

How Stibbe used contentCrawler to index 28 million documents and emails for its enterprise search engine



Stibbe

About Stibbe

Stibbe is an internationally-orientated Benelux law firm with over 375 lawyers. From its main offices in Amsterdam, Brussels and Luxembourg, together with its branch offices in Dubai, London and New York, Stibbe handles complex legal challenges for its clients both locally and cross-border. As a specialist firm, Stibbe's lawyers work in multidisciplinary teams and deliver pragmatic advice. They build close business relationships with their clients that range from local and multinational corporations to financial institutions, government organizations and public authorities. Stibbe's understanding of its clients' commercial objectives, their position in the market, and their sector or industry allows the firm to always provide clients with timely, effective, and appropriate advice on their complex local and cross-border legal challenges.

Using enterprise search for GDPR compliance

The ability to search for and find 100% of documents is required to meet data return, erase, and portability requirements under the GDPR.

"We invested in an enterprise search engine to be future-proof before the GDPR came into force in May 2018," explained Olivier Van Eesbeecq, Head of ICT & Facilities at Stibbe Belgium. "Several products we were using – including our document management system – came with their own search engines, but we found them to be lacking. So, we decided to invest in enterprise search technology."



Legal



375 lawyers



Finding Documents

The business needs:

Automatically find and convert image-based files to searchable PDFs

Maximize the value of the enterprise search engine

Comply with GDPR data return, erase, and portability requirements

Avoid impacting staff workflows with new OCR or scanning requirements

“

We were already using the DocsCorp desktop productivity solutions, so when we learned there was an automated OCR solution as well, choosing it was a no-brainer.

”

Olivier Van Eesbeecq, Head of ICT & Facilities at Stibbe Belgium

The problem with non-searchable files

For it to work effectively, enterprise search relies on the existence of a text layer in every single file in your system. But scanned files, TIFFs, JPEGs, and image-based PDFs (which Stibbe Brussels had many) – don't have that layer.

Full-text search in your documents is important because a) we don't always remember the name of a file, so it's essential that on-page content can be searched, and b) under the GDPR, you need to be able to search for and find every document that contains a name, email address, bank account number, or other personal data.

To get the maximum benefit of its enterprise search investment, Stibbe needed a solution that could find non-searchable files that were not indexed for searching and could process them, so it had the necessary text layer to be indexed for searching.

Bulk conversion into searchable PDFs

Search and assess technologies using OCR software can find non-searchable content and automatically convert them into text-searchable PDFs. Stibbe required a solution that could work “in the background,” so it wouldn't impact staff workflows or processes.

“We were already using the DocsCorp desktop productivity solutions,” said Olivier, “so when we learned there was an automated OCR solution as well, choosing it was a no-brainer for us.”

contentCrawler is configured at Stibbe to be a set-and-forget solution. Staff continue to upload documents into the document management system, for example, without worrying about their need to be OCR'd. “If our lawyers photocopy or scan a file, they simply add it to the document management system, and it's automatically made searchable. That's a big advantage,” Olivier commented.

“contentCrawler connected to all our document sources – like file servers, email servers, the document management system, SharePoint – and converted all the content into searchable PDFs,” Olivier continued. “Once contentCrawler processed the files, the search engine picked it up and indexed it within minutes.”

“We now have more than 28 million documents and emails indexed by our enterprise search engine. All that content is now searchable thanks to contentCrawler.”

Have staff noticed a difference?

“Absolutely. Our staff have certainly noticed a difference since having contentCrawler,” said Olivier. “Although it's a background process, they really see the value because they trust that their documents will be automatically indexed and made searchable. It also saves them time since they no longer need to use desktop scanners to manually OCR files.”

Summary

Stibbe used contentCrawler to unlock the benefits of its enterprise search engine since non-searchable documents were impacting its performance. Now, the firm has a solution that works silently behind the scenes, automatically catching every new document added to its file systems and adding a text layer when needed. Staff are now able to search for and find content across 28 million documents and emails, and the firm can comply with GDPR requirements for data storage and handling.