

Liberating Your Summer With Text Retrieval

You already know that technology can be liberating. You don't need to fly across the country to meet with a client; you can Zoom instead. Now, liberate your summer further with instant full-text search.

With full-text search, if you need to locate a specific client file in terabytes of archives, you can instantly do that. Or if you need to double-check a specific client email but are a little hazy on the relevant year, no problem. Or if you want to make sure that there are no stray credit card numbers sitting on your network, you can get that peace of mind.

The secret to getting instant full-text search is using text retrieval software to first build an index across all relevant data. With the text retrieval software dtSearch®, for example, each index can hold up to a terabyte of data, and there are no limits on the number of indexes that you can create and that you and your colleagues can simultaneously search.

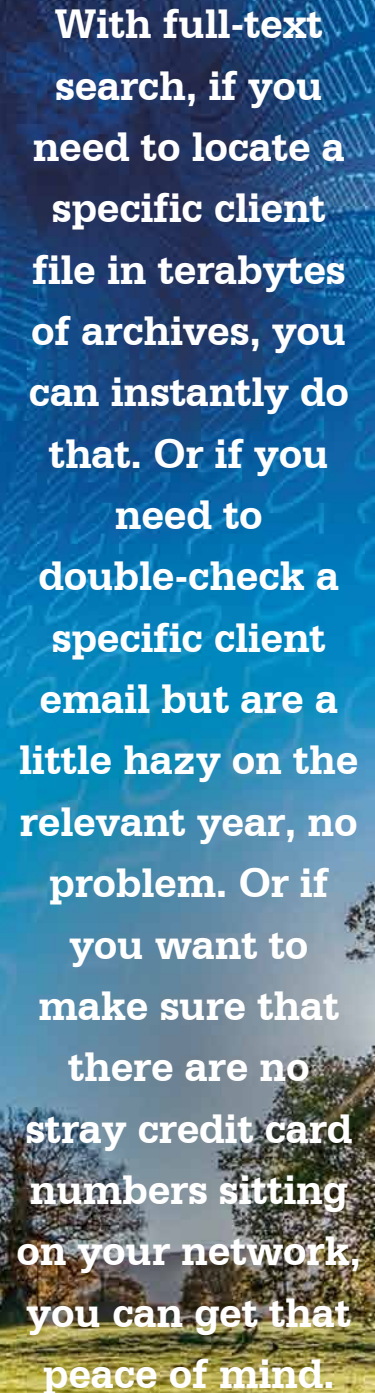
Indexing couldn't be easier. Just point to the directories you want to cover in your index, and the software will do the rest. The index can cover almost anything: PDF files; Microsoft Word, Access, Excel, PowerPoint and OneNote files; compressed archives like ZIP or RAR; even popular email formats like Outlook and Exchange. The software will automatically recognize which file formats it is working with.

If your files have "mismatched" extensions like PDF files saved with .DOCX extensions, no problem. The software uses the information inside the binary format itself to figure out the file type, not the extension. Even multilevel embedded documents are no problem for the software. If you have an email with a ZIP file and inside is an Access database and embedded within that is an Excel spreadsheet, the software will figure all of that out.

After indexing search time even across terabytes should be instantaneous. While indexing is resource intensive, concurrent searching can operate in a completely stateless manner. This means that you can have an unlimited number of search threads going at once without affecting performance. Concurrent searching can also continue while you update your indexes to account for new files and emails.

At this point, you and your coworkers can do just about any type of searching imaginable. You can enter a natural language search request like get me the Jones and Smith files from 2020. That type of search will find all files with matching search terms and relevancy rank them based on hit term density and rarity. In other words, if the word files is all over your data, but Smith and Jones are a lot less frequent, then documents with Smith and Jones and particularly denser mentions of Smith and Jones, will get a much higher relevancy ranking letting you jump right into what you are looking for.

Article contributed
by dtSearch®



With full-text search, if you need to locate a specific client file in terabytes of archives, you can instantly do that. Or if you need to double-check a specific client email but are a little hazy on the relevant year, no problem. Or if you want to make sure that there are no stray credit card numbers sitting on your network, you can get that peace of mind.

You can also enter a more structured phrase / Boolean-type search such as: (Alfred Jones or Mike Smith) and 2020 and not alphabet soup. Or you can add a proximity element, with w/ representing a word range such as: (Alfred Jones w/28 Mike Smith) and 2020 and not alphabet soup. Or you can do a directed proximity search, finding Alfred Jones but only if it comes within 35 words before Mike Smith. Or you can specify an additional metadata component like email sender and limit a search request or part of a search request to just that.

