The Ghost of Christmas Past and Enterprise Data

Article contributed by dtSearch®

You might be ready to close out this year's data. However, like the ghost of Christmas past, this year's data will still hang around in your files, emails and the like, rattling its chains and waiting to cause trouble in the coming year. Enterprise search can't stop all future hauntings. But by giving you a handle on your organization's data now, enterprise search can help blunt the impact if a ghost does emerge.
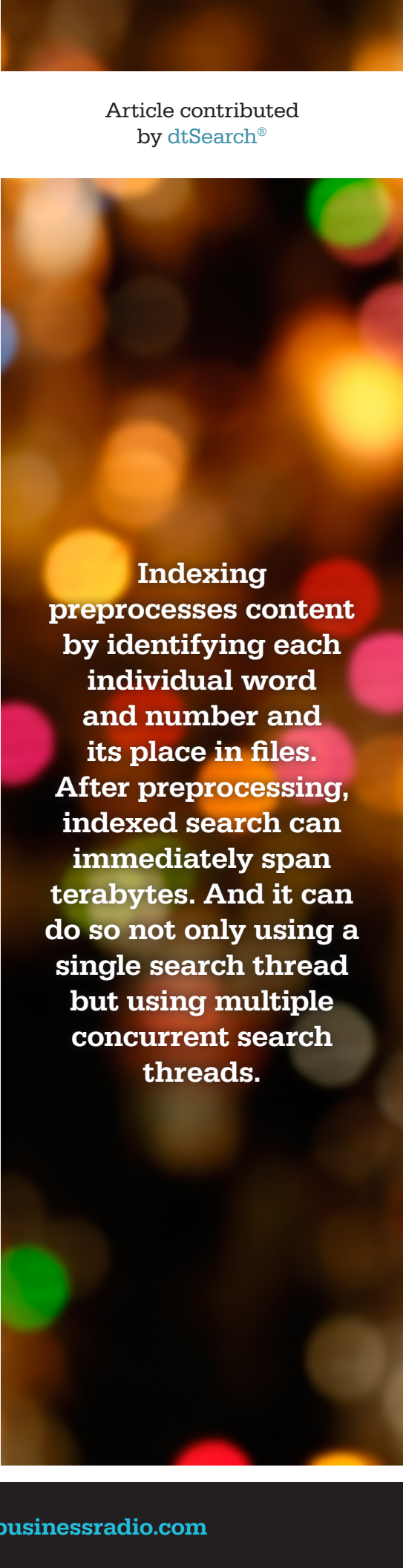
So how do you get ahead of the ghost of Christmas past in data?

Start by installing enterprise search and let it begin indexing. Let me back up a step here. Enterprise search can work in two different ways. The first is unindexed search, where you enter a search request and sit back while the software undertakes a brute force march through data for search request matches. That's fine if you are doing a one-time search for a limited number of search terms. However, if you need to repeatedly query terabytes, unindexed search is not going to cut it. The second way enterprise search can work is indexed search.

Can you explain indexing?

Indexing preprocesses content by identifying each individual word and number and its place in files. After preprocessing, indexed search can immediately span terabytes. And it can do so not only using a single search thread but using multiple concurrent search threads. Multithreaded indexed search can run from a classic Windows network, a local web server or the cloud such as Azure or AWS.

Are there any disadvantages to indexed search?

> **Indexing preprocesses content by identifying each individual word and number and its place in files. After preprocessing, indexed search can immediately span terabytes. And it can do so not only using a single search thread but using multiple concurrent search threads.**

The main disadvantage is that indexing takes time, although a 64-bit multithreaded indexing option can speed up the process substantially. But while indexing takes time, it doesn't take a lot of human exertion. With dtSearch, just point to the folders to cover and the software will take it from there. The indexer needs to identify the correct format of each file. However, the indexer can obtain that on its own from the binary format.

It doesn't even matter if a file has a mismatched extension, like a PDF with an email extension or no file extension at all. Indexing also works with multilevel nested files like an email with a ZIP or RAR attachment holding an Excel spreadsheet which itself embeds a Word document. And so long as dtSearch can see files as part of the Windows folder system, the files themselves can be local or remote.

And capacity?

A single dtSearch index can hold up to a terabyte of text data, and the software can create and enable instant simultaneous searching across any number of indexes. As end-users add new files, edit existing files and delete old files, dtSearch can update its indexes to reflect such changes without affecting continuing concurrent searching for Christmas ghosts or otherwise.

What search options does indexed searching support?

dtSearch has over 25 different search options. Enter a basic "all words" or "any words" natural language search. For more precision searching, input a structured phrase, Boolean (and/or/not) and proximity query such as *(chain or rattle)* and *(ghost pre/12 christmas past)*. This search would look for *ghost* within 12 words before the phrase *christmas past* in any file that also contains either *chain* or *rattle*.

