Dynamic Metadata Search in a Web-Based Document Management Application

Article contributed by dtSearch®

In previous articles, I've gone on at length extolling the value of integrated full-text and metadata searching. (Please see, for example, USA Daily Chronicles.)

Like what?

Suppose that you are looking for *ProjectX and ProjectY and San Francisco and not Cleveland* in the full-text of all of your documents, emails plus attachments and the like. Sometimes, a full-text only search query is good enough. But other times, you may want to limit that search to only files with certain metadata, such as *subject field contains sensitive office memo* to hone in immediately on the most relevant files. I'd like to showcase our dtSearch Engine customer FileHold which recently featured an advanced variation on metadata search in one of their blog posts.

Can you describe dtSearch's relationship with FileHold for our listeners?

Functionally, dtSearch is just the search engine. But while dtSearch can find and display retrieved data with highlighted hits, it does not have important features that enterprises need, like document management, archiving and workflow capabilities. This is where our dtSearch Engine developer customers come in.

Can you tell us a little bit about FileHold?

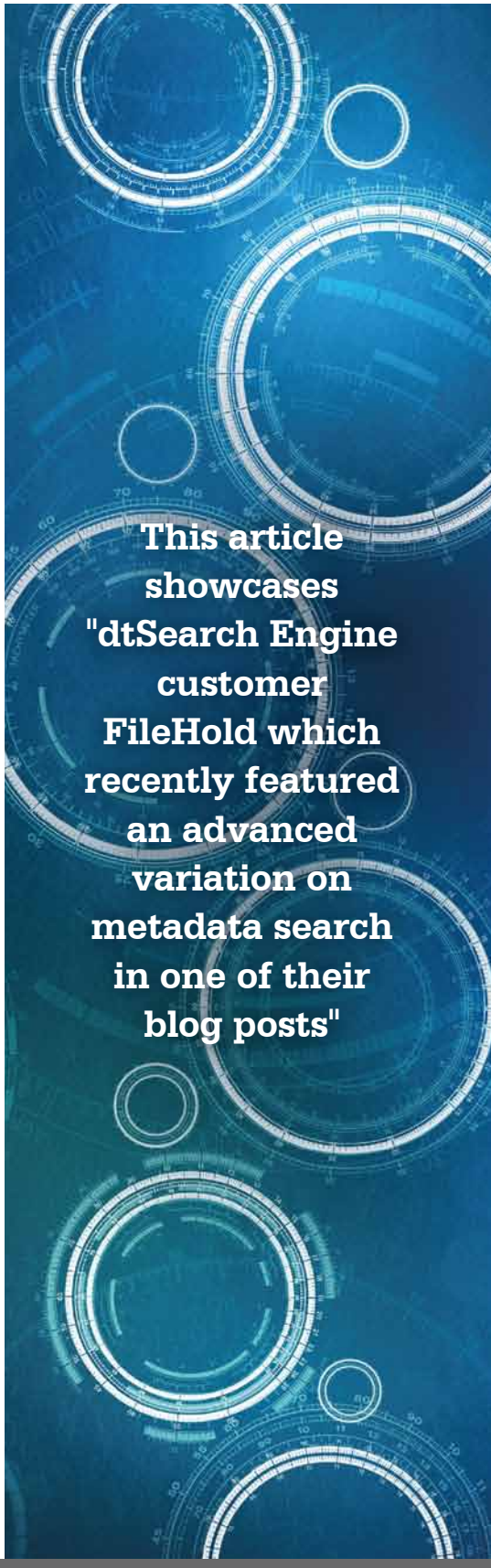
FileHold is full-featured document management software, including such options as document scanning, capture, indexing, storage, versioning, workflow, collaboration, audit logging and tracking, records management and lifecycle, individual user-defined security, information governance customization and other features. FileHold can run on a server installed "on premises" or can run in the cloud under Microsoft Azure. In either capacity, FileHold provides the security and efficiency its customers need. And indeed, FileHold's customers include some huge names in the telecom industry, in electronics, in manufacturing and many more industries as well as government agencies.

Sounds like a truly impressive customer list.

It is. And I encourage listeners to visit FileHold for more details on that. But their CTO also just featured a new blog post highlighting their "on the fly" dynamic metadata search.

What does that do?

Take the example I cited above, of limiting the *ProjectX and ProjectY and San Francisco and not Cleveland search to subject field contains sensitive office memo*. That works if every file has standalone subject field with key metadata like *sensitive office memo*. But sometimes key metadata can be buried in a file itself, instead of in a specific field.



Reprinted with permission of The Times USA, thetimesusa.com | Original article link

Can you provide an example?

