

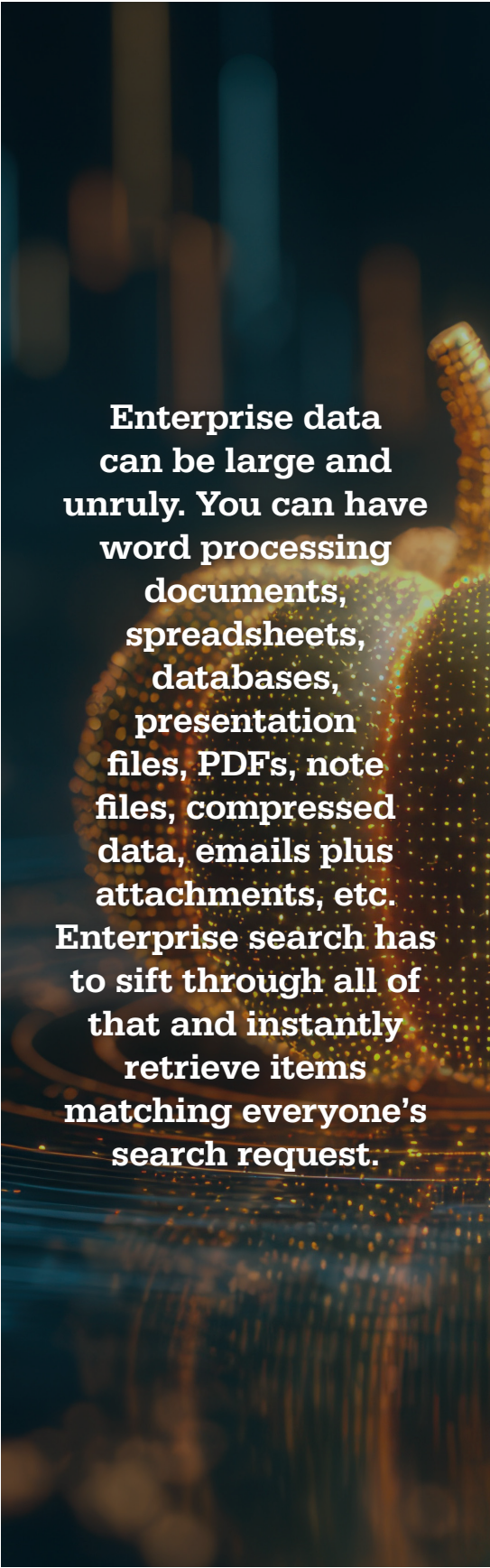
So Many Things Are Just out of Reach; Enterprise Data Shouldn't Be One of Them

So much is just out of reach. Is it pumpkin spice time at my favorite coffee shop? Still not in season; no pumpkin spice for me. But while so many things can seem out of reach, enterprise data shouldn't be one of them. Enterprise search lets everyone across the organization instantly and simultaneously find the right data. To maximize access, enterprise search can run in a classic network environment, from a local Internet or Intranet server, or from the cloud such as Azure or AWS.

Enterprise data can be large and unruly. You can have word processing documents, spreadsheets, databases, presentation files, PDFs, note files, compressed data, emails plus attachments, etc. Enterprise search has to sift through all of that and instantly retrieve items matching everyone's search request. Indexing is the secret ingredient that makes this possible. Indexing pre-processes each unique word and number across the dataset, recording the location of each word and number in the data. While labor-intensive for enterprise search, indexing is a piece of pumpkin spice cake for you. All you need to do is check off the folders and the like to cover and the indexer will take it from there.

When you normally review a file, you retrieve that file in its associated application like viewing a PDF inside Adobe Reader or a Microsoft Word document in Word. But pulling up each file in its associated application isn't feasible for the indexer. That would take way too long. Instead, the indexer needs to parse each file as it sits there in binary format. To do that, the indexer needs to apply just the right parsing specification for each specific file format. File format parsing specifications can be hundreds of pages long, so matching the right one to each file is essential for accurate text parsing.

Article contributed
by [dtSearch®](#)



Enterprise data can be large and unruly. You can have word processing documents, spreadsheets, databases, presentation files, PDFs, note files, compressed data, emails plus attachments, etc. Enterprise search has to sift through all of that and instantly retrieve items matching everyone's search request.

