opentext™

Data sheet

OpenText™ Image Crawler for eDOCS

Quickly find relevant image-based documents within eDOCS

As law firms and corporate law departments invest in legal content management, matter management systems, and search technologies, the expectation is that any document can be found instantly. The reality is different when more than 20% of documents remain "obscured" to search. These non-searchable, unfindable image-based files (email attachments, faxes, image PDFs, scanned documents, and the like) can be and are profiled in content repositories. But how do knowledge workers, including lawyers, easily access the information held within them without spending inordinate amounts of time opening files to view the content stored inside?

An OpenText eDOCS enhancement, Image Crawler is an integrated analysis, processing and reporting tool that automatically and intelligently assesses image-based documents in the eDOCS content repository for conversion to a searchable format, which is then re-profiled and consequently discoverable.

Fast processing

- Utilize available CPU cores with concurrent optical character recognition (OCR) processing

- Leverage default 4 CPU cores, additional licensing for 8, 16, and 32 CPU cores

- Process Search & Assess, OCR, and Save stages simultaneously

Content repository search

- Define searches to identify image documents and image PDFs, including those in email attachments

- Support TIFF, JPG, PNG, and BMP image types

- Refine searching on date ranges for Active Monitoring and Backlog mode

- Support multiple content repository databases or libraries

Assess text searchability

- Assesses image-based content in a repository for text searchability

- Identify PDFs with little or no text. Ignore text-based PDFs as they are already text-searchable

- Check email attachments in the Content Repository for text searchability

- Apply OCR to documents that meet the configurable text-searchability threshold

Make searchable

- Process documents to generate PDFs with a hidden text layer

- Ensure document fidelity with intelligent OCR technology

- Search and index hidden text layer. A separate text file to the image or PDF file is not required

- Use Search feature in Acrobat® or Reader® to find and review exact content

---

Improve productivity by making image-based documents findable

Increase search efficiency through automated indexing

Reduce non-compliance risks with accurate, on-demand search results

Leverage investments in DMS, search and OCR technology

List of associated OpenText products the component works with

- **OpenText eDOCS**

opentext™

- Support multi-language recognition with over 180 languages
- Unlimited page processing

Save to DMS

- Use DMS API for all connectivity, business logic, security models, and privileges honored
- Render image documents to text-layered PDFs before Save
- Replace the email attachments with OCR PDFs if appropriate before Save
- Save into DMS as a New Version, Related

Audit and reporting

- Monitor, configure, report via centralized administrative dashboard
- Control with 'Hold for Review' options prior to OCR and/or Save to Content Repository stages
- Embed Microsoft® SQL database for access to richer reporting if required
- Email notifications for periodic processing statistics and error reporting

Monitoring modes

- Automate workflows to make documents searchable
- Assess and OCR newly profiled or edited document profiles on a user-defined schedule using Active Monitoring mode
- Process legacy documents using Backlog mode

Integration

- OpenText eDOCS 10 and 16
- MS SharePoint® 2007, 2010, and 2013
- OpenText Content Server CS10 up to patch 7
- OpenText Livelink 9.7.1

Windows File System

- Search MS Windows® folders for non-searchable content in both Active Monitoring and Backlog modes
- Search for image-based PDFs, JPG, TIFF, PNG, BMP, and email messages with attachments
- Save as either

opentext™

Failure to locate a document can undermine productivity. Image Crawler adds a layer of text to every image-based document in the eDOCS library, whether you use SearchServer or OpenText™ MindServer™, making them available for indexing. This capability eliminates productivity losses caused by time spent looking for misfiled documents and skimming through others to determine context. Users can quickly locate specific words in specific documents using a simple, search function.

As businesses have invested heavily in legal content management, matter management and OCR and search technologies, Image Crawler extends the capabilities of these systems to drive business efficiency, increasing the value of the investment.

About OpenText

OpenText, The Information Company, enables organizations to gain insight through market leading information management solutions, on-premises or in the cloud. For more information about OpenText (NASDAQ: OTEX, TSX: OTEX) visit: opentext.com.

Connect with us:

- OpenText CEO Mark Barrenechea's blog
- Twitter | LinkedIn

opentext.com/contact Twitter | LinkedIn