

A Haunted House Guide to Enterprise Data

Article contributed
by [dtSearch®](#)

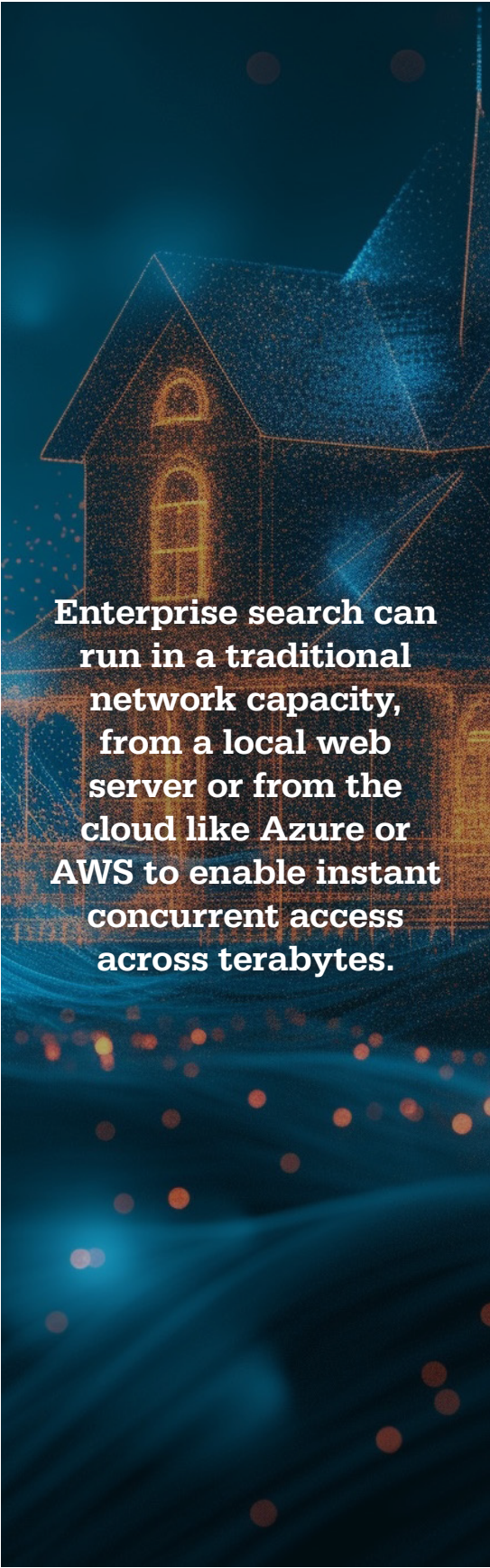
Enterprise data should be like a sunny atrium with beams of light guiding end-users instantly to the right information. But sometimes enterprise data can feel less like a sunny atrium and more like a haunted house with cobwebbed crannies and trapdoors. Enterprise search can get you out of the cobwebs and trapdoors and back to the sunbeams.

How enterprise search works. Enterprise search can run in a traditional network capacity, from a local web server or from the cloud like Azure or AWS to enable instant concurrent access across terabytes. But it does so only after first indexing the data. Indexing might sound like a lot of effort. But with [dtSearch®](#) for example all you need to do is check off the folders and the like to cover, and the software will do the rest.

To efficiently get through terabytes, the indexer bypasses pulling up each file in its associated application and instead goes directly to each file's binary format. To parse that correctly, the indexer needs to identify the exact right file type—PDF, Microsoft Word, Excel, Access, PowerPoint, OneNote, email formats, etc. But the indexer can do this identification on its own using the binary format. It doesn't even matter if a file has a misleading file extension like a Word document saved with a .PDF extension or a PDF with a .DOCX extension.

The indexing process is seamless in other respects as well. So long as the indexer can see folders as part of the Windows folder system, [dtSearch](#) can work interchangeably with both local and remote files like Office365, SharePoint or DropBox. The indexer can also automatically work with multilevel formats like an email with a ZIP or RAR attachment containing a Word document with an Excel spreadsheet embedded inside.

In terms of capacity, a single [dtSearch](#) index can hold up to a terabyte of text. There are no limits on the number



Enterprise search can run in a traditional network capacity, from a local web server or from the cloud like Azure or AWS to enable instant concurrent access across terabytes.

of indexes that the software can create and end-users instantly concurrently search. For changing data, such searching can continue while indexes automatically update to account for new files, edited files and deleted files.

On crannies and trapdoors. Now enterprise data can still contain some cobwebbed crannies and trapdoors. For example, certain metadata can be very hard to spot in a file's associated application. You can click and click around and still miss it. But all metadata appears in the binary format and is thus available to enterprise search.