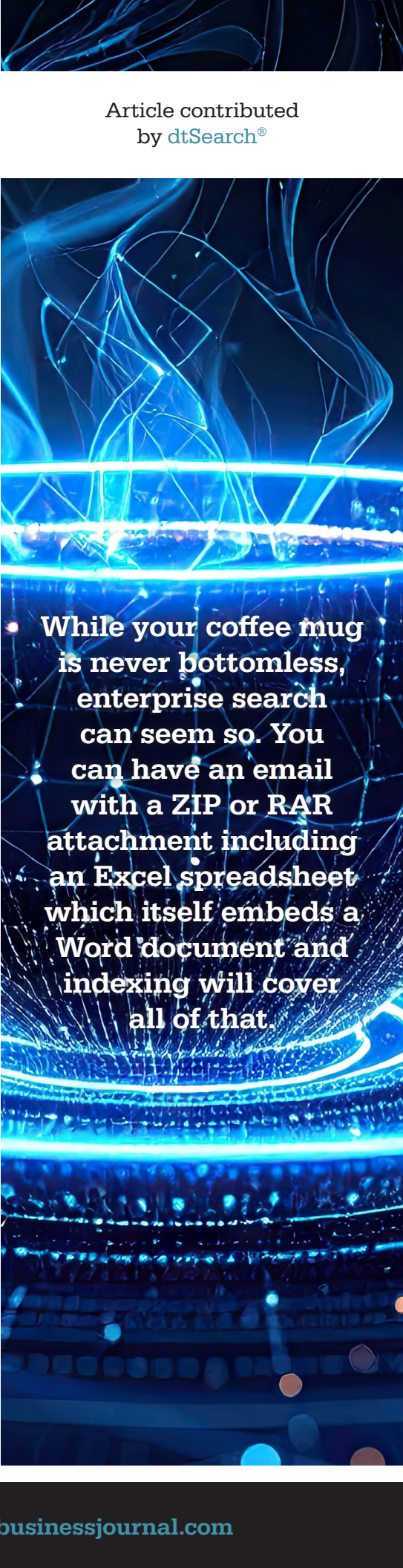
If you were trying to figure out a file type, you'd probably start with the file extension. A .PDF file extension would suggest a PDF. A .ONE extension would suggest a OneNote file. But that isn't sufficiently foolproof for the indexer as it is always possible for someone to save an Access database with a PDF extension or a PowerPoint with a .ONE extension. For accuracy, the indexer needs to look inside each binary format to figure out the parsing specification to apply regardless of the file extension.

As long as remote data like Office 365 documents, SharePoint attachments or DropBox files appear as part of the Windows folder system, the indexer can handle these just like local files. In terms of capacity, a single dtSearch index can hold up to a terabyte of text and there are no limits on the number of indexes that the indexer can build and the software simultaneously search. And searching can continue from a network, a local web server or the cloud while indexes automatically update to reflect new content.

While your coffee mug is never bottomless, enterprise search can seem so. You can have an email with a ZIP or RAR attachment including an Excel spreadsheet which itself embeds a Word document and indexing will cover all of that. Indexing will pick up all metadata in indexed files, including deep metadata that you might miss clicking around a file in its associated application. Indexing can also find text that might hide in an associated application because it blends in with its background color like mocha-colored text against a mocha-covered background. If track changes or other redactions are not fully accepted, enterprise search can still find text in phrases or paragraphs marked for deletion. Enterprise search can also tell you if you have PDFs that are image-only, requiring processing through an OCR program like Adobe Acrobat for full-text searching.

For international languages, enterprise search supports Unicode, the standard for modern data supporting hundreds of different international languages. You can have a single email or other file that goes from English to other European alphabets and text to double-byte Chinese, Japanese or Korean, to right-to-left text like

Article contributed
by dtSearch®



**While your coffee mug
is never bottomless,
enterprise search
can seem so. You
can have an email
with a ZIP or RAR
attachment including
an Excel spreadsheet
which itself embeds a
Word document and
indexing will cover
all of that.**

Hebrew and Arabic, and back to English and Unicode and enterprise search will follow all of that.

dtSearch has over 25 different search types. An “any words” search for *pumpkin spice* would look for any file containing even one mention of *pumpkin* or *spice*. An “all words” search for *pumpkin spice* would look for files including both *pumpkin* and *spice*. An “exact phrase” search for *pumpkin spice* would only retrieve files encompassing at least one instance of the exact phrase *pumpkin spice*.

