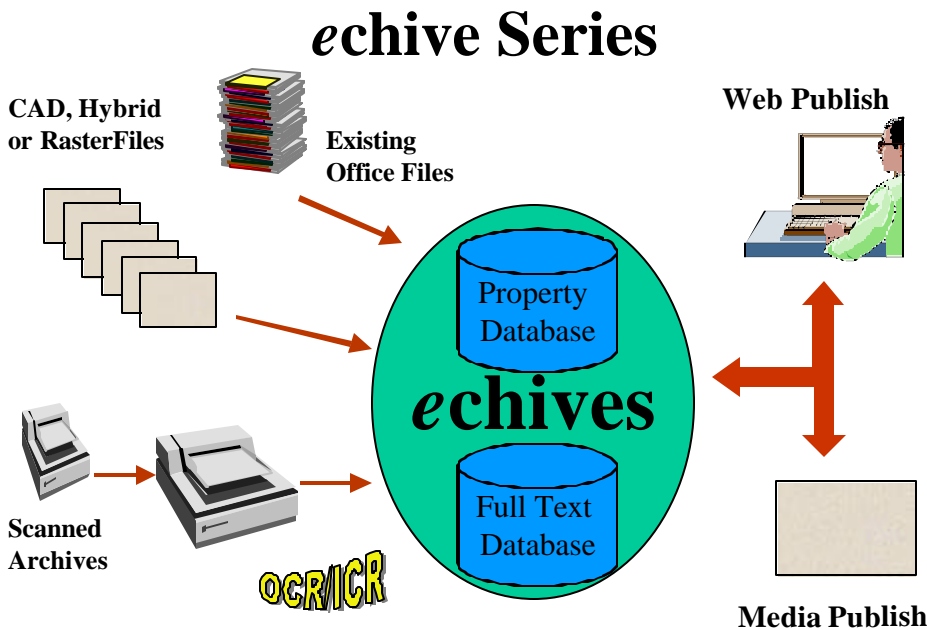


Patent Analysis gets a new Image

By David J. Wilson
for Imaging Service Bureau News

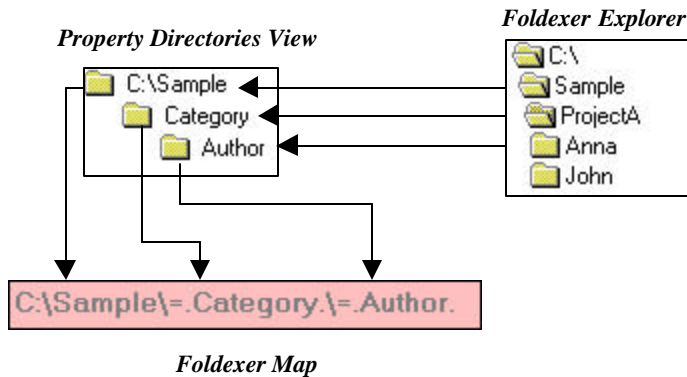
While companies strive to protect their investments, the U.S. Patent Office is challenged with approving or rejecting patent applications for many new approaches, business models, and unheard of ideas. Imaging technology and tools can significantly impact patent application and infringement research resulting in a more accurate and validation process.

My firsthand experience with patent analysis began when I was contracted as an expert witness for a patent infringement case. The specifics will remain nameless but the situation required in-depth analysis of legacy documents to prepare and defend a patent invalidation case. As part of the project I was required to review and analyze source code, users manuals, and marketing materials provided to me in both paper and electronic form. This included thoroughly searching for “prior art” products. Products related to the patent and available at or before the time of the filed patent are defined as “prior art” products.



The challenge is to accurately analyze a vast amount of information, much of it unfamiliar material. A recently completed software product for our company provided new tools to make the job much more efficient. The product included technologies that could be applied to my patent infringement research.

The first step was to collect the materials from prior art products and other related patents. Source code and some patents were already in electronic form (text files and PDF files). I organized these files in a traditional file directory structure then applied property field indexing tools. Any property field (categories such as author, title, creation date, etc.) can be populated with the sub-directory names themselves. This allows custom sorting, categorizing, and property searching.



Secondly, scanned images, which represented a majority of material, needed to be imported and indexed. Scanning and subsequent ICR text extraction, an imaging bureau staple, prepares the scanned image for indexing into a search and retrieval engine for later retrieval. Similarly, binary files like CAD and PDF documents can also be extracted into the full text database in preparation for full text indexing.

Once all the information is captured into the system, we can then index the full text content of *all* the documents including scanned documents, scanned images, PDF, and electronic files. This simple indexing process significantly impacts file search and retrieval ensuring accurate and complete document research on native file formats. Full text indexing allows advanced search engine technologies, successfully being used in document retrieval across the World Wide Web, to apply to our own subset of documents.

Since a full text extraction is being performed on much of the material, it is imperative that the best of quality and highest degree of performance is utilized. This provides an opportunity for high quality scanning and ICR services to assure companies that their documents will be properly indexed.

A full text search system provides efficient information location (word processing files, CAD drawings, scanned images, PDF files and more) based on any word or phrase. As each character of a search word is entered, the system dynamically displays the words contained within the various sources of information that were indexed. A hit count is displayed along side each word.

Advanced search tools including fuzzy searching, proximity, searching, noise word ignoring, and stemming can provide even more accurate results and improve document location efficiency.

