LITERA

contentCrawler
Business Use Case

More than 30% of documents in your content repositories are invisible to search. That's a big problem.

Research indicates that as much as 30% of documents in a content repository are "invisible" to search. The culprits usually turn out to be image-based documents - JPGs, TIFs, PNGs and image PDFs. While many of these documents get OCR'ed, some do not, and since they are image files with no text, they do not get indexed. Instead they become invisible to search technology.

Mobile technology, document ingestion, and staff workarounds have punched huge holes in OCR'ing processes and workflows. This poses significant risk to businesses globally.

Failure to find documents erodes user confidence, leads to poor decision-making, reduces organizational productivity, and risks breaching regulatory compliance.

contentCrawler delivers huge benefits to businesses in terms of automation, productivity, and cost savings.

100% searchability - all image-based documents in a content repository are OCR'ed, ensuring that all documents are indexed.

Increased organizational productivity - by ensuring that every document is text-searchable, businesses will be able to eliminate productivity losses and downtime looking for lost or misfiled documents.

Increased efficiency through automation - businesses will be able to automate the entire process so that OCR'ing can take place 24/7.

Simplified management of image-based documents - firms will be able to do away with multiple OCR'ing processes and workflows, replacing them with a single, centralized approach.

Reduced costs - firms will be able to reduce OCR'ing hardware and software requirements.

Reduce non-compliance risks

Over the last several years we have seen a growing trend in the expansion of government regulations on privacy and data security.

Last year the European Union implemented the General Data Protection Regulation (GDPR) to give its citizens better control over their personal data. Similar laws have been passed in Australia, Canada, and the United States.

Failure to produce documents on demand can expose a business to unnecessary risks; sanctions, fines, and reputational damage from which it might never recover.

contentCrawler reduces this risk by ensuring every document in every content repository is fully searchable and retrievable.

Reduce workplace inefficiency

contentCrawler analyzes the documents in a content repository based on criteria set up by the IT department. Documents that meet the criteria are converted and automatically saved back into the content repository.

The process is completely automated and runs 24/7 without any staff intervention.

contentCrawler handles jobs of all sizes—whether it is the day-to-day profiling of documents or the ingestion of 100,000s of documents into the system as a result of a merger or legal discovery process. It can also process legacy documents.

Staff no longer have to worry about OCR as a process or workflow, freeing them up to work on other projects.

24/7

Reduce operating costs

Business are always looking for ways to reduce their operating costs without sacrificing the quality of their offering or making the lives of their employees more difficult.

contentCrawler delivers on both counts. It enables businesses to immediately eliminate OCR'ing software and hardware costs, while removing staff from processes that are inefficient and labor-intensive.

