

NetDocuments is your central library for business documents, and yet our experience shows up to 30% of them may be invisible to search. Do you know how much non-searchable content you have?

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

contentCrawler analyzes documents 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 NetDocuments. This is an automated back-end process that does not impact the desktop user.

MAKE INFORMED BUSINESS DECISIONS

Find the 30% of image-based documents that are invisible to search, even in email attachments.

GET COMPLIANT, GET LEGAL

Ensure you comply with regulatory and legal requirements such as full disclosure in eDiscovery.

STREAMLINE OCR OPERATIONS

Replace multiple OCR endpoint solutions with a single, automated system that runs 24/7 with minimal staff intervention.

ELIMINATE INEFFICIENCY

Run OCR as a backend process, no need to have staff wasting time with inefficient OCR workflows that often get circumvented.

REDUCE FRUSTRATION

Ensuring documents get found first time, every time reduces workplace frustration and inspires confidence in NetDocuments search capabilities.

AUTOMATED END-TO-END PROCESSING

Document Management Administrators determine document types for conversion, and contentCrawler analyzes documents based on search criteria. Only documents that meet the criteria are processed.



The screenshot displays the contentCrawler interface within a NetDocuments environment. At the top, there are navigation tabs for 'File', 'Dashboard', 'Backlog', and 'Report'. Below these, a 'Testing Cabinet' is shown with a green progress bar. The status indicates '54 of 109 documents complete' and '29 documents ready'. The interface lists several configuration settings: 'Schedule' is set to 'Always active' and 'Every day'; 'Search' criteria include 'BMP,JPG,MSG,PDF,PNG,TIFF' and 'edited on or before 6/10/2015'; 'Operation' is set to 'OCR & Compress, Save as New Version'. A section titled '1. Search & Assess' provides a summary: '109 of 109 documents assessed (351 pages assessed)', '60 documents require processing [61.29 MB]', and a breakdown of 'OCR - 60 documents (97 pages)' and 'Compress - 22 documents (22 pages)'. It also shows '0 exception(s)' and a 'Hold for Review' checkbox.

Audit

Run a free Audit to determine the volume of non-searchable content in NetDocuments.

Search/assess

contentCrawler searches and intelligently assesses documents across all sites and libraries for non-searchable documents.

Process

contentCrawler processes these documents, converting them to smaller, text-searchable PDFs.

Save

contentCrawler saves the documents back into NetDocuments as a New Version or as a Replace the Original – ready to be found.

MULTI-FUNCTION PROCESSING	<p>OCR documents to produce text-searchable PDFs (OCR module)</p> <p>Compress PDF documents to reduce file size (Compression module)</p>
FAST PROCESSING	<p>Concurrent processing utilizes available CPU cores</p> <p>Default 4 CPU cores, additional licensing for 8, 16, and 32 CPU cores</p> <p>Simultaneous Search & Assess, Processing, and Save stages</p>
CONTENT REPOSITORY SEARCH	<p>Definable searches to identify image documents and PDFs, including those in email attachments</p> <p>Supports TIF, JPG, PNG, and BMP image types</p> <p>Refined searching on date ranges for Active Monitoring and Backlog modes</p> <p>Supports multiple content repository databases or libraries</p>
MONITORING MODES	<p>Automates workflows to make documents searchable and/or compressed</p> <p>Assesses and processes newly-profiled or edited document profiles on a regular schedule using Active Monitoring mode</p> <p>Legacy document handling and processing using Backlog mode</p>
ASSESSES DOCUMENTS	<p>Assesses content in a repository for text searchability (OCR module) and/or compression capability (Compression module)</p>
OCR MODULE	<p>Documents are OCRd to generate PDFs with a hidden text layer</p> <p>Intelligent OCR technology ensures document fidelity</p> <p>No requirement for a text file separate to the image or PDF file, hidden text layer is searchable and indexable</p> <p>Use Search feature in PDF viewer to find and review exact content</p> <p>Multi-language recognition with over 180 languages supported. Unlimited page processing</p>
COMPRESSION MODULE	<p>Compress using standard JPEG, JPEG2000, and JBIG2 formats</p> <p>Resizes and downsamples PDF image documents to reduce file size</p> <p>Automatically converts image files to PDF prior to compression, and converts to text-searchable PDF when used in conjunction with OCR module</p> <p>Configurable compression and file size reduction settings</p> <p>Reduces file size for downloads and email size limitations</p>
SAVE TO DMS	<p>Uses DMS API for all connectivity – all business logic, security models, and privileges honored</p> <p>Image documents rendered to PDFs and compressed before Save</p> <p>Replaces the email attachments with processed and compressed PDFs if appropriate before Save</p> <p>Save into DMS as a New Version, Related Document, New Rendition or as an Attachment (options depend on DMS)</p>
AUDIT AND REPORTING	<p>Centralized administrative dashboard for monitoring, configuring, and reporting</p> <p>Maximum control with 'Hold for Review' options prior to Processing and/or Save to content repository stages</p> <p>Email notifications for periodic processing statistics and error reporting</p>

Active Monitoring Mode

Processing can run in one of two (or both) modes: Backlog or Active Monitoring modes. Active Monitoring processes documents as soon as they are profiled into the document management system.

Backlog Mode

Backlog mode processes all legacy documents based on the process type.

Available Integrations

- iManage Work 10
- NetDocuments
- SharePoint Online

The application makes use of the following recognition technologies: ABBYY® FineReader® Engine 11.0 © 2008. FINEREADER, ABBYY & ABBYY FineReader are registered trademarks of ABBYY Software Ltd.

'contentCrawler' and the contentCrawler logo are trademarks of DocsCorp Group Pty Ltd.

contentCrawler's technology is protected under US Patent 8745084