

## Digitale Zeitreisen

Universitätsbibliothek Frankfurt am Main macht mit ABBYY historische Dokumente online verfügbar

Die Universitätsbibliothek Frankfurt am Main zählt mit ihren umfangreichen Beständen und Sammlungen von derzeit etwa 9 Millionen Medieneinheiten zu den zentralen wissenschaftlichen Bibliotheken in Deutschland. Um den Zugriff auf bedeutendes historisches Textmaterial zu erleichtern, baut die Universitätsbibliothek das Web-Angebot aus und stellt Informationen über vorhandene Medien sowie umfangreiche Datenbanken digitalisierter historischer Dokumente online zur Verfügung. Um die digitalen Textdokumente für die Recherche zu optimieren, nutzt die Bibliothek ABBYY FineReader Engine zur Texterkennung, sodass auch in Fraktur gedruckte Texte über Stichwörter gefunden und per Volltextsuche nach einzelnen Wörtern durchsucht werden können.

### Herausforderung

Die Universitätsbibliothek Frankfurt am Main besitzt als besonders wertvolle Dokumentensammlung die historisch einmalige Rothschild-Sammlung aus der „Freiherrlich Carl von Rothschild'schen öffentlichen Bibliothek“ von 1888. Sie umfasst zwei Sammlungen historischer Zeitungsausschnitte: zum einen rund 20.000 Artikel der nationalen und internationalen Presse aus den Jahren 1886-1926, die sich auf die Familie Rothschild und das Bankhaus beziehen, zum anderen Beiträge zur Geschichte der Rothschild-Bibliothek aus der Zeit von 1890-1928. Die Texte werden im Rahmen des Projektes „Judaica Europeana“ online bereitgestellt, dessen Ziel unter anderem ist, die Zeugnisse jüdischer Kultur in Europa zu dokumentieren, als digitale Versionen zur Verfügung zu stellen und zur Verbreitung der Dokumente in Forschung und Lehre beizutragen. Für das Judaica-Portal der Universitätsbibliothek Frankfurt am Main wurden bisher bereits über 10.000 Titel online gestellt, die meisten bisher allerdings ohne Texterkennung.

„Bei der digitalen Bereitstellung von Textdokumenten ist es wesentlich, dass die Dateien per Volltextsuche durchsucht werden können, sodass innerhalb der Zeitungsartikel gezielt nach einzelnen Wörtern und Begriffen gesucht werden kann, damit sie bei einer wissenschaftlichen Recherche nicht übersehen werden“, erklärt Rachel Heuberger, die Leiterin der Hebraica- und Judaica-Sammlung der Universitätsbibliothek Frankfurt. Dafür müssen die digitalen Dokumente durch eine Texterkennungssoftware (OCR) in ein durchsuchbares Dateiformat umgewandelt werden. Die Textdokumente der Rothschild-Sammlung stellen besondere Herausforderungen an die OCR-Verarbeitung, weil sie in zahlreichen verschiedenen Drucktypen, Bildqualitäten und Schriftarten – von Fraktur bis Antiqua – vorliegen. Daher mussten bei der Auswahl der OCR-Lösung besonders hohe Anforderungen in Bezug auf die Erkennungsqualität unterschiedlicher Schriftarten gestellt werden. Ein weiterer wichtiger Punkt bei der Auswahl der OCR-Software war die Möglichkeit, Dokumente nach Stichwörtern zu durchsuchen, was essentiell für die wissenschaftliche Recherche und Arbeit mit den Texten ist.