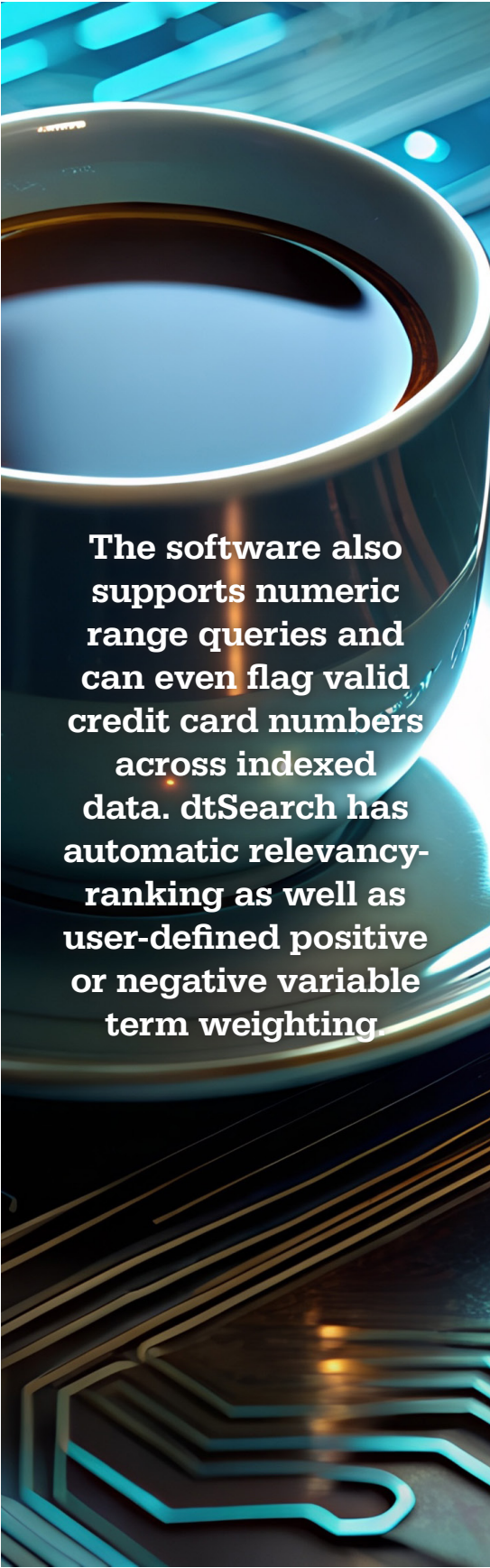
More advanced searchers can enter highly structured Boolean and proximity queries like *pumpkin spice* within 31 words of *coffeehouse* in a file that also has (*latte or cappuccino*) and not *chai tea*. By default, searching will look for search terms anywhere in files or you can limit portions of a search request to certain metadata, like specifying that subject metadata must include the exact phrase *pumpkin spice*.

Fuzzy searching in dtSearch adjusts from 1 to 10 to sift through potential misspellings like *pumpkin* mistyped *pumpplen* in an email or mis-OCR'ed as *pumpqin*. Concept search looks for synonyms like *java* for *coffee*. Date searching picks up popular date variants, enabling *date(9/20/25 to 11/15/25)* to pick up *10/10/25* as well as *October 10 2025* and *Oct 10 2025*.

The software also supports numeric range queries and can even flag valid credit card numbers across indexed data. dtSearch has automatic relevancy-ranking as well as user-defined positive or negative variable term weighting. For a different view on search results, end-users can instantly re-sort by a new criterion like file date or file location. And whatever the sorting, search results can show retrieved files with highlighted hits for convenient review.

While pumpkin spice might not forever be in season, enterprise search is. Please visit dtSearch.com for fully-functional 30-day evaluation downloads.

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



The software also supports numeric range queries and can even flag valid credit card numbers across indexed data. dtSearch has automatic relevancy-ranking as well as user-defined positive or negative variable term weighting.