Innovative Computing Systems

contentCrawler

- **Increase organizational productivity**
- **Simplify management of image-based documents**
- **Reduce non-compliance risks**
- **Increase efficiency through automation**
- **Leverage investment in DMS and search technology**
- **Reduce costs managing OCR and Compression technology**

contentCrawler is an integrated analysis, processing and reporting framework that intelligently assesses documents in a content repository for bulk processing.

contentCrawler assesses and analyses documents in a Content Repository based on search criteria as well as text and compression thresholds configured by the IT Administrator. Documents are then processed based on the process type (OCR, Compression, or both), and saved back into the Content Repository. This is an automated back-end process that does not impact the desktop user.

MAKE EVERY DOCUMENT SEARCHABLE

Businesses have invested heavily in Content Repositories such as Document Management Systems as well as in search technology to ensure they have instant access to business-critical documents.

Despite this investment, 20% of documents in Content Repositories may be non-searchable and therefore "invisible" to search technology.

Failure to locate a business-critical document can undermine efficiency and productivity as well as put an organization's reputation and financial well-being at risk.

The contentCrawler framework can identify non-searchable content in a Document Management System database or a subset of documents based on specific queries.

The OCR module converts this content to text-searchable PDFs, saving them back into the Content Repository as new or replacement documents.

REDUCE FILE SIZE

Storage space in a Document Management System can be expensive. Large files can be costly to download and slow to send by email.

The contentCrawler Compression module can help reduce storage costs, speed up file transfers when downloading or sending via email by reducing the size of the files in your Document Management System.

The contentCrawler Compression module will enable Administrators to compress image and PDF documents in their Document Management System. Converting image documents to PDF and applying compression and downsampling to the files will reduce overall file size.

MULTI-PROCESS SERVICES

IT Administrators are able to combine the OCR and Compression modules into a single service.

Image documents will be converted to text-searchable PDFs to ensure the highest image quality. The Compression module will then reduce document file size through compression and downsampling.

EFFICIENCY THROUGH AUTOMATION

contentCrawler is an end-to-end automated solution that runs 24/7 without staff intervention. Staff do not have to worry about OCR or Compression processes or workflows. Instead, contentCrawler works in Backlog mode for legacy documents and Active Monitoring for recently-profiled documents.

It can work in both modes simultaneously.

"We use contentCrawler to ensure that newly profiled and legacy PDFs are fully text-searchable. DocsCorp has worked closely with us and has been very responsive to our requests for program enhancements."

Jeff Hutchinson: Mendes & Mount, LLP - Director of Information Technology

"contentCrawler has uncovered a range of documents, including PDFs that had previously not been searchable within our DMS. The solution has greatly enhanced our ability to find documents quickly with the use of our DMS search functionality."

Mark Turner: Lubbock Fine - Managing Partner

contentCrawler

MULTI-FUNCTION PROCESSING	OCR documents to produce text-searchable PDFs (OCR module) Compress PDF documents to reduce file size (Compression module)
FAST PROCESSING	Concurrent processing utilizes available CPU cores Default 4 CPU cores, additional licensing for 8, 16, and 32 CPU cores Simultaneous Search & Assess, Processing and Save stages
CONTENT REPOSITORY SEARCH	Definable searches to identify image documents and PDFs, including those in email attachments Supports TIF, JPG, PNG and BMP image types Refined searching on date ranges for Active Monitoring and Backlog modes Supports multiple content repository databases or libraries
MONITORING MODES	Automates workflows to make documents searchable and/or compressed Assesses and processes newly-profiled or edited document profiles on a regular schedule using Active Monitoring mode Legacy document handling and processing using Backlog mode
ASSESSES DOCUMENTS	Assesses content in a repository for text searchability (OCR module) and/or compression capability (Compression module)
MAKE SEARCHABLE (OCR)	Documents are OCR'd 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 PDF viewer to find and review exact content Multi-language recognition with over 180 languages supported. Unlimited page processing
REDUCE FILE SIZE (COMPRESSION)	Compresses using standard JPEG, JPEG2000 and JBIG2 formats Resizes and downsamples PDF image documents to reduce file size Automatically converts image files to PDF prior to compression, and converts to text-searchable PDF when used in conjunction with OCR module Configurable compression settings and file size reduction save threshold Reduces risk of download and email size limitations
SAVE TO DMS	Uses DMS API for all connectivity – all business logic, security models and privileges honored Image documents rendered to PDFs and compressed before Save Replaces the email attachments with processed and compressed PDFs if appropriate before Save Save into DMS as a New Version, Related Document, New Rendition or as an Attachment (options depend on DMS)
AUDIT AND REPORTING	Centralized administrative dashboard for monitoring, configuring, and reporting Maximum control with 'Hold for Review' options prior to Processing and/or Save to content repository stages Email notifications for periodic processing statistics and error reporting
WINDOWS FILE SYSTEM	Searches MS Windows folders for non-searchable content in both Active Monitoring and Backlog modes Searches for image-based PDF, JPG, TIF, PNG, BMP and email messages with attachments Save as either Replace Original or New Document as well as a New Location option

Innovative Computing Systems

SYSTEM REQUIREMENTS

OPERATING SYSTEMS

Windows 7 SP1
Windows 8
Microsoft® Windows Server® 2012 R2*
Microsoft® Windows Server® 2012*
Microsoft® Windows Server® 2008 R2 with SP1*
Microsoft® Windows Server® 2008 with SP2*
MS .NET Framework 3.5 and 4.5/4.5.1

* Not supported on Server Core Role

INTEGRATIONS

DMSforLegal/DMSforSharePoint
HP TRIM
iManage
MS SharePoint
OpenText Content Server
OpenText eDOCS DM
OpenText LiveLink
Worldox

Innovative Computing Systems

DocsCorp
Your documents. Integrated.

SYDNEY
LONDON
PORTLAND (OR)
MANILA

info@docscorp.com
www.docscorp.com