**Name:** Universitätsbibliothek JCS Frankfurt am Main

**Hauptsitz:** Frankfurt am Main, Deutschland

**Branche:** Bildungswesen, Forschung

**Produkte und Services:** Bibliotheksdienstleistungen

**Mitarbeiter:** 270

**Web:**

[www.ub.uni-frankfurt.de](http://www.ub.uni-frankfurt.de)

**Projektseite:** [sammlungen.ub.uni-frankfurt.de/rothschild/nav/index/all](http://sammlungen.ub.uni-frankfurt.de/rothschild/nav/index/all)

**Projekt realisiert mit:**

**semantics**  
Kommunikationsmanagement GmbH  
[www.semantics.de](http://www.semantics.de)

### PROJEKTÜBERBLICK

**Herausforderung**

Optimierung umfassender digitaler Textsammlungen in verschiedenen Schriftarten einschließlich Fraktur für eine Volltextrecherche bei Bereitstellung online

**Projektverlauf**

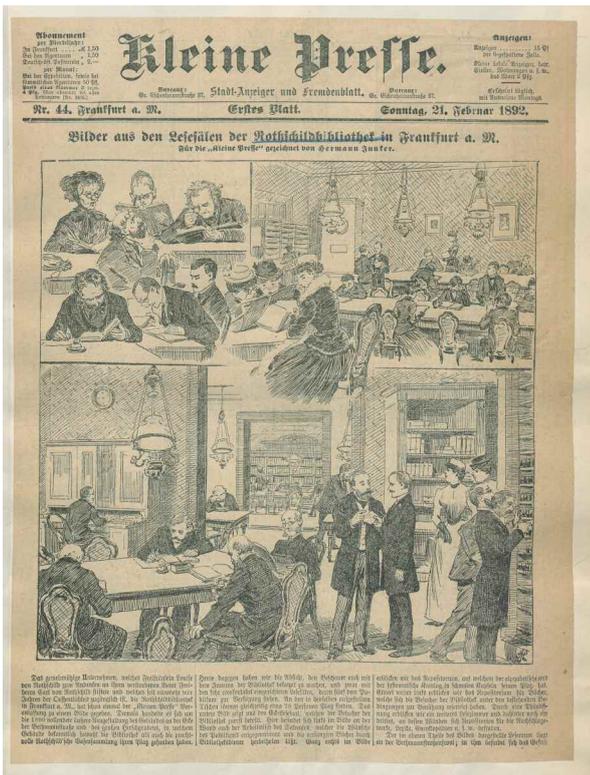
Volltexterkennung (OCR) der digitalen Textdokumente und Umwandlung in durchsuchbare PDFs mit ABBYY FineReader Engine

**Ergebnis**

Verfügbarkeit und vollständige Durchsuchbarkeit von mehreren 10.000 historischen Textdokumenten der Rothschild-Sammlung online

## Lösung

Die Universitätsbibliothek Frankfurt am Main nutzte ABBYY FineReader Engine zur Volltexterkennung der Dokumentenbilder, um die wertvollen Textdokumente für die wissenschaftliche Recherche zu optimieren. „ABBYY FineReader Engine konnte sowohl in Bezug auf die Erkennungsqualität bei unterschiedlichen Schriftarten und Bildqualitäten als auch in Bezug auf die Durchsuchbarkeit der Ergebnisdokumente überzeugen“, freut sich Rachel Heuberger. Die Software war der Universitätsbibliothek Frankfurt am Main schon durch das ähnliche Digitalisierungsprojekt „Compact Memory“ aus den Jahren 2000-2006



bekannt, bei dem allerdings nur ein Teil der online veröffentlichten Sammlung von Zeitschriften per Volltexterkennung erschlossen wurde. Das ABBYY OCR-SDK war daher bereits in die vorhandene Bibliotheks-Softwarelösung Visual Library der Firma Semantics integriert und wurde nach kurzer Evaluierung ohne Testphase direkt für das Projekt eingesetzt. Die historischen Dokumente, die bereits digital im Tiff-Format mit einer Auflösung von 300-400 dpi in Farbe vorlagen, wurden vollständig OCR-erkannt. Die OCR-Ergebnisse wurden ohne Verifizierung als Volltexte einschließlich der Zeichenkoordinaten des erkannten Textes für das „Highlighting“ und die PDF-Generierung direkt an die Visual Library übergeben. Die optimierten Dokumente können dann online zur Verfügung gestellt werden und sind über die Suchfunktion jederzeit schnell und einfach wieder auffindbar.

