

PDFDOCS CONTENT CRAWLER - OCR

Uncover non-searchable content in your Document Management System

Do you know what your exposure is to non-searchable content in your DMS? What is the risk to your organization in terms of reputation and financial well-being?

Organizations have invested heavily in Document Management Systems (DMS). Users should be able to search and find any document in the system using simple text-based searching. But what about documents such as scanned image PDFs, JPEGs and TIFFs, which are not text-searchable? Consider these same documents as attachments in email messages.

pdfDocs Content Crawler is a framework for searching an entire DMS database or a subset of documents based on specific DMS queries. The Content Crawler OCR module identifies non-searchable content, converts it to a text-searchable PDF using DocsCorp OCR technology and saves it back into the DMS. Users can convert legacy documents as well as newly-profiled documents.

Documents identified as being image documents are saved as either New Versions, Attachments (Open Text eDOCS only) or Related Documents. A layer of text is added to the new documents to facilitate searching. All documents in the DMS can now be searched.

SOURCE OF NON-SEARCHABLE DOCUMENTS

- Scanned images saved as TIFF or image PDF
- Emails with TIFF or image-based PDF attachments
- Electronic faxes saved as TIFF or PDF
- Legacy documents retained over many years
- Documents from business acquisition or litigation file ingestion

RISKS OF NON-SEARCHABLE DOCUMENTS

- Failure to find a critical document for a case
- Failure to find a document required for e-discovery in time
- Failure to realize a conflict of interest
- Failure to locate an old document template, which needs to be recreated
- Users lose confidence in DMS search capability to locate profiled documents

BENEFITS AT A GLANCE

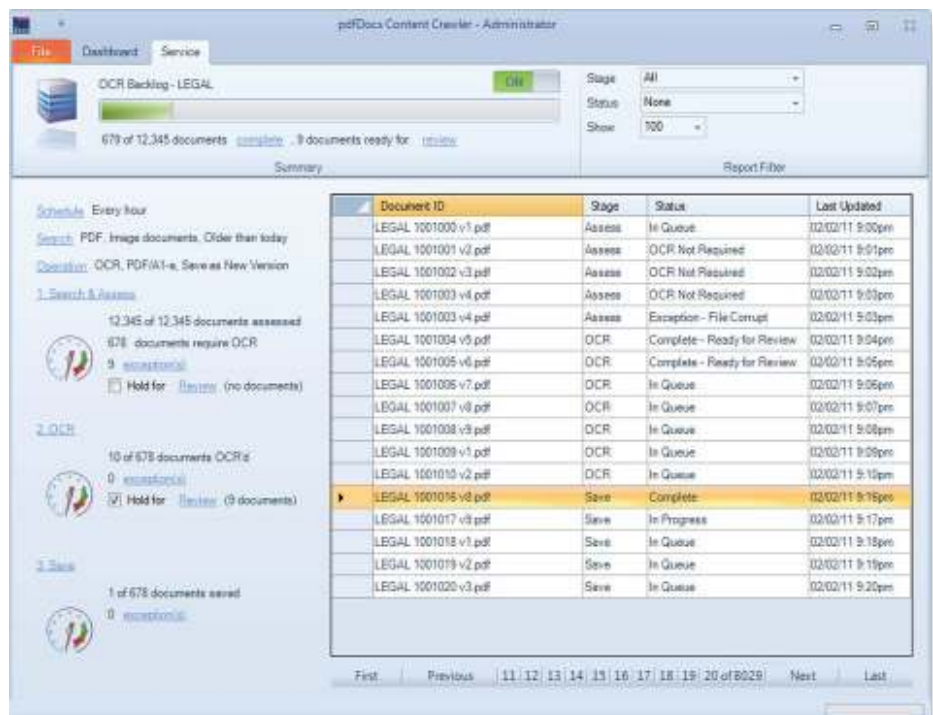
Expose the size of the problem of non-searchable content in your DMS

Unlock content in image files and image-based PDFs

Maximize the value of your investment in DMS and search technology

Tight integration with Autonomy iManage and Open Text eDOCS DM

Fully compliant PDF and PDF/A format, multi-language including Asian languages



FLEXIBLE CONFIGURATION

- Create any number of definable 'Services' to gain access to your DMS repositories
- Assess the text searchability of documents and if required OCR the documents to text-searchable PDF and save them back into the DMS

DMS SEARCH

- By default all image type documents, PDFs and email message files will be identified in the DMS
- Using standard DMS search queries or saved queries, Content Crawler can create complex search criteria to target specific areas
- Caters for multiple databases or libraries

ASSESS TEXT-SEARCHABILITY

- Checks all non-searchable content such as image files, image PDF and emails with attachments
- Identifies PDFs with little or no text. Text-based PDFs will not be processed as they are already text-searchable
- Uniquely looks inside emails stored in the DMS to assess text searchability of the attachments and replaces the attachment with OCR'd PDF

MAKE SEARCHABLE (OCR)

- Documents are processed using OCR technology to convert to a new PDF with a hidden text layer. No requirement for a text file separate to the image or PDF file
- The text layer is searchable by the DMS, and when viewed in Adobe Reader you can use the search feature to find and review content
- Options include creation of the PDF in conformance with the PDF/A standard (if required)

SAVE TO DMS

- Integrates with Autonomy iManage 8.2 or higher, Open Text eDOCS DM 5.1.05 or higher
- Uses DMS API for all connectivity – all security models and privileges honored
- Save the OCR PDF into DMS as a New Version, Related Document or as an Attachment (eDOCS only)

AUDIT AND REPORTING

- Rich administrative dashboard to monitor, configure and report on progress
- Maximum control with 'Hold for review' options prior to OCR and/or Save to DMS steps
- Embedded Microsoft SQL database for access to richer reporting if required

ACTIVE MONITORING

- Automate the protection provided by Content Crawler with the Active Monitoring service
- Newly-profiled or edited document profiles are assessed and OCR'd on a regular schedule of your choosing
- Documents are reliably and quickly made searchable in the DMS, with no interaction – set and forget!

SYSTEM REQUIREMENTS
 Microsoft® Windows® 7
 (32 and 64-bit)
 Microsoft® Windows
 Server® 2008
 Microsoft .NET Framework
 4.0
 Dual Core minimum
 processor
 4GB RAM minimum
 (8 GB recommended)
 DMS database/library
 connectivity via DMS API



SYDNEY
 MELBOURNE
 NEW YORK
 PORTLAND
 LONDON

info@docscorp.com
 www.docscorp.com