Despite a significant investment in Autonomy iManage, AJ Park, a New Zealand intellectual property law firm, discovered that 30% of its content was invisible to search technology. This represented a significant and unacceptable risk to the firm.

Read how the firm turned this around to ensure content was 100% searchable with contentCrawler...

AJ Park goes with contentCrawler to minimize the risks of non-searchable content in Autonomy iManage

AJ Park specializes in intellectual property law in New Zealand, Australia and the Pacific region. The firm works with clients across a range of sectors including biotechnology, chemical, electrical and electronics, mechanical and manufacturing, materials and nanotechnology, IT and software industries. With offices in Auckland, Wellington and Sydney, the firm counts over a third of New Zealand's top 100 companies and almost half of the global Fortune 500 companies as clients.

The business challenge

It is a fact that Document Management (DMS) and Enterprise Content Management (ECM) systems are full of image-based content. The problem is, there is no easy way to determine the size of the problem, or how much it will cost to fix it. This was precisely the dilemma that AJ Park faced in 2011, according to Stephen Field, a System Engineer at the firm. "We concluded that image-based documents in the Autonomy iManage DMS represented a serious risk to the firm. So, we started to look for a solution."

> "We concluded that image-based documents in the Autonomy iManage DMS represented a serious risk to the firm. So, we started to look for a solution."
>
> Stephen Field
> System Engineer,
> AJ Park

The firm has been a DocsCorp client for several years using their pdfDocs products for creating and editing PDF documents. It was through this relationship that they became aware of the contentCrawler product. Stephen recalls how they obtained the contentCrawler audit tool to put some actual numbers on the scale of the problem and to build the business case for resolving it. "After running the audit tool on a section of the Autonomy iManage database, we concluded that there was about 30% of non-searchable content. When you have 4 million plus documents stored in Autonomy iManage, this is a sizeable number of documents being omitted from searches," says Stephen.

The firm had two concerns initially. They wanted reassurance that contentCrawler would not modify or change the actual appearance of the document. DocsCorp assured them that it didn't. In fact, contentCrawler does not modify the original documents, instead it simply adds a text layer to facilitate indexing and searching. Further assurances were given that it would also preserve any annotations that might have been on the original, and that it was 99.9% accurate, supporting more than 180 languages.

Secondly, the firm did not want to double up on storage. Again, contentCrawler provided the firm with a number of options for saving documents back into Autonomy iManage. Documents could be

DocsCorp

saved as a new version or replace the original. AJ Park decided to replace the original with the new searchable PDF. The IT department conducted a period of testing. They were happy with the results of the test. Stephen claimed that "the documents showed up in the searches as described in the brochure."

Our solution

The firm decided to proceed with the purchase and deployment of contentCrawler. But before commencing, the IT department made a number of decisions on how contentCrawler would tackle the enormous library of over 4 million documents. The first decision was to automate the entire process.

The process would be an end-to-end, automated process with contentCrawler assessing, converting, saving and replacing the original documents with no intervention from staff. This would allow them to run the contentCrawler service 24/7 to complete the task as quickly as possible. contentCrawler can also be run as a manual process with built-in "Hold for Review" options prior to the OCR and/or "Save to" stages.

In addition to running the crawl as an automated process, AJ Park decided to tackle the problem in two stages. The first stage would focus on the conversion of all the legacy documents year by year, and the second would handle all newly-profiled documents. contentCrawler provides organizations with the flexibility to work in one of two (or both) modes precisely for this reason.

The Backlog mode handles all legacy documents whereas Active Monitoring processes recently profiled documents. Once the legacy PDF documents in the firm's Autonomy iManage database had all been converted to searchable PDFs, the IT department turned their attention to ensuring all newly-profiled documents would be handled in a similar way. This provided the firm with a single, back-end OCR solution that eliminated the need for multiple OCR workflows and processes. It also allowed end users to forget about OCR and focus on other tasks.

Other benefits

Initially, the AJ Park IT department was concerned that such a solution would require them to buy and support new hardware. However, this was not the case as the firm was able to run contentCrawler on their existing hardware and operating systems.

"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," concludes Stephen.

In summary

New Zealand IP law firm, AJ Park sought an automated solution that would convert legacy and newly-profiled PDF documents in its Autonomy iManage DMS to searchable PDFs to ensure that every document was 100% text searchable.

One hundred percent searchability would minimize risks associated with failing to produce documents on demand, or failing to recognize any conflicts of interest in taking on new clients. The service provided the firm with a single, enterprise-wide OCR solution that eliminated the need for multiple OCR workflows and processes.

"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

DocsCorp

KEY BENEFITS
Leveraged investment in Autonomy iManage and in search technology to drive business efficiency.
Reduced non-compliance risks by ensuring full content searching on every document in every content repository.
Reduced costs managing OCR technology by eliminating the need for other OCR'ing solutions.
Increased organizational productivity by eliminating productivity losses and downtime looking for misfiled documents.
Simplified management of image-based documents through a centralized monitoring and reporting dashboard.