

# Rule Your Data With Enterprise Search

Data can rule you, requiring you to spend vast amounts of time and energy just to find the critical item you need. Or you can rule data, deploying an enterprise search engine to instantly find whatever you need, whenever you need it.

Enterprise search can instantly search terabytes after it first indexes the data. Using dtSearch as the example, a single index can hold up to a terabyte of text and there are no limits on the number of indexes that dtSearch can create and instantly search. To index data, just point to the folders and the like to index, and the indexer will take it from there.

No need to tell the dtSearch indexer what type of content it is covering. The indexer has built-in document filters to automatically recognize popular file formats like PDF, Microsoft Word, Access, Excel, PowerPoint and OneNote, as well as compressed formats like RAR and ZIP. The document filters will also work with email formats like Outlook and Exchange. The indexer can even cover cloud-based files like Office 365 or files in SharePoint so long as these present through the Windows folder system.

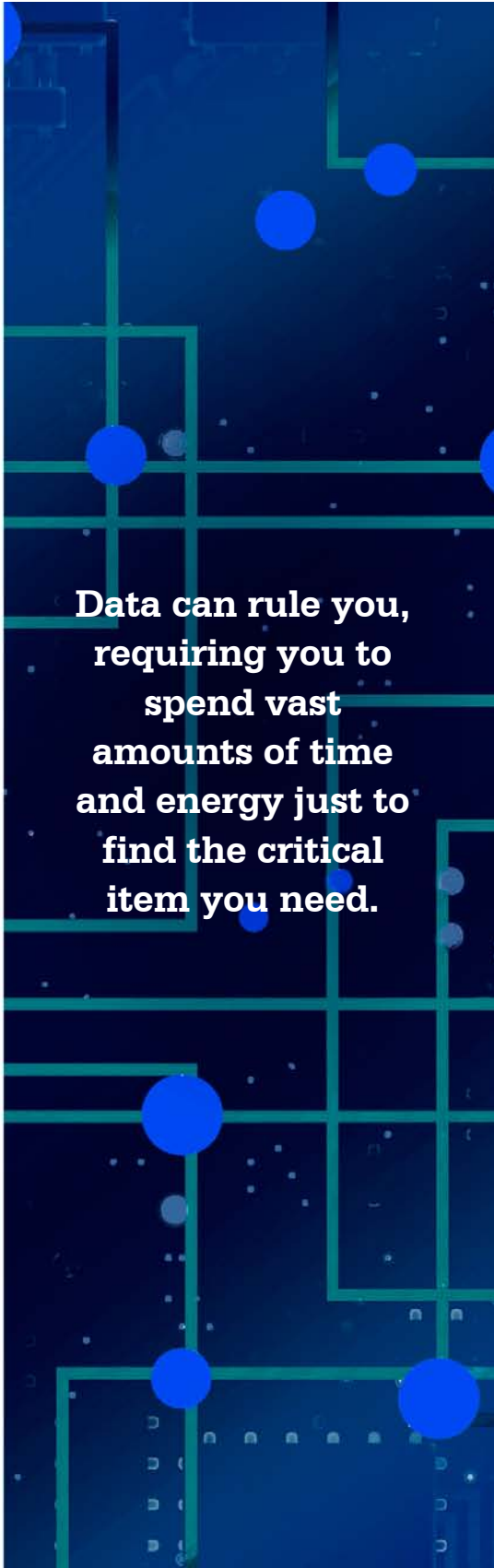
The indexer's document filters use information from inside a binary format to determine the applicable file format, so even a misplaced file format extension will not trip up the indexer. The document filters can even support recursively embedded formats. Once the indexer figures out the correct file format, it can apply the applicable parsing specification to recognize all text and metadata.

The index itself stores each unique word and number across the data along with the location of each. Importantly, while indexing is resource intensive, searching is quite light when it comes to resources. Internet or Intranet search—whether operating from an on-premises server or from the cloud like Azure or AWS—can run statelessly, with no built-in limit on the number of simultaneous and instant search threads. Automatic index updates to accommodate new, modified or deleted content can proceed without disrupting multithreaded concurrent search.

Indexed searching supports over 25 different search types. Natural language searches are the easiest. An “all words” search will look for files, emails and the like that have every one of the words in a search request. An “any words” search is broader, finding items that contain even one word in a list of search terms. Enterprise search also supports more structured search requests.

A summertime Boolean phrase search example might be *ice cream cone and chocolate chip or mocha chip and not rum raisin*. You can also add in a proximity element like *ice cream scoop w/17 sundae*, requiring that *ice cream scoop* and *sundae* appear within 17 words of each other. Or you can add in an even more

Article contributed  
by dtSearch®



**Data can rule you,  
requiring you to  
spend vast  
amounts of time  
and energy just to  
find the critical  
item you need.**

precise directed proximity requirement. *Ice cream scoop pre/9 sundae* requires that *ice cream scoop* appear within 9 words just before *sundae* for the item to be a match.

