

You might be done with last year's data; it might not be done with you

You close out one year, looking for a fresh start on the next. But old content doesn't just disappear when you hang up a new calendar. There's always a chance of something in previous data reemerging to challenge the plans you have for the upcoming year. While nothing can completely counter that possibility, enterprise search can help keep tabs on all information, past and present, to mitigate such a risk.

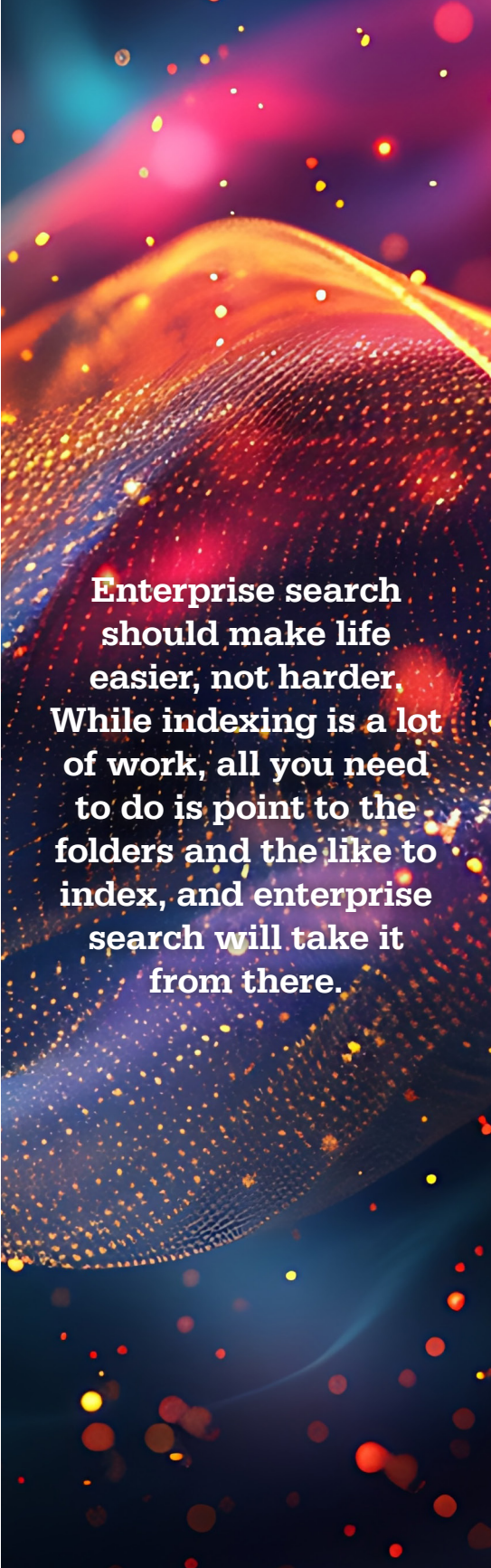
Enterprise search enables instant concurrent searching across terabytes after first indexing the data. A single index can hold up to a terabyte, and there are no limits on the number of indexes enterprise search can create and end-users simultaneously query.

Enterprise search should make life easier, not harder. While indexing is a lot of work, all you need to do is point to the folders and the like to index, and enterprise search will take it from there. In fact, the indexer works automatically with not only local data, but also remote data, like DropBox, SharePoint attachments and OneOffice files, that present through the Windows folder system.

To correctly parse a file, the indexer needs to pinpoint its file format. Fortunately, the indexer can do this on its own through the binary format of each item, figuring out if it is a word processing document, a spreadsheet, a presentation file, a database file, a PDF, etc., plus the applicable file version. This works with nested files as well. An email can have a ZIP or RAR attachment with a Word document and a recursively embedded Excel spreadsheet and the indexer will support all those layers of text and metadata.

The structure of the index enables multiple end-users to instantly query the data without affecting each other's

Article contributed
by dtSearch®



Enterprise search should make life easier, not harder. While indexing is a lot of work, all you need to do is point to the folders and the like to index, and enterprise search will take it from there.

