

# Let Enterprise Search Be Your Lighthouse, Illuminating the Path Through Data Fog

National Lighthouse Day comes in early August. But don't wait. Deploy enterprise search now to start lighting a path through the fog of organizational data.

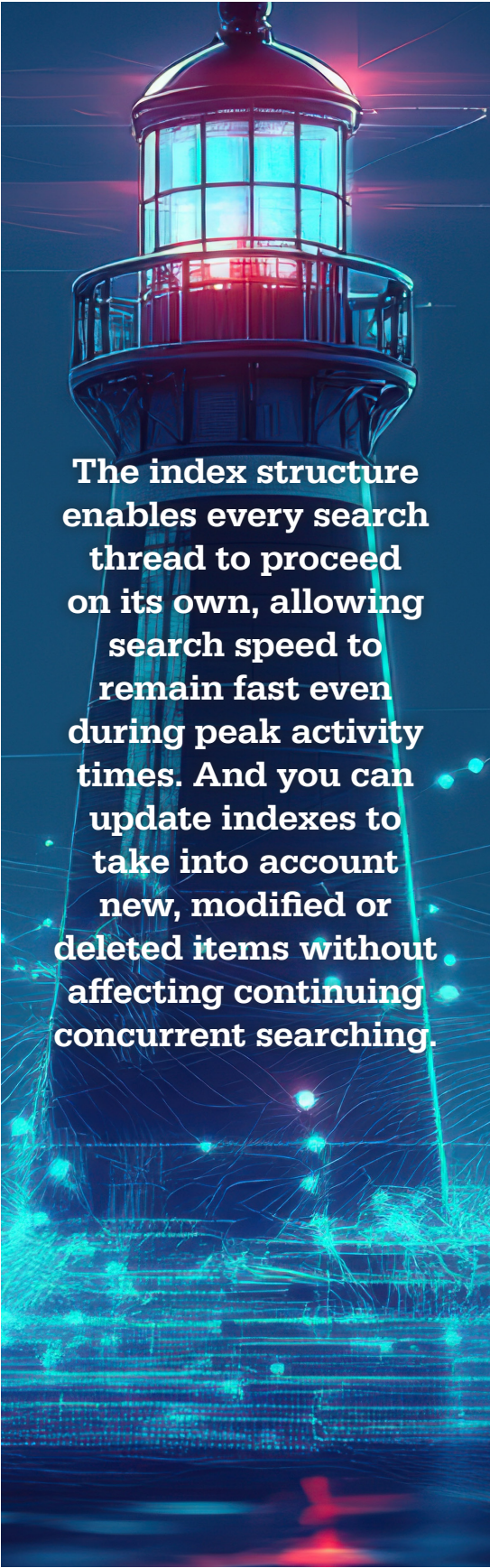
Enterprise search like dtSearch® lets end-users instantly and concurrently query terabytes after first indexing the data. Indexing may sound labor-intensive and it is for the enterprise search indexer. But all you need to do is check off the folders you want the indexer to cover. That's it. With dtSearch, each index can hold up to a terabyte of text, with no limits on the number of indexes the software can create and cover in a search.

Indexing records each unique word and number and the place of each in the data. The index structure enables every search thread to proceed on its own, allowing search speed to remain fast even during peak activity times. And you can update indexes to take into account new, modified or deleted items without affecting continuing concurrent searching.

Indexes can reside in the same general location as the indexed files so only those with access to the original files can access the indexes. For more complex security requirements, the dtSearch Engine developer SDK enables granular data classification. The classification can leverage metadata in a database like SQL or NoSQL, metadata inside the files themselves, or full-text content. If the Nuclear Option is a top-secret code word, then anything that mentions Nuclear Option anywhere can be "eyes only" for the executive suite.