A built-in English language thesaurus can automatically extend search terms to similar concepts, covering either default thesaurus entries or user-defined synonyms. Stemming can find different endings on the same root word. Wildcard searching can find any number of missing letters in a word, such as *coffe\* ice cream*. Fuzzy searching adjustable from 1 to 10 looks for deviations in typographical or OCR data. A search for *ice cream sprinkles* with a low level of fuzziness would still pick up a mistyping of *ice cream sprinGles*.

A search request can require that one or more terms appear in specific metadata. Enterprise search can also support numeric-oriented queries for a specific number or a numeric range. Date ranging searching can pick up a date range like *July 5, 2021 through August 4, 2023* in full-text or metadata, including covering different date formats like *July 31, 2023* and *7/31/23*. The software can further identify certain numeric sequences, like finding any credit card numbers buried in the data.

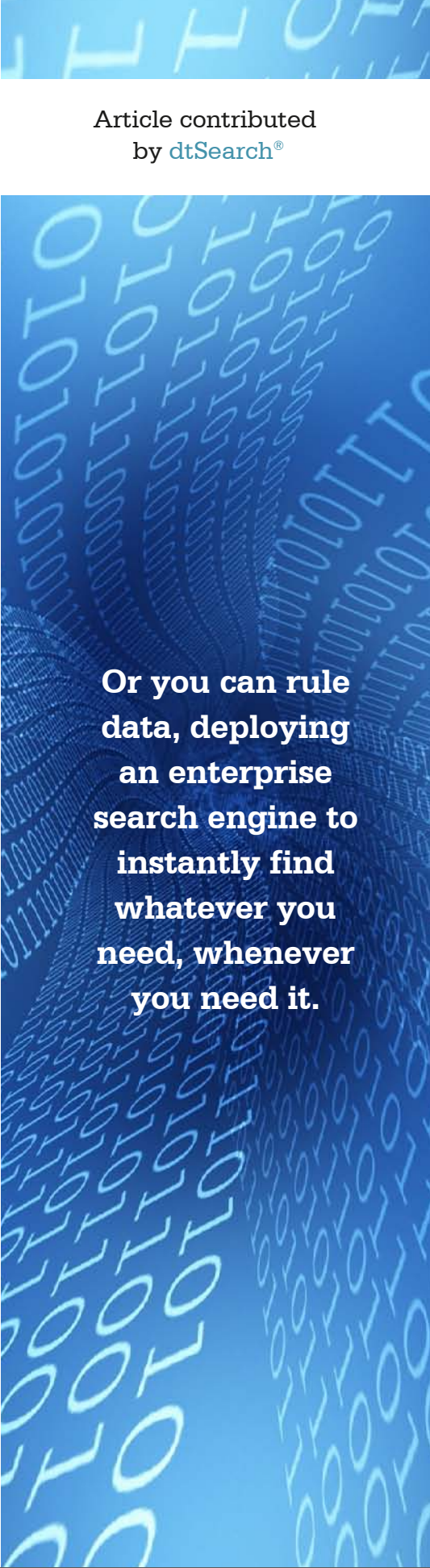
Unicode support works with hundreds of international languages. This support covers not only English and other European languages, but also right-to-left languages like Hebrew and Arabic and double-byte character languages like Chinese, Japanese and Korean. A file or email can switch languages multiple times, and Unicode and enterprise search will track the progression.

By default, relevancy-ranking will use the vector-space relevancy model. That way, if *vanilla* is all over indexed data but *peach* is relatively rare, then *peach* would get a higher relevancy ranking, and items with denser *peach* mentions would get a higher ranking still. Positive or negative variable term weighting works on top of the default ranking, letting you further customize search results. For example, you can add an additional positive or negative weight to *chocolate* regardless of its prevalence in the data, or an optionally greater weight if *chocolate* appears in certain metadata or at the top or bottom of a file.

For a different view of search results, you can instantly re-sort by some other unrelated metric like file name, file location or file date. Whatever the sorting, search results will display a complete copy of retrieved items with highlighted hits for convenient browsing. So don't let your data rule you. Rule your data!

*About dtSearch.* dtSearch has enterprise and developer products that run "on premises" or on cloud platforms to instantly search terabytes of "Office" files, PDFs, emails along with nested attachments, databases and online data. Because dtSearch can instantly search terabytes with over 25 different search features, many dtSearch customers are Fortune 100 companies and government agencies. But anyone with lots of data to search can download a fully-functional 30-day evaluation copy from [dtSearch.com](https://www.dtsearch.com)

Article contributed  
by [dtSearch®](https://www.dtsearch.com)



**Or you can rule  
data, deploying  
an enterprise  
search engine to  
instantly find  
whatever you  
need, whenever  
you need it.**