

contentCrawler Compression

Free up space

Save on cloud storage, online collaboration and data migration costs

Expedite your cloud DMS migration

Faster upload/download times for files in the cloud

Reduce the size of discovery bundles

Include compression in your scanning workflow

LARGE FILES, LARGE STORAGE COSTS

File storage can be expensive, especially if you are storing on high-availability storage devices or paying by data volume to store files in the cloud.

One issue that contributes to the cost is bloated files that are larger than they need to be. This is often because their image content is unnecessarily large, or they were scanned at a higher resolution than required.

Document creators commonly include hi-res images in their documents without realizing the impact on file size, or they scan documents at the default resolution of their scanner, which is often too high and results in large file sizes.

REDUCE FILE SIZE

contentCrawler helps reduce storage costs by compressing unnecessarily large files. It looks for image file types and PDFs that contain image content, assessing them to determine if they can be successfully reduced in size.

Using lossless compression technology, it compresses only the files that need it and saves the resulting smaller file back into the content repository.

The system can be configured to replace the original file in the repository, add a new version or related document if processing documents in a DMS (document management system), or save the compressed documents in a separate location.

contentCrawler is a fully automated software solution that intelligently assesses documents to determine if they can be compressed to reduce their file size. contentCrawler also incorporates OCR technology to ensure all documents are fully text-searchable.

contentCrawler assesses and analyzes documents based on search criteria and predetermined compression thresholds. Where appropriate documents are processed using lossless compression technology and the resulting new files are then saved. contentCrawler runs as an automated server-based process.

REDUCE STORAGE COSTS

contentCrawler can significantly reduce storage costs, speed up file transfers and migrations, and reduce upload and download times. One law firm reduced the total size of its document repository by 35% with contentCrawler negating the need to expand their expensive SAN. This bought the firm enough time to complete its office move, which included an upgrade to higher capacity storage.

"With contentCrawler, staff time has been freed up to focus on creating an even better relationship with our clients, and to show them the firm is efficient and innovative with our use of technology."

Luke Schnoebelen
Learning & Development
Manager, Shook, Hardy & Bacon
L.L.P

"In fact, contentCrawler was easy to install—intuitive and simple. There were no new processes or staff training required. Everything just worked in the background. Staff members were completely unaware of any changes other than the fact that more documents started to show up in the search results."

Stephen Field
System Engineer, AJ Park

Another client compressed 100,000 documents within 48 hours, significantly reducing the total data size before uploading to a cloud collaboration platform. This expedited the project by reducing the upload time and cost as the collaboration vendor's charges were based on total data volume.

END-TO-END AUTOMATED SOLUTION

contentCrawler provides end-to-end automated processing that runs 24/7 to reduce the size of bloated files. The system works in two modes:

Backlog to process all files, or files from within a specified date range

Active Monitoring to watch for new files being saved into the repository and assess them for compression processing

It also provides OCR (Optical Character Recognition) converting image-based files into fully discoverable ones that your DMS or Enterprise Search System can index.

FAST PROCESSING	<p>Compression processing utilizes CPU cores</p> <p>Simultaneous Search & Assess, Processing and Save stages</p>	<p>SYSTEM REQUIREMENTS</p> <p>Operating Systems</p> <p>Microsoft® Windows Server® 2019, 2016, 2012 R2, 2012*, or 2008 R2 SP1* (all only 64-bit)</p> <p>*Unsupported on Server Core Role</p> <p>MS .NET Framework 4.5.2</p> <p>Hardware</p> <p>8 GB RAM</p> <p>100 GB free disk space</p> <p>1-2 GB per CPU core over 4*</p> <p>* Recommended: 4 dedicated CPUs</p> <p>INTEGRATIONS</p> <p>Elite ProLaw</p> <p>File System</p> <p>HP TRIM/Records Manager</p> <p>iManage FileSite/DeskSite</p> <p>iManage Work 10</p> <p>Micro Focus Content Manager</p> <p>Microsoft SharePoint</p> <p>Microsoft SharePoint Online (O365)</p> <p>NetDocuments</p> <p>OpenText Content Server</p> <p>OpenText eDOCS</p> <p>OpenText LiveLink</p> <p>Worldox</p> <p>The application makes use of these recognition technologies: ABBYY® FineReader® Engine 11.0 © 2008. They are registered trademarks of ABBYY Software Ltd.</p> <p>'contentCrawler' and the contentCrawler logo are trademarks of DocsCorp Group Pty Ltd.</p> <p>contentCrawler's technology is protected under US Patent 8745084</p>
CONTENT REPOSITORY SEARCH	<p>Define searches to identify image documents, PDFs, emails and their 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 compress documents</p> <p>Assesses and processes newly-profiled or edited document profiles on a regular schedule using Active Monitoring mode</p> <p>Handling and processing of legacy documents using Backlog mode</p>	
ASSESSES DOCUMENTS	<p>Assesses content in a repository for text searchability (OCR module) and/or compression capability (Compression module)</p>	
COMPRESSION	<p>Compresses using standard JPEG, JPEG2000, and JBIG2 formats</p> <p>Auto-converts image files to PDF before compression</p> <p>Resizes and downsamples PDF image documents to reduce file size, before Save</p> <p>Configurable compression and file size reduction settings</p> <p>Reduces risk of download and email-size limitations</p>	
SAVING TO DMS	<p>Uses DMS API for all connectivity and business logic, where security models and privileges are honored</p> <p>Replaces email attachments with processed and compressed PDFs before Save</p> <p>Includes Save As options for New Version, Related Document, New Rendition or Attachment</p>	
AUDIT AND REPORTING	<p>Centralized administration console for monitoring, configuring, and reporting</p> <p>'Hold for Review' options prior to Processing and Save to content repository stages</p> <p>Email notifications for periodic processing statistics and error reporting</p> <p>Exception reporting and 'Not supported' list filtering</p> <p>Optionally use a Microsoft SQL Server database for processing and reporting</p>	
WINDOWS FILE SYSTEM	<p>Searches MS Windows folders for non-searchable content in both Active Monitoring and Backlog modes</p> <p>Searches for image-based PDF, JPG, TIF, PNG, BMP, and emails with attachments</p> <p>Save as Replace Original or New Document, as well as to a New Location</p>	