

Enterprise Data Still Needs Enterprise Search

Article contributed
by dtSearch®

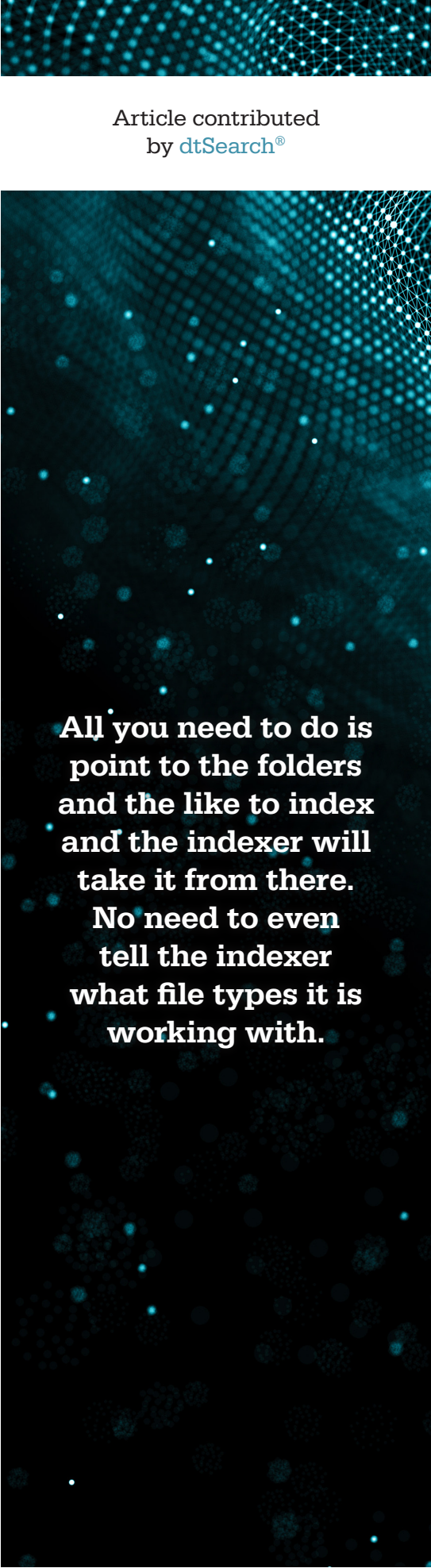
Whatever else you may have in your toolbox for organizing data, you still need enterprise search. Nothing matches the deep dive enterprise search can provide through every corner of an organization's content. Enterprise search like dtSearch® offers instant multithreaded concurrent searching spanning terabytes running from a traditional network, a local web server, or the cloud like Azure or AWS. However it runs, enterprise search enables instant concurrent searching across terabytes only after indexing the data.

How do you get enterprise search to index the data?

All you need to do is point to the folders and the like to index and the indexer will take it from there. No need to even tell the indexer what file types it is working with. The indexer on its own can figure out the exact mix of PDFs, word processing documents, spreadsheets, databases, presentation files, note files, compressed files, emails plus attachments, etc.

What about remote files?

So long as files appear in the Windows folder system, the indexer can handle any mix of local and remote data like DropBox files, SharePoint attachments and Office 365 documents. Importantly, indexing doesn't move around or otherwise alter the original data. The indexer doesn't even need to retrieve files in their associated applications. Instead, indexing goes directly to the binary formats of files. Accessing the binary formats alone, the indexer parses each file to identify all specific words and numbers and record the location of each in the data.



All you need to do is point to the folders and the like to index and the indexer will take it from there. No need to even tell the indexer what file types it is working with.

What about capacity?

A single dtSearch index can hold up to terabyte of text covering both local and remote files, and there are no limits on the number of indexes that the software can create and end-users can instantly and concurrently search. While indexing is resource-intensive, multithreaded searching is resource-light. In fact, automatic index updates can accommodate new files, deleted files and edited files while concurrent searching continues uninterrupted.