The blog post cited an example from a collection of scientific papers that have their own internal structure with parts like Abstract, Categories and Subject, General Terms, Keywords, and Introduction. In the example, these sections are in the full-text of the files, as opposed to in structured metadata. The goal is to use these section markers as attributes in a search.

How would that work?

Quoting from the FileHold blog post here:

> A segmented search will allow us to create a field on-the-fly using the regular structure of the document. In this case, we know that the keyword we are looking for will be between the heading "KEYWORDS" and the heading "INTRODUCTION". If we are looking for documents with the keyword "sort", we could form our query like (KEYWORDS to INTRODUCTION) contains SORT.

In other words, this type of dynamic metadata search uses the structure of the scientific papers themselves to come up with an "on the fly" dynamic metadata element.

How would you summarize the take-away?

What this does is extend the advantages of adding a metadata element to a full-text search request in winnowing search results to files that may effectively, but not formally, contain such metadata elements.

Obviously, a lot of enterprises rely on Microsoft in some form or another. Can you explain how FileHold works with Microsoft products and platforms?

FileHold is a Gold level partner, the highest level of partnership Microsoft provides. FileHold is 100% based on the Microsoft platform and FileHold is committed to releasing products that harmonize and complement Microsoft technology including Microsoft SharePoint and Microsoft Office. As mentioned before, the FileHold cloud-based option runs on Microsoft Azure for the highest level of performance and security.

How does an enterprise or government agency "get started" with FileHold?

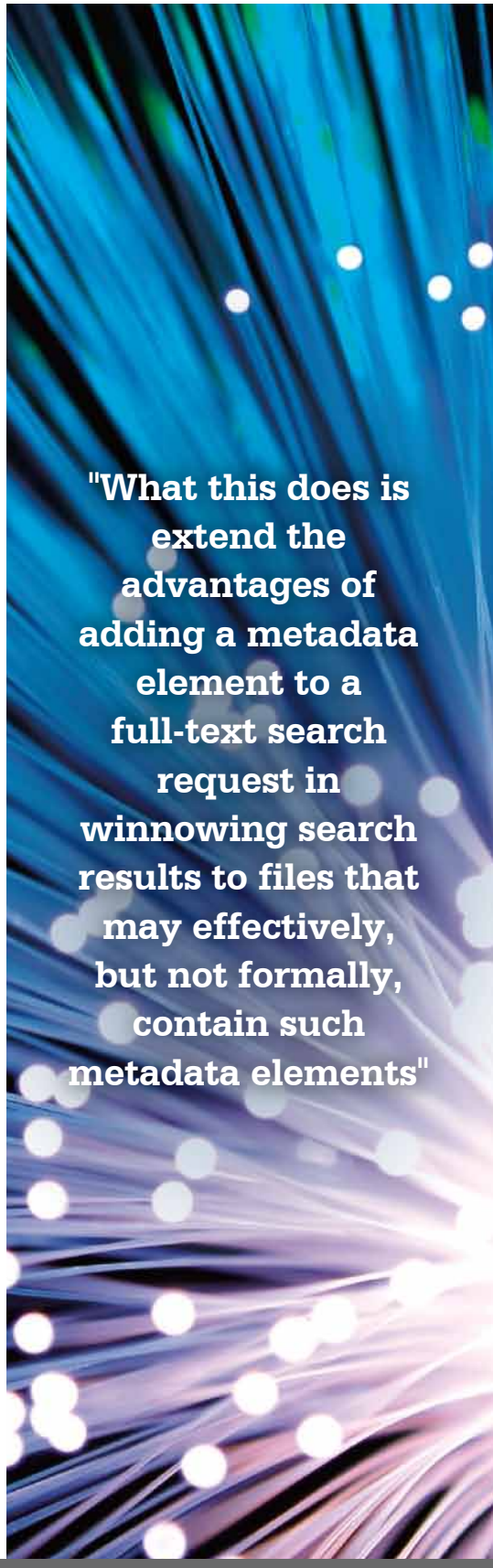
FileHold has no-charge evaluation options, etc. Anyone can go to FileHold for more details on evaluation and also to see a full list of available features.

Anything else you'd like to add?

Please visit dtSearch anytime with any questions, and to see the FileHold write-up on the dtSearch site. And of course please see FileHold for your document management needs.

About dtSearch® dtSearch has enterprise and developer products that run "on premises" or on cloud platforms to instantly search terabytes of "Office" files, PDFs, emails along with nested attachments, databases and online data. Because dtSearch can instantly search terabytes with over 25 different search features, many dtSearch customers are Fortune 100 companies and government agencies. But anyone with lots of data to search can download a fully-functional 30-day evaluation copy from dtSearch.com

Article contributed by dtSearch®



Reprinted with permission of The Times USA, thetimesusa.com | Original article link