

Failure to find image-based documents in their HP WorkSite document management system was impacting staff productivity and confidence at the Electricity Authority, New Zealand's electricity market regulator.

Learn how contentCrawler was able to shine some light on the "invisible" document problem...



contentCrawler empowers Electricity Authority to find "invisible" documents in their DMS

The Electricity Authority of New Zealand (the Authority), based in Wellington, is an independent Crown entity that was formed in 2010.

Prior to that there was an Electricity Commission. The Authority is the electricity market regulator and responsible for the efficient operation of the New Zealand electricity market. It promotes competition in, reliable supply by, and the efficient operation of the electricity industry for the long-term benefit of consumers.

The business challenge

Board papers originating with the Electricity Commission for which the Authority is now responsible, started life as electronic documents that were printed to form part of a 'Board pack' and later scanned and saved into a content repository to make them more accessible to everyone within the organization. The authoritative document was therefore the scanned version.

However, scanned documents are basically like pictures. You can see what is on the page, but you cannot search what is on the page. Therefore, if you need to search for a particular document or phrase within a document, you will not find it. This was precisely the issue that confronted the staff at the Electricity Authority.

Searching for board papers, or looking for any document that had been scanned, often produced the same result. They couldn't easily find papers that had been scanned and profiled into their HP WorkSite document management system, which contained more than 100,000 documents. What's more it was not easy to know what had been scanned and what had not. Time was wasted looking for documents the old fashioned way. The result was that staff were becoming increasingly frustrated with the system.

Suzanne Jones, Knowledge and Information Manager at the Electricity Authority, recalls a meeting with their HP WorkSite consultants, Next Page, in which the issue of finding non-searchable content in HP WorkSite was discussed. It was at this meeting that contentCrawler from DocsCorp was suggested as a possible solution to the problem.

Our solution

contentCrawler is an integrated analysis, processing and reporting framework that makes content in a document management system or Windows file system 100% searchable.

It is a fact that DMS and ECM content repositories are full of non-searchable 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. The contentCrawler audit tool can provide the numbers to build the business case for solving this problem.

The Electricity Authority decided to run the audit tool in a test environment that was set up to put contentCrawler through its paces. This proved to be successful and also enabled the Authority to size the extent of the issue. It discovered that almost 10% of existing documents were scanned images. Consequently the application was put on the live system.

contentCrawler can run in one of two, or both modes simultaneously. Backlog mode assesses every image-based document in the content repository. It then converts only those documents that meet the criteria to text-searchable PDFs. The documents are then profiled back into the content repository overwriting the originals or creating a new copy of the document. Active Monitoring looks for recently profiled documents only and processes them.

The Electricity Authority decided to run contentCrawler first in Backlog mode on its legacy documents. Once this was complete, contentCrawler ran in Active Monitoring mode.

Other benefits

The Electricity Authority also decided to automate the process. contentCrawler can run as an automated end-to-end process or manually with built-in "Hold for Review" stages. The fact that the solution was completely automated meant that it could run 24/7 without staff intervention. It also meant that there was no need for any other OCR'ing hardware or software.

By performing the conversion process at the backend, it had no impact on staff workflows or processes. They could continue to profile documents into the document management system without worrying about OCR as a process or workflow. "For us it really was a case of set and forget. Staff now comment on how easy it is to find and reuse documents," concluded Suzanne.

In summary

New Zealand Electricity Authority staff grew increasingly frustrated looking for documents in their HP WorkSite document management system. Scanned documents saved into the system were essentially "invisible" to the search engine. Upon the recommendation of their consultants, the Electricity Authority tested and deployed contentCrawler to automate the process of converting and profiling these image-based documents into text-searchable PDFs.

"For us it really was a case of set and forget. Staff now comment on how easy it is to find and reuse documents."

Suzanne Jones
Electricity Authority
Knowledge and Information
Manager

SYDNEY
LONDON
PORTLAND (OR)
PITTSBURGH
MANILA

info@docscorp.com
www.docscorp.com



KEY BENEFITS

- Leveraged investment in HP WorkSite 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 documents.
- Simplified management of image-based documents through a centralized monitoring and reporting dashboard.