

OpenText™ Image Crawler for eDOCS

Automatic, intelligent analysis and processing of image-based documents within eDOCS.

Your investment in content repositories such as Document Management Systems (DMS) and search technologies is an absolute necessity. And you expect to find whatever document you want instantly. The reality is somewhat different. More than 20% of documents in your content repositories remain “obscured” to search.

While it is true that the PDF format paved the way for the paperless office, electronic filing, and the iPad as a briefcase, the content within these documents is not easy to retrieve. These non-searchable, “unfindable”, image-based files (email attachments, faxes, image PDFs, scanned documents and others) can be and are profiled in your content repositories. Research has shown that knowledge workers spend 20% of their day looking for information in documents and only finding what they need 50% of the time. But how do you easily access the information held within them without spending inordinate amounts of time?

OpenText Image Crawler for eDOCS

Integrating OpenText Image Crawler for eDOCS changes everything. It is not a search engine and it does not modify the original document. As an enhancement to OpenText Document Management Systems, eDOCS Edition, Image Crawler is an integrated analysis, processing and reporting framework 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.

Reduce non-compliance risks

Failure to produce documents on demand can have an impact on regulatory compliance and exposes your organization to unnecessary risks. Specifically, sanctions or a dismissal of claims can undermine your organization's reputation. Image Crawler enables full content searching on every document in all your content repositories.

Increase organizational productivity

Despite significant investment in Enterprise Content Management systems, more than 20% of documents in these systems are “invisible” to search. Failure to locate a document can undermine productivity as well as put an organization's financial well-being at risk.

FEATURES

- *Fast processing*
- *Content repository search*
- *Assess text-searchability*
- *Make searchable (OCR)*
- *Save to DMS*
- *Audit and reporting*
- *Monitoring modes*
- *Integration*
- *Windows file system*

BENEFITS

- *Ensures searchability of all image-based documents*
- *Increases organizational productivity*
- *Simplifies the management of image-based documents*
- *Reduces non-compliance risks*
- *Increases efficiency through automation*
- *Leverages investment in DMS and search technology*
- *Reduces costs managing OCR technology*

Image Crawler adds a layer of text to every image-based document in your library, making them available to your search technology 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.

Leverage your investment in eDOCS

Businesses have invested heavily in Enterprise Content Management Systems as well as in search technology. Image Crawler extends the capabilities of these systems to drive business efficiency, increasing the value of the investment.

Features

OPPORTUNITY	DESCRIPTION
FAST PROCESSING	<ul style="list-style-type: none"> Concurrent OCR processing utilizes available CPU cores Default 4 CPU cores, additional licensing for 8, 16, and 32 CPU cores Simultaneous processing of Search & Assess, OCR, and Save stages
CONTENT REPOSITORY SEARCH	<ul style="list-style-type: none"> Definable searches to identify image documents and image PDFs, including those in email attachments Supports TIFF, JPG, PNG, and BMP image types Refined searching on date ranges for Active Monitoring and Backlog mode Supports multiple content repository databases or libraries
ASSESS TEXT-SEARCHABILITY	<ul style="list-style-type: none"> Assesses image-based content in a repository for text searchability Identifies PDFs with little or no text. Ignores text-based PDFs as they are already text-searchable Checks email attachments in the Content Repository for text searchability Applies OCR to documents that meet the configurable text-searchability threshold
MAKE SEARCHABLE (OCR)	<ul style="list-style-type: none"> Documents are processed to generate PDFs with a hidden text layer Intelligent OCR technology ensures document fidelity No requirement for a text file separate to the image or PDF file; hidden text layer is searchable and indexable Use Search feature in Acrobat® or Reader® to find and review exact content Multi-language recognition with over 180 languages supported Unlimited page processing
SAVE TO DMS	<ul style="list-style-type: none"> Uses DMS API for all connectivity, business logic, security models, and privileges honored Image documents rendered to text-layered PDFs before Save Replaces the email attachments with OCR PDFs if appropriate before Save Save into DMS as a New Version, Related
AUDIT AND REPORTING	<ul style="list-style-type: none"> Centralized administrative dashboard for monitoring, configuring, reporting Maximum control with 'Hold for Review' options prior to OCR and/or Save to Content Repository stages Embedded Microsoft® SQL database for access to richer reporting if required Email notifications for periodic processing statistics and error reporting
MONITORING MODES	<ul style="list-style-type: none"> Automates workflows to make documents searchable Assesses and OCRs newly-profiled or edited document profiles on a regular schedule of your choosing using Active Monitoring mode Legacy document handling and processing using Backlog mode
INTEGRATION	<ul style="list-style-type: none"> MS SharePoint® 2007, 2010, and 2013 OpenText Content Server CS10 up to patch 7 OpenText eDOCS DM 5.3.x and DM 10 OpenText Livelihood 9.7.1
WINDOWS FILE SYSTEM	<ul style="list-style-type: none"> Searches MS Windows® folders for non-searchable content in both Active Monitoring and Backlog modes Searches for image-based PDFs, JPG, TIFF, PNG, BMP, and email messages with attachments Save as either

www.opentext.com

NORTH AMERICA +800 499 6544 • UNITED STATES +1 847 267 9330 • GERMANY +49 89 4629-0
 UNITED KINGDOM +44 (0) 1189 848 000 • AUSTRALIA +61 2 9026 3400