You can add in other search options on top of these basic searches as well. If you activate stemming, the search will find different variations on the same root word, like file, files, filing, filed. Concept searching uses a built-in English language thesaurus and/or your own custom synonym rings so if CompanyX has a street name of SpeedyFastCo, you can find those interchangeably. Fuzzy searching will even find a word that is slightly misspelled. If Smith is mistyped as Smich in an email, you can still find that with a low level of fuzzy searching.

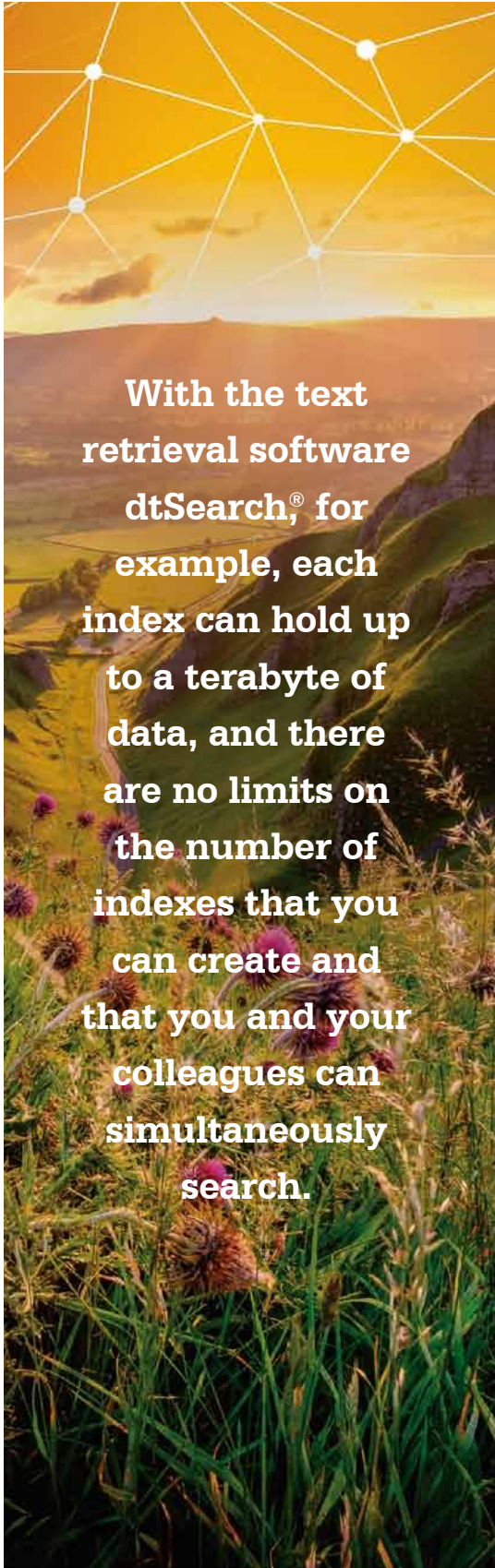
Multi-language text is no problem. The software has Unicode support covering European languages, double-byte character languages and right-to-left Middle Eastern languages. You can even search ancient Egyptian hieroglyphics with Unicode. And the same document can switch back and forth across multiple languages without affecting searching.

Beyond that, you can pile on more advanced search options as needed. For example, you can employ numeric range searching or date range searching, including automatic recognition of multiple formats of the same date, such as July 4, 2018 and 7/4/18. Or you can go through all of the data looking to see if it contains any valid credit card numbers.

After a search, you can see matching files with highlighted hits, jumping from "hit" to "hit" as you browse through full copies of the documents. Suddenly, you have a lot of free time ... at least until the fall crunch sets in.

Elizabeth Thede of dtSearch is director of sales. The company offers enterprise and developer products running "on premises" or in the cloud to instantly search terabytes with over 25 search options. dtSearch's own document filters support files, emails, databases and web data.

Article contributed
by [dtSearch®](#)



**With the text
retrieval software
dtSearch®, for
example, each
index can hold up
to a terabyte of
data, and there
are no limits on
the number of
indexes that you
can create and
that you and your
colleagues can
simultaneously
search.**