

ABBYY FineReader Gives KnowledgeLake Recognition Muscle

Client: KnowledgeLake Inc.

Location: St. Louis, Mo.

Business Goal:

To extend the capabilities and enhance the recognition accuracy of its Capture 2008 and Capture Server 2008 products.

Products Used:

ABBYY's FineReader® Engine and FormReader™ Developer Edition with FlexiCapture Studio add-on

Webs:

www.knowledgelake.com,
www.abbyy.com.

Leading ECM Vendor Selects ABBYY for Integration in Capture Solution

KnowledgeLake has made a name for itself as a leading provider of Enterprise Content Management (ECM) solutions for Microsoft SharePoint products and technologies. As it developed the newest iterations of its KnowledgeLake Capture 2008 and KnowledgeLake Capture Server 2008 (both part of KnowledgeLake Capture Suite), the company knew that it wanted to continue to extend the capabilities and quality of its results. It turned to ABBYY to provide additional technology to do just that.

"ABBYY's FineReader technology allows our users to create searchable PDFs," said Chris Caplinger, Chief Technology Officer of KnowledgeLake. "From inside of our Capture products, users can pull text of images and ready it for input from anywhere."

Products with a Purpose

KnowledgeLake's Capture Server 2008 puts end-to-end capabilities that empower organizations to convert paper documents into electronic data using any device (scanners, multi-function devices, or fax servers) from any location (central office, branches, or external service providers).

"Capture Server 2008 offers tight integration with SharePoint to enable enterprises to finally implement a complete capture solution around the leading ECM platform," said Caplinger. "With the KnowledgeLake suite of imaging, capture and transactional content management products, information workers from the remote office to the mail room can convert paper documents to digital content inside any SharePoint implementation."

In its search to extend the future capabilities of its products, KnowledgeLake considered several options. Another of its technology partners, Fujitsu Computer Products of America, recommended ABBYY's unmatched technology. "We have a strong relationship with Fujitsu and they recommended ABBYY over any other company," said Caplinger. "ABBYY offered more accurate recognition results than the others."

Bringing It All Together

KnowledgeLake used the ABBYY FineReader Engine® and FormReader® Developer Edition with FlexiCapture Studio add-on to meet its needs. The technologies offer image pre- and post-processing, document layout analysis, advanced PDF conversion and forms processing.

ABBYY FineReader Engine delivers complete full-page and field-level recognition capabilities to support key-word classification, and keyword indexing that KnowledgeLake is currently utilizing in its Capture 2008. The company also plans to plug in data capture functionality into its Capture 2008 product by integrating FormReader Developer Edition, its software development kit of choice.

FormReader Developer Edition together with FlexiCapture Studio helps users extract data from semi-structured forms and documents from a wide range of document types, including invoices, claim forms, resumes, contracts, and loan documents. In fact, even the smallest details, such as a hyperlink in a document are picked up, even if the text is in color rather than black and white, said Caplinger. It allows for the creation of a formalized description, called FlexiLayout, which tells ABBYY FormReader how to look for required fields on documents with similar data but different layouts.

“We were a little overwhelmed because the sheer amount of things that ABBYY does is more than we imagined,” said Caplinger. “There were so many possibilities within the engine that we were amazed...but we soon understood everything that was possible so we are using most of the capabilities. The amazing feature set blew us away.”

ABBYY's technology supports several important capabilities for optical character recognition (OCR) in the KnowledgeLake solutions. Now, the solution allows users to recognize digital text-based data and to create searchable PDF files. In addition, image cleanup capabilities (including rotating, straightening, and elimination of “noise”, etc.) allow for further improvement of recognition results.

Looking to the Future

ABBYY's technology has also allowed KnowledgeLake to position itself for growth into new areas in the future. “We want to move into forms processing and to start that process, ABBYY's engine does a much better job at organization of recognized data than other engines. It's much more intelligent. The engine will organize the data and tell you if it's on the same line as another word or character.”

The future product, dubbed KnowledgeLake Intelligent Document Recognition, will be a module of Capture Server. “It will allow us to automate the processing of invoices, explanation of benefits and more,” said Caplinger.