

Looking to a New Day in OCR



“Incorporating ABBYY’s APIs into our software was a smooth process. We got great technical support—when we had a question we could send them a code fragment and they would identify the problem. It has been a great association.”

- Ratheesh Nair, Vice President of Technology, APEX CoVantage

Apex CoVantage Taps FineReader® Engine for Improved Accuracy and Sophisticated Features

Apex CoVantage was no neophyte when it came to using optical character recognition (OCR) technology to maximize efficiency in its business. The Herndon, VA-based provider of global Knowledge Process Outsourcing (KPO) has long used sophisticated technology and workflow systems to ensure that important content was available quickly to its customers who bought its engineering and content solutions. “We are always adding more customers and getting more data from existing customers,” said David Case, service delivery manager for the company. “OCR is absolutely necessary, a mainstay of what we do.”

However, the company eventually found that the system it had initially chosen could not adequately address the millions of pages that it scanned each year. “We knew that we needed to look for something that did a better job,” said Case. “This need was particularly driven by our customers’ requirements for better accuracy.”

Apex CoVantage scans and recognizes content from newspapers, journals and other sources for customers that use its content solutions. These customers receive their scanned data from Apex in an XML format, which they then put into the content solution and sell on a subscription basis to academic research libraries.

Fast, Formattable and Color Friendly

In searching for an alternative to its existing solution, the company had a number of important items on its shopping list. “Overall, we were looking to update our technology,” said Ratheesh Nair, vice president of technology at Apex CoVantage. “We knew we needed better bandwidth and improved storage. In addition, everything was moving to color so we wanted a system that would work on both color and grayscale.”



Client: Apex CoVantage

Location: Herndon, VA, USA

Vertical Market: Global Knowledge Process Outsourcing (KPO) Solutions

Challenge: Already using OCR, Apex CoVantage realized that it needed a more sophisticated solution that supported color scanning and could handle millions of pages per year.

Solution: ABBYY® FineReader® Engine

Results: FineReader Engine allowed the service provider to improve speed and accuracy substantially, particularly when recognizing old fonts.

Since the company was working with a variety of documents, the ability to accurately analyze document formatting and overall recognition accuracy was of prime importance. Finally, although many of the documents being handled by the system were in English, it also handled documents in Spanish and French, so multi-lingual support was necessary," said Nair.

High Tech for Historical

The company found that ABBYY FineReader Engine, a comprehensive software development kit (SDK) for document recognition, PDF conversion and data capture, offered all the features it needed to integrate robust capabilities into its system. FineReader supports almost 200 languages for OCR and more than 110 languages for ICR. The FineReader XIX module provides a unique capability to recognize texts published in the period from 1600 till 1937 in English, French, German, Italian and Spanish. It supports old fonts such as Fraktur, Schwabacher and the majority of Gothic fonts.

"One of the tools we were most intrigued by in ABBYY FineReader Engine was FineReader XIX and its ability to read hard-to-read fonts," said Harrison Yee, solutions architect manager at Apex CoVantage. "Some older documents used very ornamental black letter font faces and there is currently nothing on the market that does recognition of that with the exception of ABBYY. Without support for those old fonts, our results were terrible—complete garbage without any accuracy whatsoever."

As another plus, ABBYY FineReader Engine allowed for the recognition and creation of editable tables. "We want to preserve the document while preserving the formatting of tables," said Nair. "Now, we extract the tables and submit it to FineReader and it recognizes the text and outputs it as formattable text. Then it's simple for users to go in and modify that text."

Help at All Levels

Apex CoVantage also looked to ABBYY for help in getting the most out of FineReader Engine as possible. "We contacted ABBYY's technical support team, who helped us identify which product we should be buying, as well as samples and more insight into how the tool could be configured for the types of applications we have," said Nair. "Incorporating ABBYY's APIs into our software was a smooth process. We got great technical support—when we had a question we could send them a code fragment and they would identify the problem. It has been a great association."

Although the company had already reaped many benefits in its initial forays into using OCR technology, its move to ABBYY further streamlined the process. "There weren't quantifiable savings, but using ABBYY just made everything better," said Nair.

Improved recognition results did save time later in the company's process. For example, after recognition, the results were manually compared to the originals and corrected. "Depending on the quality of the OCR, we spend more or less time cleaning it up," said Micky Nijhawan, production manager at Apex CoVantage. "Our cleanup time was improved by ten to twenty percent."

As data recognition continues to be at the center of the work of Apex CoVantage, the company will continue to look for ways to continue to leverage the accuracy and speed of ABBYY FineReader Engine to get better and faster results for its customers.

ABBYY USA
47221 Fremont Blvd
Fremont, CA 94538, USA
Tel +1 866.463.7689
Fax +1 510.226.6069
sales@abbyyusa.com

www.ABBYY.com

ABBYY®