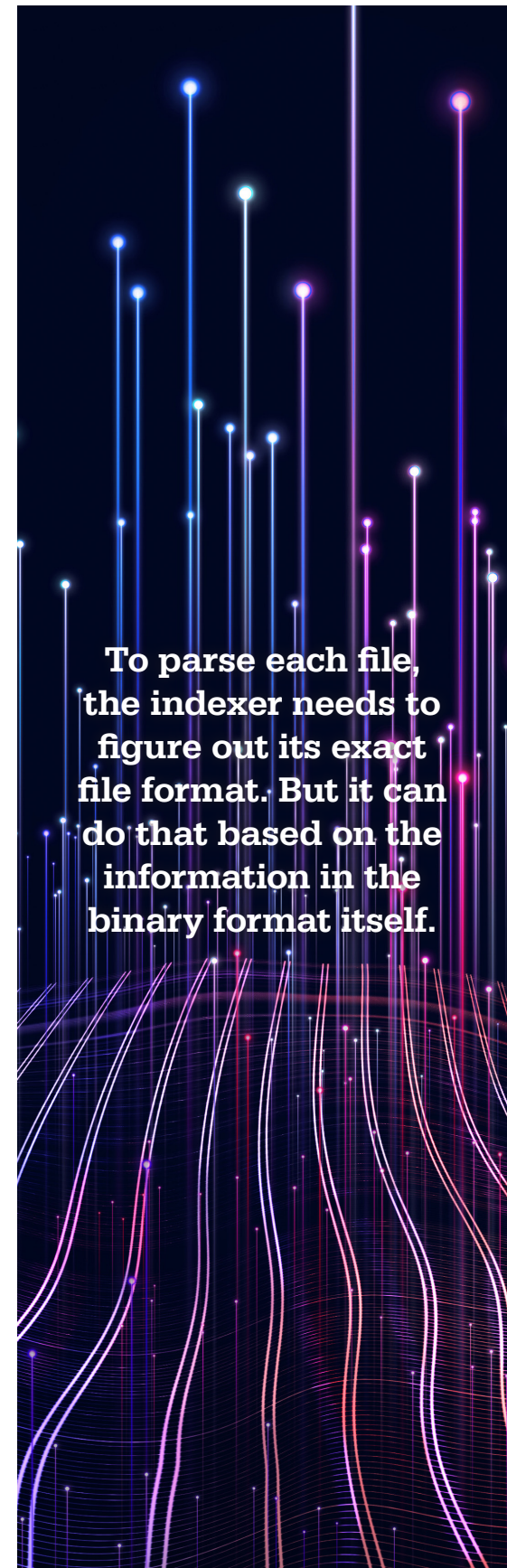
So the whole process can proceed automatically?

That's the idea. To parse each file, the indexer needs to figure out its exact file format. But it can do that based on the information in the binary format itself. It won't even matter if a file has a mismatched file extension like a Word document saved with a .PDF extension or a PowerPoint saved with a .DOCX extension. Indexing can also work automatically with both individual files as well as recursively nested files. The data can include an email with a ZIP or RAR attachment including a Word document recursively embedding an Excel spreadsheet and the binary format access will let the indexer pick up all of that.

And metadata?

File metadata, even very obscure metadata that you might miss clicking around a document in its associated application, is readily available in the binary format for the indexer to process. The indexer can also retrieve items like tracked changes or redactions where the original text persists in a file, even if that text would not ordinarily appear in the file's associated application. The indexer can further find text that someone may have actively tried to hide like white text against a white background or blue text against a blue background. And the indexer can also identify image-only PDFs requiring OCR through a program like Adobe Acrobat for full-text processing.

Article contributed
by [dtSearch®](#)



**To parse each file,
the indexer needs to
figure out its exact
file format. But it can
do that based on the
information in the
binary format itself.**

Can the indexer automatically handle text in different languages?

Current file formats all use Unicode to encode hundreds of international languages. A file or an email can proceed through any number of Unicode encodings: English, other European languages, right-to-left languages like Hebrew and Arabic, double-byte text like Chinese, Japanese or Korean, and other international languages. Unicode will track all of that, enabling enterprise search to identify the Unicode encodings from binary format access.