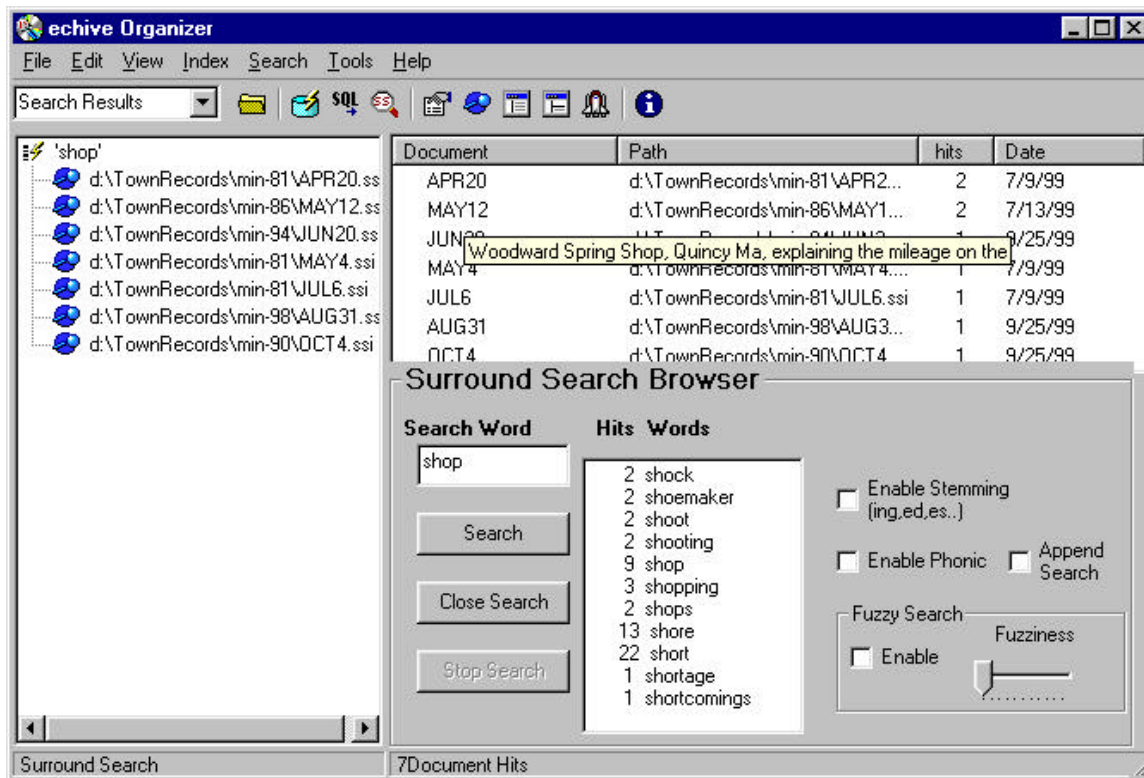
- Proximity searching allows the user to locate documents based on two identified words being within a specified number of words of each other. For example, you can tell the system to find all documents where “month” is within 3 words of “May”.
- Fuzzy searching is common in applications using OCR to help compensate for poor quality originals and hence poor results. Fuzzy search locates words similar, but not an exact match, to your search word and reports a confidence factor back to the user. The original scanned image is always displayed to ensure you have the exact document regardless of any possible ICR/OCR inaccuracies.



Open Archive Systems
603-890-9980
info@openarchive.com
www.openarchive.com

- Another advanced search technique, stemming, allows you to define the search rules to match a specified word regardless of its suffix. You can search for the word “shop” and receive hits on the words “shopping”, “shopper”, “shops” but it would not include a word like “shoplift” as a valid hit.
- Common words used in every day language (“a” or “the”) can be ignored and treated as noise. A noise dictionary can be used to eliminate unnecessary indexing of words too common to add to a searching system.

When a word from the hit list is selected, SurroundSearch™ displays the document names containing your word/phrase and the number of occurrences within that document. As you move your mouse over the list of documents, Tooltips shows the sentence within that document that contains your search word/phrase. This extracted text and all occurrences of the matching word are provided in a unique color for easy identification. This allows you to preview the sentence and document prior to viewing it using the built-in viewer or launching it into its native application.



When you have found a good match, you can use the Windows cut and paste buffer to collect information. Highlight a phrase and copy it to the Windows clipboard as a “legal quote”. The document name and quotes are automatically added to the text string identified by the Windows selection cursor. You can keep appending quotes, thus preserving previous selections. When you are done, simply paste all the collected quotes into a word processing application along with your other thoughts related to the topic.



Open Archive Systems
 603-890-9980
 info@openarchive.com
 www.openarchive.com

The benefits with tools like this are significant. Not only is time saved, but accuracy is improved in the location and analysis of multiple information sources. It is now possible to quickly search and view sentence content without having to view and read an entire document. With many documents, the cost saving is very measurable. In the world of patent application, validation, or infringement, these tools can help assure companies that their technology investments are protected.

The service opportunities go well beyond pure scanning and image clean-up. A full text system allows bureaus to add value to traditional scan and save services. Using full text extraction from images and CAD files, you can provide indexed and populated databases on CD and DVD or explore Web based service provider solutions.

The advent of the Web and ASP service models introduces an exciting opportunity to imaging service providers to organize and manage documents for legal or other clients applications as data repository providers. Maybe it's something worthy of a patent.



Open Archive Systems

603-890-9980

info@openarchive.com

www.openarchive.com