

For Those of Us Who'd Rather Avoid Spring Cleaning Their Data

Some people embark on spring cleaning with zeal. Others shudder at the thought. If you're mired in emails, PDFs, "Office" files and the like and hate the idea of spring cleaning data, consider enterprise search. A key point to spring cleaning is to get a handle on the clutter so you can quickly find what you're looking for going forward. But with enterprise search, there's no need to spring clean to get instant concurrent searching across terabytes. And enterprise search has over 25 different simultaneous-user search options, so everyone can jump directly to the right information.

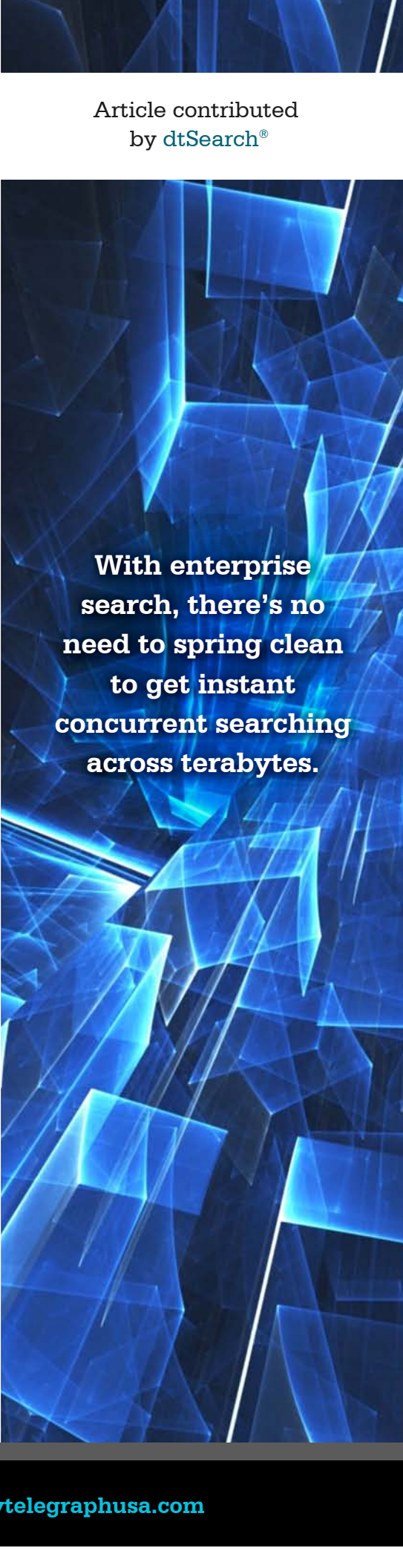
How does enterprise search work?

Enterprise search offers instant concurrent searching across terabytes only after it first indexes the data. Indexing is a lot of work for enterprise search, but no work for you. In dtSearch® for example all you need to do is point to the folders where the data resides, actually or virtually, and enterprise search will do the rest. The data can span different email archives, files on the file server or in Office 365, SharePoint attachments, etc.

So enterprise searching can cover remote data as well as local files?

As long as enterprise search like dtSearch can see local and remote files as a part of the Windows folder system, enterprise search can index and search them. No need to even identify various file types. The indexer on its own will identify whether each item is a PDF, Word document, Access database, Excel spreadsheet, PowerPoint, OneNote file, email, etc. The indexer will even work if files are in a compression archive like ZIP or RAR or have incorrect file extensions. PDFs can have .DOCX extensions and Word documents can have .PDF extensions, and it won't matter.

Article contributed
by dtSearch®



With enterprise search, there's no need to spring clean to get instant concurrent searching across terabytes.

What about capacity?

With dtSearch, a single index can hold up to a terabyte of text and there are no limits on the number of terabyte indexes the software can create and instantly search. And enterprise search has a deep reach. For example, the software can retrieve obscure metadata that would be very hard to run across clicking around a file in its native application. It can also find camouflaged text like black writing against a black background. And it can drill down multiple layers, like an email with a ZIP or RAR attachment that includes an Excel spreadsheet with a Word document embedded inside.

What about speed?

Relative to search, indexing is definitely the slower part of the operation. But after indexing, search is typically instantaneous. And enterprise search optimizes its indexes for concurrent access, allowing multiple end-users to instantly query the data at the same time without a delay in response. As data evolves, enterprise search can automatically update its index or indexes without impeding continuing concurrent searching.

