



Customer Overview

Name

Genealogy Indexer

Industry

Not For Profit

Web

<http://genealogyindexer.org/>

The Genealogy Indexer website enables users to make full-text searches of over one million pages of historical records. But their data must first be converted into searchable digital files originating from paper documents that are often of poor quality, hundreds to thousands of pages long and in hard-to-recognize typefaces. A task made possible with sophisticated, accurate and automatic Optical Character Recognition (OCR) from ABBYY Recognition Server.

“Without Recognition Server, I would simply not be able to do any of this. No other solution I have tested comes close to delivering acceptable accuracy.”

— Logan Kleinwaks, Founder, Genealogy Indexer

Challenge

Make the content from over one million pages of paper documents, spanning three centuries and 20 languages, available to genealogists — by using advanced and fully-automated OCR to convert the records into searchable digital files.

Results

ABBYY Recognition Server enables the swift, accurate, and automatic conversion of scanned historical documents into text files, which are easily integrated into an online full-text search engine for genealogists.

Opening the past to genealogists, historians and families

For those who seek insights into the history of Jewish communities, as well as individuals researching their own ancestry, Genealogy Indexer provides an invaluable resource. A unique innovation in the field of Jewish Genealogy, the free website makes it possible to search original documents that have not been previously indexed. Created and maintained by Logan Kleinwaks as a service to historians and genealogists, Genealogy Indexer utilizes source materials from around the world — but primarily from Central and Eastern Europe, as Kleinwaks describes:

“Genealogy Indexer makes searchable more than a million pages of historical European directories, books commemorating Jewish communities destroyed in the Holocaust, military lists, school records and other documents of interest to genealogists and historians. Most of the material is not searchable elsewhere. The website is also free to use and completely non-commercial.”

Advancing genealogical research with ABBYY OCR

In 2008, Kleinwaks began the process of converting documents into fully searchable files and integrating them into Genealogy Indexer. “Even with many volunteers,” says Kleinwaks, “manually

“Recognition Server offers enormous benefit. The automation features are incredibly valuable and they save a lot of time...they reduce manual intervention to a minimum.”

— Logan Kleinwaks, Founder,
Genealogy Indexer

transcribing documents took a very long time. So OCR was key.” Initially, Kleinwaks tried a mix of OCR solutions. But the accuracy and versatility of ABBYY FineReader led him to standardize on it.

“Many of our documents,” explains Kleinwaks, “are from business directories, address directories, or telephone directories. They may arrive as paper — or as DjVu or PDF files of between 200 to 3,000 pages each, or multiple JPG or TIFF files. Often these are challenging for OCR because of poor print and paper quality, small dense text, complex layouts, and the high percentage of non-dictionary words such as surnames. ABBYY’s software,” Kleinwaks states, “was very good at meeting those challenges.

“Plus,” he adds, “ABBYY’s language capabilities were really valuable. We have documents in 20 languages and there was no problem recognizing them.”

Meeting the demand for automated high-volume Fraktur OCR with Recognition Server

However, Kleinwaks’ vision for Genealogy Indexer also extended to adding thousands of historical directories from Germany and German-speaking areas printed in Fraktur Gothic fonts during the 18th to early 20th centuries. He especially wanted to make directories from the 1930s searchable, to assist researchers of families separated during World War II. “Because of the large numbers involved,” says Kleinwaks, “finding a highly-automated OCR solution was essential - there are millions of pages that need to be converted.”

So, after discussing options with ABBYY, Kleinwaks decided to adopt ABBYY Recognition Server. “I discovered it is capable of handling high-volume Fraktur recognition,” states Kleinwaks, “thanks to its inclusion of the FineReader XIX module.”

ABBYY Recognition Server: Opening a new chapter in genealogical research

As a server-based document conversion solution, Recognition Server automatically converts high volumes of paper, image-only digital files and electronic documents into searchable records. Moreover, the software is capable of recognizing over 190 languages in a wide variety of fonts — including Fraktur.

Using Recognition Server, Kleinwaks performs OCR tasks on a single PC that hosts the server manager, processing station and verification station. Software developed by Kleinwaks then automates the post-OCR workflow — integrating the output files and document metadata with the site’s search engine.

“After OCR,” explains Kleinwaks, “I upload the output and a spreadsheet featuring metadata about the documents to my website and search engine server. From there, software I created integrates the OCR output and metadata into my search engine automatically — making the information available to users of Genealogy Indexer.”

The results

According to Kleinwaks, users are performing between 4,000 to 5,000 searches every day. The searchable content at their disposal now includes: 900,000 pages of 1,800 historical directories; 114,000 pages from 256 yizkor books; 32,000 pages of military lists; 43,000 pages of community and personal histories; and 24,000 pages of Polish secondary school reports and other school sources.

“Generally,” says Kleinwaks, “it is fair to say that OCR has greatly increased the use of Central and Eastern European directories as a genealogical source. And without Recognition Server I would simply not be able to do any of this.

“Being able,” he concludes, “to OCR Fraktur documents using Recognition Server has brought new users to my site and allowed existing users to search documents they never could before. No other solution I have tested comes close to delivering acceptable accuracy. And since I’ve been using Recognition Server its automation features have proven incredibly valuable. It saves a lot of time.”

About ABBYY

ABBYY is a leading global provider of technologies and solutions that help businesses effectively action information.

North American Headquarters
880 N. McCarthy Blvd. Suite 220
Milpitas, California 95035, USA
Tel +1.866.463.7689
Fax +1.408.457.9778
sales@abbyyusa.com



Copyright © 2017, ABBYY. All rights reserved. ABBYY, the ABBYY logo are either registered trademarks or trademarks of ABBYY Software Ltd. All other trademarks are the sole property of their respective owners. Information in this material is subject to change without notice. Please check the ABBYY web site for updated information. Part #8495e

www.ABBYY.com