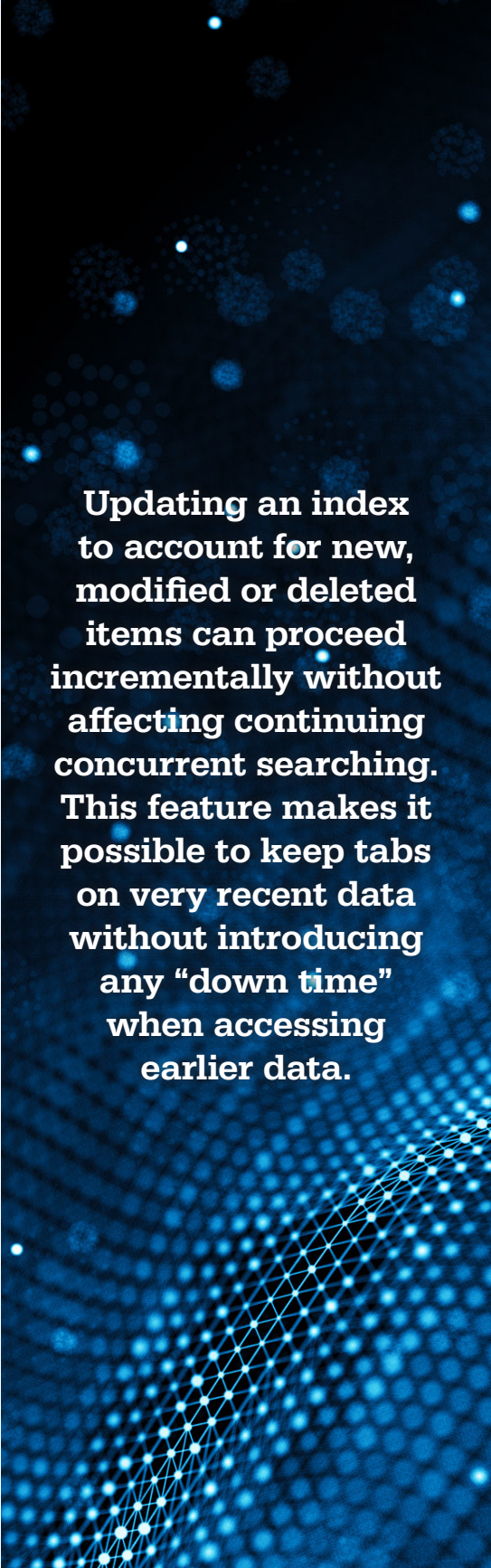
search threads. Searching can proceed in a classic Windows network environment, from an on-premises Intranet server, or from the cloud such as Azure or AWS. Updating an index to account for new, modified or deleted items can proceed incrementally without affecting continuing concurrent searching. This feature makes it possible to keep tabs on very recent data without introducing any “down time” when accessing earlier data.

Over 25 search options make it hard for data to hide. For example, enter *auld lang syne* as a natural language “any words” search request to find items mentioning even just one of these terms. Enter *auld lang syne* as a natural language “all words” search request to locate only items with all 3 terms. Or enter *auld lang syne* as a phrase search to retrieve just items with this exact phrase. Search across all full text content or limit a search to specific metadata.

Enterprise search can also handle complex Boolean and proximity formulations: (*auld lang syne and new year's eve*) *w/31 (champagne or confetti) and not party poppers*. Concept searching can pick up *bubbly* for *champagne*. Fuzzy searching adjusts from one to 10 to sift through OCR or typographical mistakes like *champain* for *champagne*. Add on a specific date like *date(December 31, 2024)* or a date range like *date(December 26, 2024 to January 3, 2025)*. Both will pick up common variants such as *Dec. 31, 2024* or *12/31/24*. Or add on a specific number or a numeric range.

Credit cards that remain in files are a constant problem with aging data. Fortunately, enterprise search has a credit card recognition feature to flag any valid credit card numbers across data. Enterprise search can even find data that someone may have intentionally hidden in a file like black text against a black background or white text against a white background. And enterprise search will retrieve super-obscure metadata that you might otherwise never spot.

