DocuShare OCR Crawler
Convert image files to searchable PDFs
for use with Xerox DocuShare

Finally, a really versatile solution for all your OCR needs including the ability to convert existing image files and image-only PDFs into fully searchable PDFs. If you find that you have hundreds to thousands to millions of images on your DocuShare server that you'd like to make fully content searchable, you can now use our DocuShare OCR Crawler to methodically walk through your nested collections to find images and convert them to compact, searchable PDFs as an ongoing background operation you can run once or schedule for weekly crawls for new content to convert.

All of this leverages the tremendous accuracy and versatility of ABBYY Recognition Server which can be used in a multitude of other ways. This is in contrast to Xerox' offering of the DocuShare OCR Module which must be used with Content Intake to trigger conversion of newly added image files. If you want to convert existing images you will also need the DocuShare Lifecycle Manager to establish policies to sweep your collections for image files to convert. That's an expensive proposition that costs $16,500 plus annual support and maintenance. That's all it will do for you.

ABBYY Recognition Server allows you to set up 32 workflows for use with fax servers, email, FTP input, website uploads, and shared folders for digital copiers, and input and output folders on user desktops for easy drag and drop conversions. This is a much more versatile approach that everyone in your company can use. See our information on ABBYY Recognition Server to better understand all your options. For example, you can use Recognition Server's Indexing Station to tag documents with metadata and use in conjunction with DAVupDoc to file new documents where they belong.

Both approaches will let you choose other output options such as Word, Excel, Rich Text Format, and plain text. Our approach adds output to PDF/A, XML, HTML, CSV, and ePub for mobile devices. Our approach allows you to install the OCR engine on your DocuShare server or on another server. We recommend installing your OCR engine on a separate server as OCR is very CPU and disk intensive and you don't want to negatively impact DocuShare performance in this way. However, workflows can be scheduled to run after hours.

Features & Options
Works in the background without affecting normal DocuShare use.
No noticeable impact on DocuShare performance when the OCR engine is installed on a separate server or run after peak usage hours.
Target specific collection sets for conversion. (i.e. parent collection and all subcollections)
Filter files to be converted by words in the title or other metadata.
Choose the output format for the conversion per collection set.
Run once or schedule for hourly/daily/weekly/monthly crawls of new files.
Generates new versions for documents so that the original image is retained for reference.
Can work with other document submission methods including DocuShare scan cover sheets, DocuPage Pro, FS2DS, and data capture solutions.

DocuShare OCR Crawler requires purchase of ABBYY Recognition Server. Both components are easy to install and configure. This solution is affordable, versatile, and scalable. Criteria First also carries other document imaging software such as our own DocuPage Pro and ABBYY FlexiCapture for use in conjunction with DocuShare. Call if you require other functionality.

There is no trial for this software. Call if you'd like an overview and demonstration and to properly configure your system. You can purchase this directly from our online store.

Criteria First
Solutions for managing knowledge in content & print
972-492-4428
http://www.criteriafirst.com
info@criteriafirst.com