What types of search features does dtSearch offer?

dtSearch has over 25 different search options. These include natural language search request processing as well as handling highly structured phrase, Boolean (and/or/not) and proximity search requests. Search across the full text of everything or require that certain search elements appear in specific metadata. Concept searching extends a search request to similar words or concepts. Fuzzy searching adjusts from 1 to 10 to pick up potential typographical or OCR errors like toolbox for toolbox.

What about numbers?

Numeric range searching looks for individual numbers or numbers in a range like *18* to *36*. Date and date range searching operates automatically across popular date formats, enabling a date range search covering *November 30, 2024* to *October 9, 2025* to retrieve not only *January 15, 2025* but also *Jan 15 2025* and *1/15/25*. dtSearch can even identify any valid credit card numbers across indexed data.

And relevancy ranking?

dtSearch can automatically relevancy rank search results. Say you search for *red*, *blue* or *purple*. If *red* and *blue* are common across indexed files but *purple* relatively rare, then *purple* mentions will receive

Article contributed
by [dtSearch®](#)

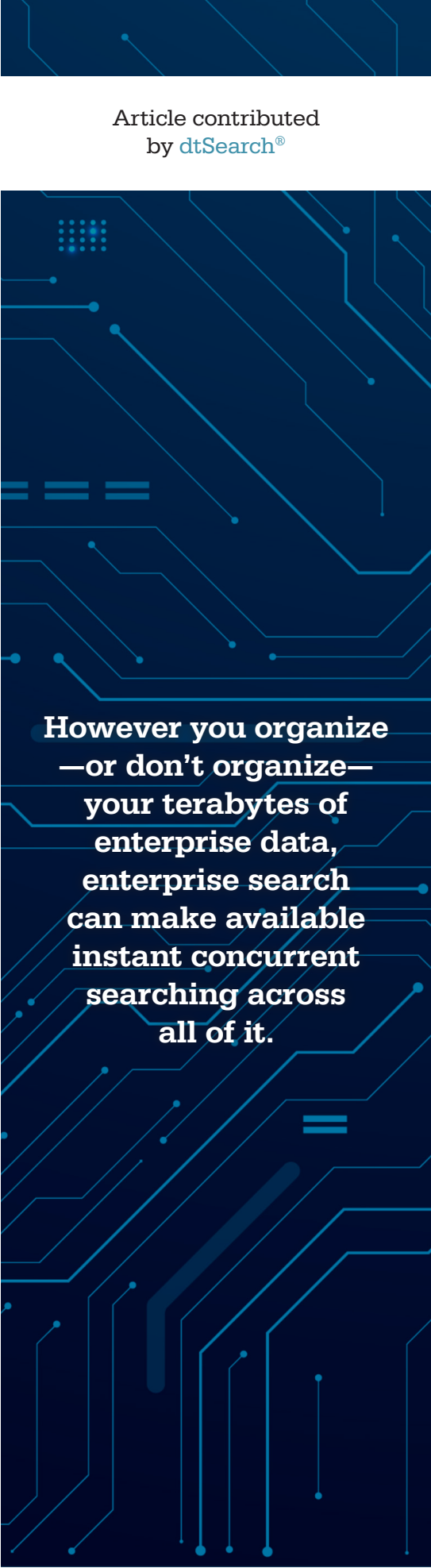
dtSearch has over 25 different search options. These include natural language search request processing as well as handling highly structured phrase, Boolean (and/or/not) and proximity search requests.

a higher relevancy rank, with documents with the densest *purple* mentions coming out on top. dtSearch also supports user-defined variable term weighting, like giving *red* a positive weight of 7, *blue* a negative weight of 3 and *purple* a positive weight of 8 but only if it appears in certain metadata or positionally towards the top or bottom of a file. For a different view of search results, instantly re-sort by a new metric like file date or file location. Whatever the sorting, browse through retrieved files with highlighted hits for easy navigation.

Final thoughts?

However you organize—or don't organize—your terabytes of enterprise data, enterprise search can make available instant concurrent searching across all of it. Get started at [dtSearch.com](https://dtsearch.com) with fully-functional 30-day evaluation downloads.

Article contributed
by [dtSearch®](https://dtsearch.com)



**However you organize
—or don't organize—
your terabytes of
enterprise data,
enterprise search
can make available
instant concurrent
searching across
all of it.**