## Ergebnis

Nach einer Phase der Adaption lief die OCR-Erkennung reibungslos zur vollsten Zufriedenheit des Bibliothekspersonals. In kurzer Zeit konnten 22.000 Einzelseiten bearbeitet und im Netz freigeschaltet werden. Pünktlich zum 125. Jahrestag der Gründung der „Freiherrlich Carl von Rothschild'schen öffentlichen Bibliothek“ im Januar 2013 standen die voll erschlossenen Dokumente der Rothschild-Sammlung den Nutzern online zur Verfügung.

„ABBYY FineReader Engine konnte sowohl in Bezug auf die Erkennungsqualität bei unterschiedlichen Schriftarten und Bildqualitäten als auch in Bezug auf die Durchsuchbarkeit der Ergebnisdokumente überzeugen.“

Dr. Rachel Heuberger, Leitung Hebraica- und Judaica-Sammlung Universitätsbibliothek Frankfurt

„Wir freuen uns, damit auch heute noch dem Gründungsauftrag der Rothschild-Bibliothek treu bleiben zu können und allen Interessierten kostenlos den Zugang zu wichtigen Texten und damit zur ‚ernsthaften Bildung und wissenschaftlicher Belehrung‘ (wie es in einem zeitgenössischen Text heißt) zu ermöglichen“, sagt Rachel Heuberger, die Leiterin der Hebraica- und Judaica-Sammlung der Universitätsbibliothek Frankfurt.

Der große Erfolg des Projektes schlägt sich nicht nur in einem äußerst positiven Presseecho auch überregionaler Zeitungen nieder, sondern wird besonders durch die überaus große Nachfrage durch Nutzer des Online-Angebots bestätigt: Die digitalen Judaica-Sammlungen werden um die 1 Millionen Mal pro Monat aufgerufen. Nach dem großen Erfolg des Digitalisierungsprojektes hat die Universitätsbibliothek großes Interesse, weitere digitale Sammlungen mit OCR zu erschließen, um die digitale Recherche in ihren Beständen weiter zu verbessern und damit eine stärkere Nutzung der digitalen Quellen zu erreichen.

## ÜBER ABBYY

ABBYY ist ein führender Anbieter von Technologien und Services für Texterkennung und Dokumentenumwandlung. Das umfangreiche Produktportfolio für Dokumentenverarbeitung und Informationserfassung ist verfügbar für unterschiedliche Plattformen und Geräte. ABBYY bietet eine große Bandbreite von Lösungen für spezifische Geschäfts- und Branchenbedürfnisse. Unternehmen in der ganzen Welt vertrauen auf ABBYY-Lösungen zur Optimierung papierbasierter Geschäftsprozesse. Mehr Informationen unter [www.ABBYY.com](http://www.ABBYY.com)



**ABBYY Europe GmbH**  
Elsenheimerstr. 49,  
80687 München, Deutschland  
Tel: +49 89 511 159 0  
[sales\\_eu@abby.com](mailto:sales_eu@abby.com)  
[www.ABBYY.com](http://www.ABBYY.com)

**ABBYY UK**  
[sales\\_UK@abby.com](mailto:sales_UK@abby.com)

**ABBYY France**  
[sales\\_france@abby.com](mailto:sales_france@abby.com)

**ABBYY Spain**  
[sales\\_es@abby.com](mailto:sales_es@abby.com)

**ABBYY Italy**  
[sales\\_it@abby.com](mailto:sales_it@abby.com)

**ABBYY Benelux**  
[sales\\_benelux@abby.com](mailto:sales_benelux@abby.com)

**ABBYY Scandinavia**  
[sales\\_nordic@abby.com](mailto:sales_nordic@abby.com)