The indexer can cover PDFs; Microsoft Office files like Word, Access, Excel, PowerPoint and OneNote; emails including Outlook and Exchange; and ZIP or RAR compressed data. Files can be local or remote like SharePoint attachments or OneDrive Office 365 files

Article contributed  
by dtSearch®



**The index structure enables every search thread to proceed on its own, allowing search speed to remain fast even during peak activity times. And you can update indexes to take into account new, modified or deleted items without affecting continuing concurrent searching.**

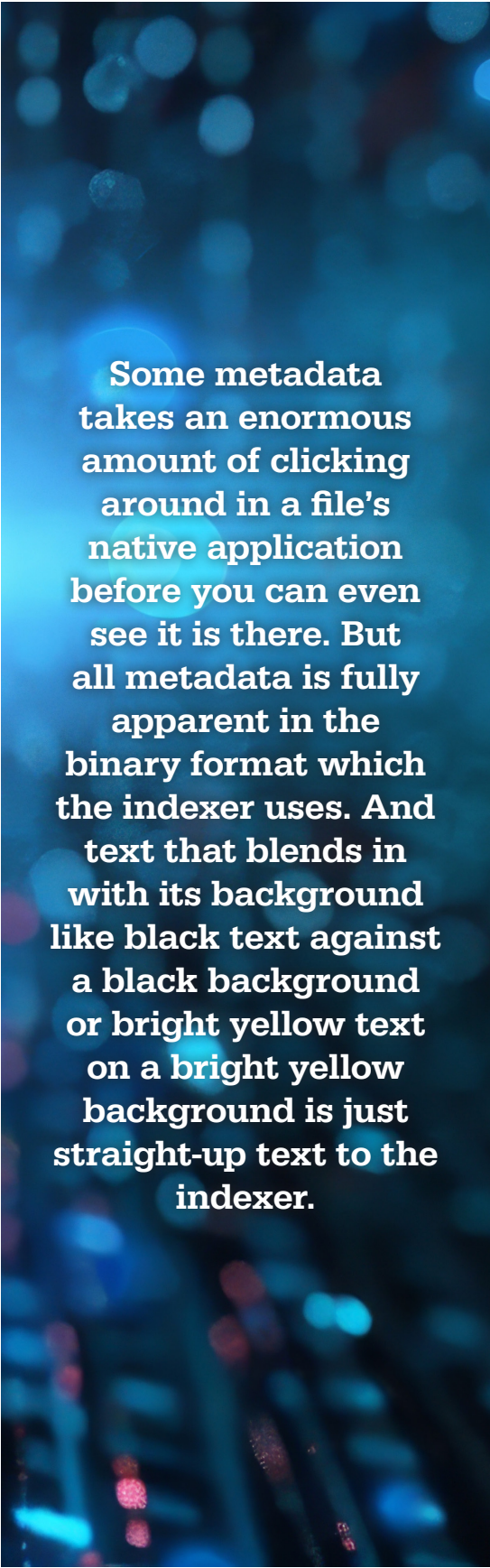
so long as these appear as part of the Windows folder system. The indexer will automatically recognize the file format of each item from its binary format, without regard to the file format extension. That way, if a Word document has a .PDF file extension or a PDF has a .DOCX file extension the indexer will still correctly handle it. The dtSearch Engine developer SDK can also work with databases like SQL or NoSQL including both referenced files and BLOB data.

Indexing goes deep into data. You can have an email with a ZIP or RAR attachment with a PDF Portfolio and a separate Word document that itself contains a nested Excel spreadsheet, and the indexer will cover all of that multilevel nested structure. Some metadata takes an enormous amount of clicking around in a file's native application before you can even see it is there. But all metadata is fully apparent in the binary format which the indexer uses. And text that blends in with its background like black text against a black background or bright yellow text on a bright yellow background is just straight-up text to the indexer.

Searching acts like a lighthouse through data fog. dtSearch, for example, supports basic queries like "any words," "all words" or exact phrase. Or it can tackle more advanced search requests using Boolean and/or/not and proximity searching: (*Operation ABC w/17 Project CDE*) and (*South Carolina or New Mexico*) and not *North Dakota*. Searches can also add on requirements for certain text in specific metadata. Concept searching extends a search request to built-in or user-defined synonyms. Fuzzy searching adjusts from 1 to 10 to sift through typographical errors as may appear in emails or OCR'ed copy. With a low-level of fuzzy searching, a query for *South Carolina* would also pick up *South Carolima*.

