a2ia

Case Study

a2ia DocumentReader Indexes and Extracts Information from Handwritten Census and Civic Records

Overview

Coutot-Roehrig, a European leader in finding unknown heirs, missing beneficiaries, real estate owners, property shareholders, and locating people or data to validate hereditary rights, developed out of the Coutot Genealogy Study of 1895. With 31 branches in France, Italy, Belgium, Poland and Switzerland, Coutot-Roehrig's services were originally utilized in the legal space, but have recently expanded to a much broader range of industries including financial and real estate institutions.

Today, Coutot-Roehrig maintains an international network and is most often tasked with finding next-of-kin, as far back as six generations, when unknown at the time of death. To date, Coutot-Roehrig manages more than half a billion documents (over 20 T-bytes) of department and municipal archive images, including public records, population census lists, and city/municipality resident information *a2ia DocumentReader*, a document classification and data extraction toolkit, was selected to automatically index and extract first and last name of individuals, as well as their family relationship, from over 350,000 tables from France's census and civic records dating from pre-1908.

a2ia DocumentReader was the only software toolkit able to overcome the complexity of this project, and alleviated the manual processes that were in place, saved Coutot-Roehrig money, sped the processing time, and ensured the accuracy of the data being recognized.

If these manual processes were not able to be automated, the project, by hand, is estimated to have required several dozen years of work and was error-prone. By implementing *a2ia DocumentReader*'s proprietary artificial intelligence and image analysis technology, Coutot-Roehrig was able to save money and allow for these historic documents and archives to be searched, online, by both genealogists and the public.

a2ia DocumentReader was the only software toolkit to overcome the complexity of the project.

Challenge

In addition to digitizing census records, Coutot-Roehrig has been compiling a collection of private archives since 2000, which have been digitized and posted on an intranet site for an easy search and review process by genealogists. The added-value of this electronic system was soon realized, and Coutot-Roehrig decided to digitize the records held by the departments and municipalities in France. Partnerships were established with many local communities who agreed to make their records available. Under these partnership agreements, the local communities benefit from the use of the digitized archives, which are available to the public and, in return, Coutot-Roehrig bears the costs of the digitization process and obtains a set of digitized images.

In order to optimize the searches of these records, a team of full-time employees was initially utilized to carry out the significant task of indexing the images manually. If the entire project were to be done by hand, it would have required several dozen years of work. These documents were handwritten, over 100 years old, varied greatly with regard to their quality because they were imaged by different service providers, and were stored in varying formats. Because of this, the manual process was extremely time consuming, costly, and error-prone.

> If the entire project were to be done by hand, it would have required several dozen years of work.

Solution

Because of these problems, Coutot-Roehrig sought to automate the processing of these 100-year-old handwritten documents.

After several tests, Coutot-Roehrig selected *a2ia DocumentReader* to automatically index France's census and civic records dating from pre-1908. The software was the only toolkit able to overcome the complexity of the project, which included handwritten documents, old-fashioned writing styles, mixed formats and poor quality. A2iA's Research and Development team was also able to provide professional services and customized development efforts for Coutot-Roehrig.

a2ia DocumentReader automatically extracted the first and last name of each individual, as well as the family relationship. First, *a2ia DocumentReader* cleaned the images and separated double-page spreads into single pages. Then, the engine detected the columns and segmented them into boxes.

Once a box was identified, it used a customized dictionary to automatically recognize the handwritten words contained inside. The extracted information was then converted into usable electronic data and transmitted to Coutot-Roehrig's system as an XML file, making these historic records and civic archives available to genealogists and the public.

Results

a2ia DocumentReader automated the indexing of over 350,000 pages of census tables, a process that if done by hand, would have required several dozen years of work. Because of the success of this project, Coutot-Roehrig plans to utilize *a2ia DocumentReader* to index other population censuses as well as other public record archives, which are typically even harder to process.

This installation worked with 100-year-old documents, handwritten in an old-fashioned style and of poor image quality. Typically, common OCR and ICR engines are unable to work with any type of cursive handwriting, especially those with varied formats. *a2ia DocumentReader* was able to perform classification/indexing based on both the layout/geometry and contents, whereas other engines only look at the layout. A2iA's ability to work with this unstructured, complex data, as well as provide professional services to Coutot-Roehrig, make this installation unique and a success.

Representative Images