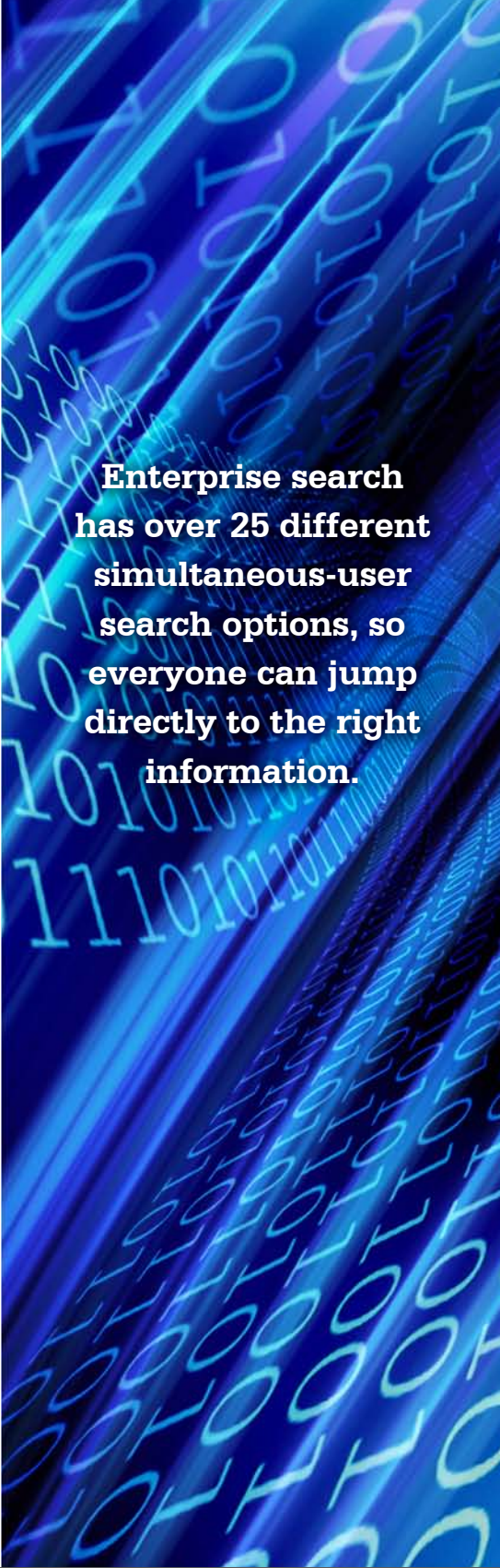
What are enterprise search options?

When you are going file by file looking for something, it is usually on an "I'll know it if I see it" basis. But an enterprise search query can draw from over 25 different search options. On the most basic level, an Any Words or All Words search can look for words or phrases anywhere across the data. Or a search can specify specific combinations of words and phrases in various Boolean and/or/not or proximity configurations like *obscure reference and (odd phrasing or word usage not w/27 words of dull ordinary)*. Search elements can cover the full-text of files, or home in on specific metadata, such as taking the just-mentioned search request and adding an additional requirement that Sender metadata must include *Albert Jones*.

And other search options?

Fuzzy searching adjusts from 0 to 10 to sift through minor typographical errors like *obscare* for *obscure* that can occur in emails or with scanned and OCR'ed copy. Concept searching can find synonyms like *unknown* for *obscure*.

Article contributed
by dtSearch®



**Enterprise search
has over 25 different
simultaneous-user
search options, so
everyone can jump
directly to the right
information.**

Numeric or numeric range search searching can locate specific numbers or number ranges across all text or in specific metadata. Date searching can look for full-text or metadata mentions of a date or date range like *date(May 1, 2023 to February 15, 2024)*—even picking up common variants in that date range like both *Jan 7, 2024* and *1/7/24*. For forensics-oriented work, the software can generate hash values across all indexed files and optionally search for specific hashes. The software can further identify any credit card numbers across indexed data.

I know you talk a lot about Unicode support.

The Unicode standard spans hundreds of international languages, and enterprise search works with Unicode text. A single file or email can go from English to Greek to a double-byte character language like Chinese, Japanese or Korean to a right-to-left language like Hebrew or Arabic and then back to English and the Unicode standard and enterprise search will cover all of that. dtSearch can even search for Unicode emojis 😊

How does search sorting work?

By default, enterprise search ranks retrieved files using vector-space relevance. Take an Any Words search for *Texas*, *Virginia* or *Nebraska*. If *Texas* and *Virginia* are prevalent across indexed data with much sparser mentions of *Nebraska*, then *Nebraska* would get a higher relevancy rank. Files with the densest *Nebraska* mentions would go to the top of the list. But an end-user can also override the defaults and customize relevancy-ranking, like giving *Texas* a positive weight of 7, *Virginia* a positive weight of 4 and *Nebraska* a negative weight of 6 for a full-text mention but a positive weight of 9 for appearances in Subject metadata or at the top or bottom of a file. Regardless of the sorting, enterprise search can show a full copy of retrieved files with highlighted hits for convenient review.

Final thoughts?

Go ahead and ditch the spring cleaning, at least of your enterprise data. You'll find fully-functional 30-day evaluation downloads at [dtSearch.com](https://dtsearch.com) so everyone at once instantly search across terabytes—no spring cleaning required.

Article contributed
by dtSearch®

**The Unicode
standard spans
hundreds of
international
languages, and
enterprise search
works with
Unicode text.**