Zeit und Ressourcen einsparen würden. Wie Tore J. Mehl, Mitglied der DIS-Hordaland-Geschäftsleitung erklärt: „Was, wenn wir die Daten mit Hilfe moderner Werkzeuge wie einem OCR-Programm konvertieren könnten? Wir könnten dann ganze Akten von Quelldokumenten mit weit weniger Mühe und in kürzerer Zeit als durch manuelle Verarbeitung veröffentlichen.“



Originaldokument (in Papierform)



OCR-Ergebnisse

Über DIS-Hordaland

DIS-Norge (Norwegische Genealogiegesellschaft) ist der größte Ahnenforscher-Verband des Landes mit fast 9000 Mitgliedern. Das Anliegen der Gesellschaft ist es, ein landesweites Forum für Familien- und Lebensgeschichten zu etablieren, das Datenverarbeitung und das Internet als Hilfsmittel einsetzt.

Kontakt

www.disnorge.no

ABBYY® FineReader

Die Organisation formulierte einige Anforderungen: Zunächst sollte die Software komplexe Layouts in der digitalen Version beibehalten, wie bei farbigen, gescannten Schreibmaschinen-seiten in DIN A4-Format mit unterschiedlichen Schriftarten. Weiter war es wichtig, dass das Programm Tabellen erkennen würde, da zahlreiche Quellinformationen in dieser Form vorlagen. Schließlich sollte die gesuchte Software editierbare Tabellenseiten erstellen, als CSV-Dateien ausgeben und direkt in die Datenbank übertragen können.

Die Lösung

DIS-Hordaland untersuchte und testete eine Reihe von OCR-Programmen. Nach reiflicher Überlegung und überzeugt durch die vielfache Leistung des Produktes in den Bestenlisten der OCR-Programme entschied sich Tore J. Mehl, eine Lizenz von ABBYY FineReader Professional Edition zu erwerben, um die Funktionen des Programms kennenzulernen. Nach weiteren Untersuchungen und Tests der Software-Funktionalitäten wählte DIS-Hordaland schließlich ABBYY FineReader als die geeignetste Lösung für ihre Anforderungen aus.

„Das letzte Argument, das uns davon überzeugte, dass ABBYY FineReader das richtige Produkt ist, war die Zuverlässigkeit in der Umwandlung von Tabellen – was für uns geschäftsentscheidend ist, da Informationen in Tabellenform so häufig in den Quellen vorkommen. Kein anderes Tool bot eine mit ABBYY vergleichbare Tabellenverarbeitungs-funktion“, erklärt Tore J. Mehl. „Ein weiterer Vorteil war die einfache Bedienung der Software. Wir benötigten ein Programm, das unser Team ohne großes Training sofort einsetzen konnte.“

Die ehrenamtlichen Mitarbeiter des Projekts belegten einen kurzen Einführungskurs zur Handhabung von ABBYY FineReader und zur Erläuterung weiterer wichtiger Aspekte des Projekts. Sie waren mit den Ergebnissen mehr als zufrieden: Die Erkennungstechnologie in ABBYY FineReader ermöglichte es ihnen, die auf Papier und in Bildern enthaltene Information auszulesen und machte das zeitaufwendige Abtippen und manuelle Formatieren so gut wie überflüssig. Der Ablauf gestaltet sich für die Mitarbeiter recht einfach: Papierdokumente werden gescannt und in ABBYY FineReader geladen. Die Bilder werden in Text umgewandelt und, gemäß DIS-Hordalands Anforderungen, im Microsoft® Excel® -Format ausgegeben. Die Tabelle wird dann in eine Datenbank exportiert und veröffentlicht, sodass sie privaten Nutzern und Forschern auf der ganzen Welt online zur Verfügung steht.

Das Ergebnis

Dank ABBYY FineReader sparte DIS-Hordaland signifikant Zeit und Ressourcen ein. Das ist besonders wertvoll angesichts der Tatsache, dass alle Projekte der Organisation von ihren Mitarbeitern ehrenamtlich und nebenberuflich durchgeführt werden. Die Umwandlung von 10 Seiten durch OCR-Verarbeitung von maschinengeschriebenem Text mit ABBYY FineReader, die zusätzliche Korrektur und die Verarbeitung in Excel dauerten ca. eine Stunde und zwanzig Minuten. Unter normalen Umständen hätte ein ähnlicher Prozess bis zu vier Stunden gedauert und wäre durch das manuelle Abtippen sehr fehleranfällig. DIS-Hordaland schätzt, dass während dieses Projekts insgesamt ungefähr 400 Stunden manueller Arbeit eingespart wurden.

„Die Datenbank konnte nur mithilfe von ABBYY FineReader so schnell und effizient aufgebaut werden. Statt neun bis zwölf Monaten manueller Arbeit schlossen wir das Projekt mit ABBYY FineReader in nur drei Monaten ab“, bekräftigt Tore J. Mehl. „Die Online-Datenbank ist ein großer Erfolg. In den ersten zwei Monaten nach der Veröffentlichung auf der Website verzeichneten wir 44.000 Suchanfragen in der Datenbank. Das sind durchschnittlich ca. 370 Anfragen täglich. Und das Interesse an unserer Datenbank wächst beständig.“

Über ABBYY

ABBYY ist ein führendes Unternehmen in der Entwicklung von Technologien für Dokumenterkennung, Dokumentumwandlung, Data Capture und Linguistik. Zum Produktportfolio von ABBYY gehören: **FineReader OCR** und **PDF Transformer** – Endanwenderprogramme zur Umwandlung von Dokumenten; **Recognition Server** – eine serverbasierte Lösung für OCR und PDF-Umwandlung; **FlexiCapture** – Data Capture Lösung zur Verarbeitung von Formularen, semi- und unstrukturierten Dokumenten; **FineReader Engine SDKs** mit dem gesamten Leistungsumfang der ABBYY OCR-Technologien; **Lingvo** – eine Serie von elektronischen Wörterbüchern.

Mehr Informationen über ABBYY unter www.ABBYY.com