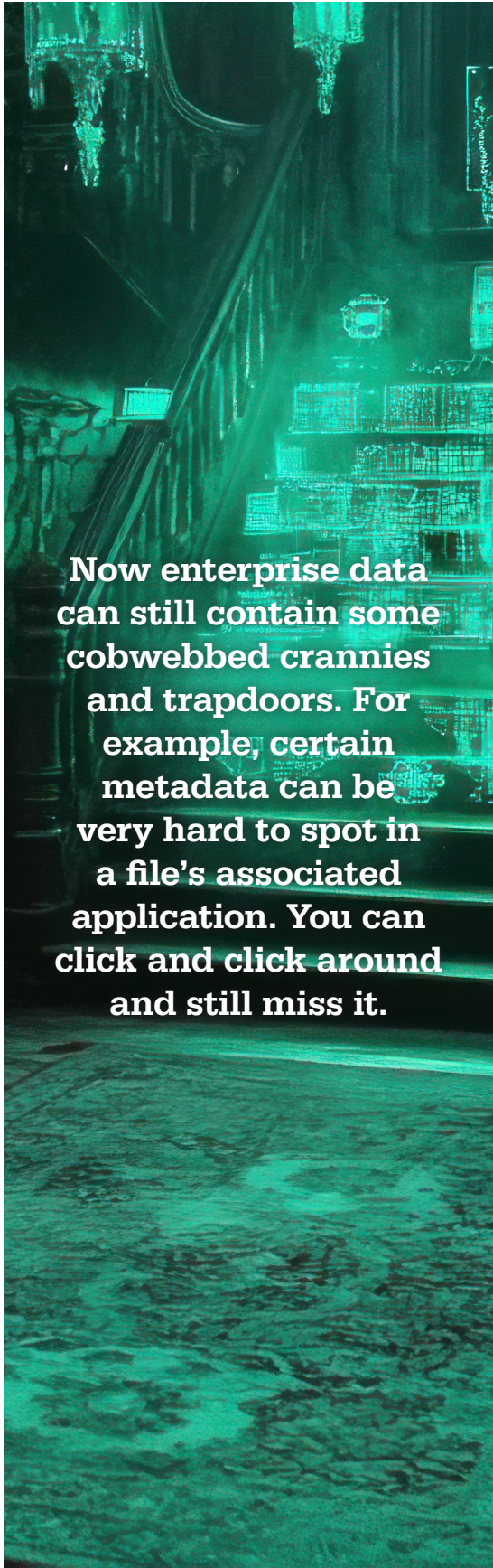
Another example would be camouflaged text like black writing against a black background or orange writing against an orange background. Such camouflaged text is not easy to spot inside a file's associated application. In fact, someone may have inserted such camouflaged text precisely to avoid detection from inside a file's associated application. But all text, camouflaged or otherwise, is on the same footing in a file's binary format, allowing enterprise search to shine a sunbeam on it.

Some redaction programs also put a black rectangle over text when marked for redaction. But the text itself can remain in the file even though you can't see it under the black rectangle. Track changes can further make it seem like certain deletions are no longer present. However, without full acceptance of the changes, the deletions will remain, even if not visible by default. Through its binary format access, enterprise search will locate both redacted text and other text marked for deletion just like any other text in files.

As a last example, you know when you are looking at a PDF file inside a PDF viewer and you try to copy and paste some text from the PDF and nothing copies out? That is likely an "image only" PDF. "Image only" PDFs are often mixed interchangeably with regular text-based PDFs in folders. But dtSearch can flag "image only" PDFs during indexing to let you know that you need an OCR program like Adobe Acrobat to process these before returning them to the dtSearch indexer and the sunny atrium of full-text searchable data.

Search options to illuminate your data. dtSearch has

Article contributed
by [dtSearch®](#)



Now enterprise data can still contain some cobwebbed crannies and trapdoors. For example, certain metadata can be very hard to spot in a file's associated application. You can click and click around and still miss it.

over 25 different full-text and metadata search options to cast light on your data. Enter a free-form natural language query or a highly structured phrase, Boolean (and/or/not) and proximity search request. With concept searching, a search for *cobweb* would find the synonym *gossamer*. Fuzzy searching adjusts from 0 to 10 to sift through typographical and OCR errors like *cobveb* for *cobweb*.

dtSearch works not only with English text but also any of the hundreds of international languages that the Unicode standard supports. A single file or email can go from a European language and alphabet to double-byte Chinese, Japanese or Korean, to right-to-left text like Hebrew or Arabic. Unicode and dtSearch will track all of that.

Along with word-based searching, dtSearch supports number and numeric range searching. Date and date range searching can automatically extend across popular date formats enabling a search for *date(10/20/25 to 1/5/26)* to pick up not only *10/31/25* but also *October 31 2025* and *Oct 31 2025*. dtSearch can even pick up credit card numbers in indexed data. Following a search, dtSearch has multiple options for relevancy-ranking and other sorting—or instant re-sorting—of search results.

One last sunbeam in the enterprise data atrium: dtSearch displays retrieved files and other search results with highlighted hits for easy navigation. So turn your enterprise data haunted house into an instantly concurrently searchable sunny data atrium. Visit dtSearch.com for fully-functional 30-day evaluation downloads.

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



dtSearch works not only with English text but also any of the hundreds of international languages that the Unicode standard supports. A single file or email can go from a European language and alphabet to double-byte Chinese, Japanese or Korean, to right-to-left text like Hebrew or Arabic.