Reduce lost productivity

Your enterprise content management system is the central library for your corporate documents, and yet 30% of those documents are typically not indexed by search.
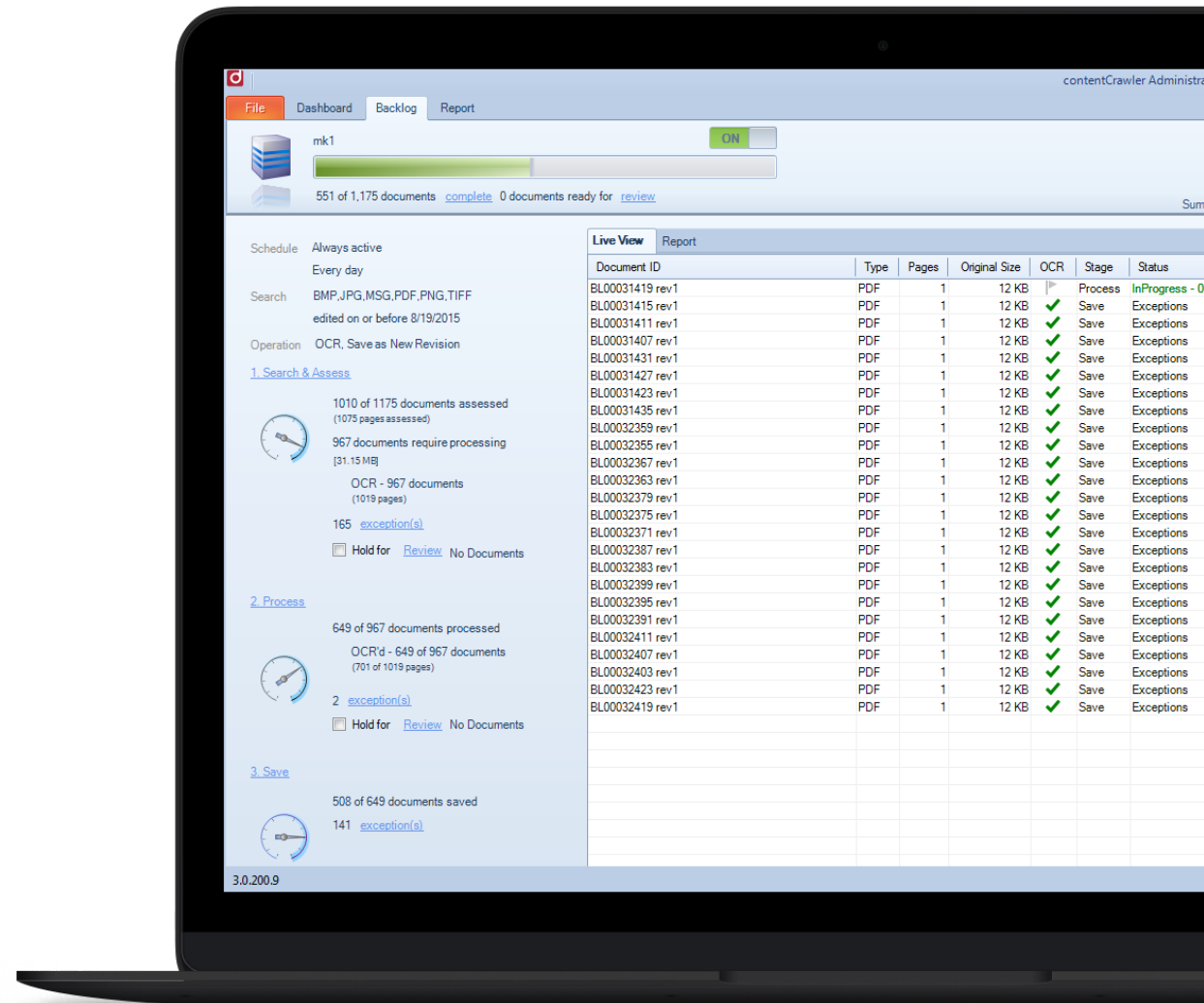
Failure to find documents erodes user confidence in the system. More importantly, staff waste significant time looking for these documents or lodging support tickets with IT for assistance. It has been estimated that an enterprise with 1,000 workers wastes between USD$2.5 and $3.5 million a year searching for — and failing to find — important documents.

When documents cannot be found, staff resort to recreating the document from scratch — wasting more time.

Reduce complexity

The contentCrawler Administration Console's dashboard provides up to the minute progress, showing the number and percentage of documents OCR'ed and compressed.

Additionally, you can see each document status in the Live View:
what needs to be assessed, requires OCR or compression, has been successfully processed, or has failed either or both OCR and compression.

What our clients are saying

"Law firms don't just store files in the DMS, firms also store documents in their network file shares. Since contentCrawler OCRs documents in all these repositories, your search results are much improved – something that is very important for GDPR".

Henrik Jarnberg
Head of IT at Delphi

"To comply with the GDPR, we needed to ensure that email attachments like scanned copies of identity documents and other image files could be searched for and deleted when necessary."

Stefan Winquist
CIO at Bech-Bruun

"The software has been invaluable. The ROI was immediately realised when users no longer had to spend 30-45 minutes OCR'ing a document. They just dropped it into ProLaw and it was fully-searchable the next morning. 30-45 minutes of OCR'ing 4 times a week ends up being 2-3 hours of productive time, 8-12 hours per month, 96-144 hours per year per assistant!"

Dan Featherman
IT Manager at Garlington, Lohn & Robinson, PLLP

contentCrawler

contentCrawler is an integrated analysis, processing, and reporting framework that intelligently assesses documents in a content repository for bulk processing. Users can bulk process documents using either the OCR or Compression modules to make image-based documents searchable or smaller in file size.

Modules

Compression
The Compression module converts image documents to PDF, applying compression and downsampling to the documents to reduce overall file size.

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

Better together
The OCR and Compression modules can be combined into a single service. Image documents will be converted to text searchable PDFs to ensure the highest image quality. The Compression module then reduces document file size.

LITERA

About Litera

Litera has been a global leader in legal technology for 25+ years, helping legal teams work more efficiently, accurately, and competitively. As a leader in document workflow, collaboration, and data management solutions, we empower legal teams with simplified technology for creating and managing all their documents, deals, cases, and data.

For more information about Litera visit litera.com or follow us on LinkedIn.