Beyond words, the search can also look for numbers or numeric ranges as well as dates or date ranges. A date range search can also pick up common date variants. A date range search for *date(July 31 2024 to September 15 2024)* would pick up both *8/23/24* and *Aug 24, 2024* in full-text or metadata. dtSearch can also generate and search for hash values across indexed data. The software can even flag any credit card numbers that

Article contributed  
by dtSearch®



**Some metadata takes an enormous amount of clicking around in a file's native application before you can even see it is there. But all metadata is fully apparent in the binary format which the indexer uses. And text that blends in with its background like black text against a black background or bright yellow text on a bright yellow background is just straight-up text to the indexer.**

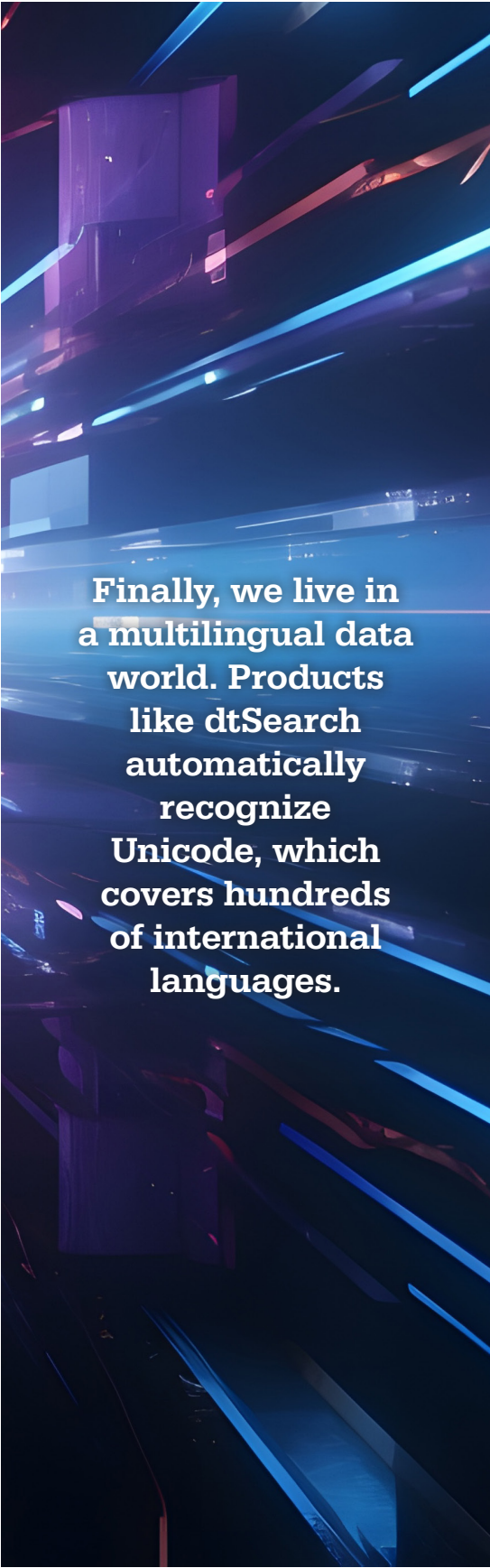
may appear in indexed data. And the product can find specific Unicode emojis. There is a navigation emoji, but no lighthouse emoji yet.

By default, the software uses vector-space relevancy ranking. Take an “any words” search for *thick*, *fog* or *lighthouse*. If *thick* and *fog* appear in thousands of files but *lighthouse* in just a few, then *lighthouse* hits will get a higher relevancy rank, with items densely referencing *lighthouse* coming out on top. Or dtSearch lets end-users override default relevancy ranking, giving *thick* a negative weight of 6 if it appears near the top or bottom of a file, and *fog* a positive weight of 8 if it shows up in certain metadata. After a search, the software can display a full copy of retrieved items with highlighted hits for convenient search results navigation.

Finally, we live in a multilingual data world. Products like dtSearch automatically recognize Unicode, which covers hundreds of international languages. These include European languages, right-to-left text like Hebrew and Arabic, as well as double-byte Chinese, Japanese and Korean text. A file or email can cycle through multiple different languages and Unicode and dtSearch will track all of that.

So don't wait for National Lighthouse Day. Light the way now through your enterprise data. You can download a fully-functional 30-day evaluation version from [dtSearch.com](http://dtSearch.com).

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



**Finally, we live in  
a multilingual data  
world. Products  
like dtSearch  
automatically  
recognize  
Unicode, which  
covers hundreds  
of international  
languages.**