



The Practice of Excellence

Manually converting image files to text searchable PDFs involves hours of work. The IT Director at DBL Law explains how a transition to batch OCR processing added value by automating the workflow.

How DBL Law convert images to searchable PDFs using batch OCR processing

The business need

- Ensure all files are searchable for regulatory compliance
- Remove the impact non-searchable files have on staff productivity
- Process non-searchable historical documents already in the document management system
- Switch to batch OCR processing for better handling of discovery intake
- Automate the processing of new files profiled into the document management system
- Implement a solution that integrates with the firm's iManage document management system

About DBL Law

Dressman Benzinger LaVelle psc, also known as DBL Law, is a full-service law firm with offices located in Cincinnati (OH), Crestview Hills (KY), and Louisville (KY). Their attorneys provide a high level of valuable legal services to private individuals, institutions, and companies in many industries and areas of law.

About non-searchable documents

Research has found that, on average, over 30% of documents in a content repository are non-searchable. Usually, these are image-based files like TIFFs, scanned PDFs, and emails with image attachments. Since there is no text in these documents, they can't be searched for using specific words or phrases.

The IT Director at DBL Law, **Rob Andres** knew failure to find documents was a significant risk to the firm. "If someone's trying to use the document management system as a research tool – to find an agreement to use as a template or see other types of case law we've worked on, for example, they're not going to be able to find what they need."

"contentCrawler gave us a much more efficient solution for batch OCR processing and, as a bonus, it automatically converts all non-searchable files in the document management system."

Rob Andres
IT Director, DBL Law

Recognizing non-searchable documents

Rob knew that the firm's old method of processing non-searchable documents wasn't a catch-all solution for making image-based files searchable. The firm had been using a PDF editor with Optical Character Recognition (OCR) functionality – technology that adds a text layer on top of an image file. But it couldn't manage the high volume of discovery the firm needed to process. "The PDF editor was great at OCR'ing, but with a large batch of documents it was worthless," said Rob. "Plus, it wasn't able to help us recognize non-searchable files in our document management system."

Rob saw the biggest need for batch OCR processing was within the medical malpractice department, "which gets a ton of discovery that is mostly scanned files." He described the previous workflow for OCR'ing this discovery:

When our litigation support staff had discovery to import into the e-discovery platform, they were splitting the files into batches of 500. They would OCR these batches one at a time using the PDF editor. This was a very manual process that needed to be tracked closely. Sometimes, something would just fail halfway through, and it would become a mess.

Switching to contentCrawler for batch OCR processing

Rob and his IT team at DBL Law [had recently deployed cleanDocs](#) when they assessed whether contentCrawler, also from DocsCorp, would meet the firm's needs for batch OCR processing. Explaining the decision to deploy contentCrawler, Rob said: "I didn't really look at anything else, because it was clear that I could take contentCrawler and point it at a folder, or at our document management system. That made it an easy sell."

Batch processing means staff at DBL Law don't have to spend any time on making image-based files searchable. Rob explains that since deploying contentCrawler, "if a discovery intake project comes up, I create a new job in contentCrawler, point it at a folder, and just let it run." OCR processing with contentCrawler is an automated service that runs in the background 24/7. "contentCrawler is hands off. I don't touch it, it just runs against our document management system all the time," said Rob.

DBL Law uses contentCrawler to OCR both historical and newly profiled documents, which has had a significant impact on the search power of their document management system. Any files with an added text layer are saved back into iManage as a new version. "Then, when our document management system goes back and indexes those newly searchable documents, it increases the power of the search two-fold," said Rob.

Summary

"We know contentCrawler is providing a lot of value," remarked Rob. Switching from manual OCR'ing with a PDF editor to batch OCR processing that is fully automated has had a real impact on the firm's ability to use their iManage document management system as a research tool. "contentCrawler gave us a much more efficient solution for batch OCR processing and, as a bonus, it automatically converts all non-searchable files in the document management system."

About DocsCorp

DocsCorp designs easy-to-use software and services for document professionals who use enterprise content management systems. The DocsCorp product suite is built to drive business efficiency and increase the value of existing technology investment. DocsCorp is a global brand with customers located in the Americas, Europe, and Asia Pacific. More than 3,500 organizations rely on DocsCorp software every day.

SYDNEY
LONDON
PITTSBURGH

info@docscorp.com
www.docscorp.com