For multilingual data, enterprise search works with Unicode covering hundreds of international languages. A file can start in English, cycle through multiple



Updating an index to account for new, modified or deleted items can proceed incrementally without affecting continuing concurrent searching. This feature makes it possible to keep tabs on very recent data without introducing any “down time” when accessing earlier data.

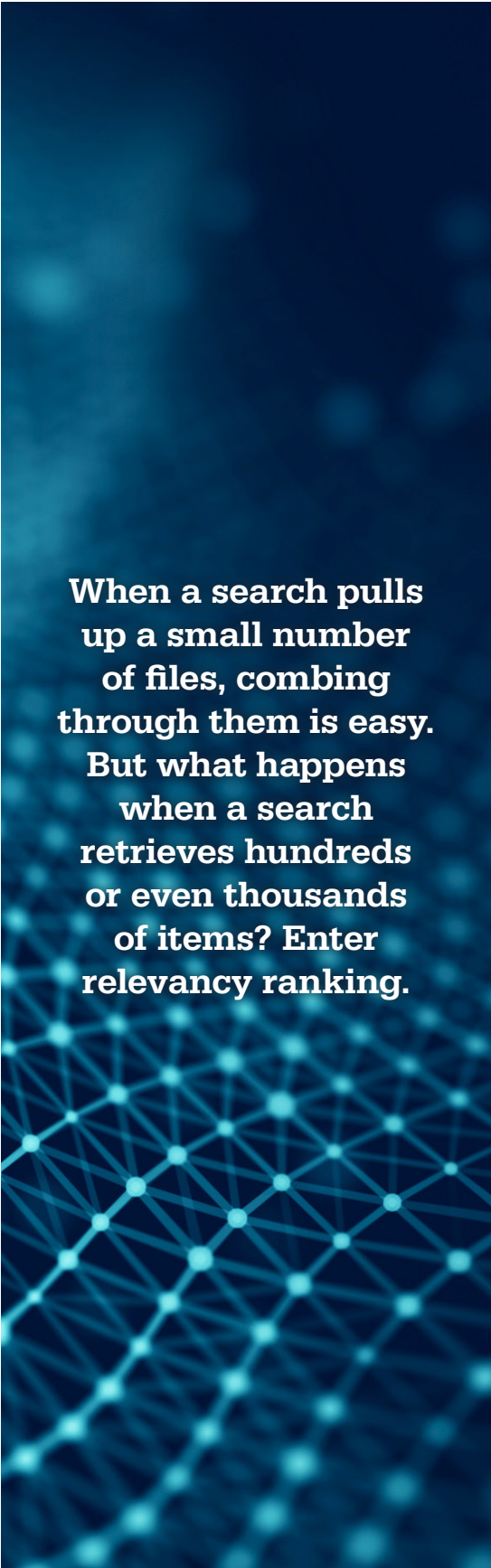
other European languages, right-to-left languages and double-byte Asian text, then back to English. Unicode and enterprise search will track the entire progression.

When a search pulls up a small number of files, combing through them is easy. But what happens when a search retrieves hundreds or even thousands of items? Enter relevancy ranking. Take an “any words” search for *auld lang syne*. If *auld* and *lang* are all over the indexed data but *syne* is comparatively rare, *syne* will get a higher relevancy rank with the densest *syne* files coming out on top, taking you to the most relevant hits.

Or customize relevancy ranking through positive or negative variable term weighting, giving *auld* and *lang* a positive weight of two and *syne* a negative weight of three with a positive weight of nine for mentions in specific metadata or near the top or bottom of a file. For a different perspective on search results, instantly re-sort by some completely different metric like file date or location. Whatever the sorting, view a full copy of retrieved items with highlighted hits for easy navigation.

So get ready to ring in the new year — and use enterprise search to keep tabs on not only this year’s data, but the data of *auld lang syne*.

Article contributed
by [dtSearch®](#)



When a search pulls up a small number of files, combing through them is easy. But what happens when a search retrieves hundreds or even thousands of items? Enter relevancy ranking.