Concept searching finds *specter* or *spook* as synonyms for *ghost*. Fuzzy searching adjusts from 1 to 10 to sift through typographical or OCR errors like *ChristNas* for *Christmas*. To further refine searching, add on a

A single dtSearch index can hold up to a terabyte of text data, and the software can create and enable instant simultaneous searching across any number of indexes.

metadata component like *Subject contains Christmas and not Spring Break*.

Are there other search features?

A date range element can find a date or date range like *date(September 30, 2023 to January 15, 2026)* spanning all file text or just specific metadata. This search will also pick up common date variants like *December 25, 2025*, *Dec 25 2025* and *12/25/25*. dtSearch can search for numbers or numeric ranges as well. Indexed searching can even flag credit card numbers across indexed data.

For multilingual text, dtSearch supports Unicode which governs the treatment of hundreds of international languages. A single email or other file can go from English to Chinese to Finnish to Japanese to Greek to Korean to Arabic to Hebrew, with Unicode and dtSearch following all of that.

And hidden text?

Hidden text can truly epitomize the ghost of Chrismas past. Indexed searching can locate not only obvious text that you would see pulling up a file in its associated application but also camouflaged text such as red lettering against a red background or green lettering against a green background. Searching also covers deleted or redacted content that still remains in a file, even if such content would not by default appear in a file's associated application. Queries can further pick up highly obscure metadata that someone browsing a file in its associated application might completely miss.

How does sorting work?

Default relevancy ranking sorts by hit density and rarity throughout indexed data. Take an "any words" search for *ghost Christmas past*. If *Christmas* and *past* are relatively common but *ghost* mentions rare, *ghost* will get a higher relevancy ranking, and files with the densest *ghost* hauntings will come out on top. dtSearch also supports custom variable term weighting, like giving *ghost* a positive weight of 9, *Christmas* a positive weight of 4 and *Easter* a negative weight of 6. Custom

For multilingual text, dtSearch supports Unicode which governs the treatment of hundreds of international languages. A single email or other file can go from English to Chinese to Finnish to Japanese to Greek to Korean to Arabic to Hebrew, with Unicode and dtSearch following all of that.
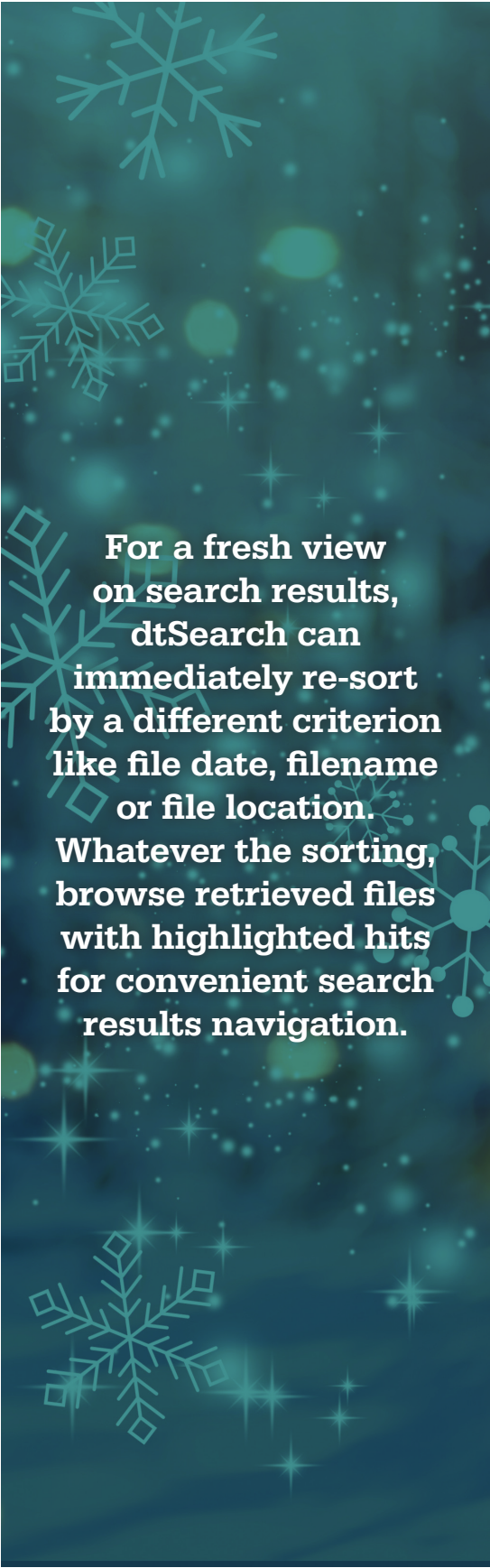
variable term weighting can further adjust for the occurrence of search terms near the top or the bottom of a file or in certain metadata.

For a fresh view on search results, dtSearch can immediately re-sort by a different criterion like file date, filename or file location. Whatever the sorting, browse retrieved files with highlighted hits for convenient search results navigation.

Final thoughts?

dtSearch.com has fully-functional 30-day evaluation downloads to start your organization now on instant multithreaded searching. Find the *ghost of Christmas past*, the *ghost of Christmas future*, or just the necessary information to get on with everyone's day.