

Spring Clean Your Closet, Not Your Enterprise Data

While it may take a spring cleaning to unearth the baseball cap in your closet, enterprise data requires a different solution. Here's how to instantly find anything across terabytes of enterprise data, minus the spring clean.

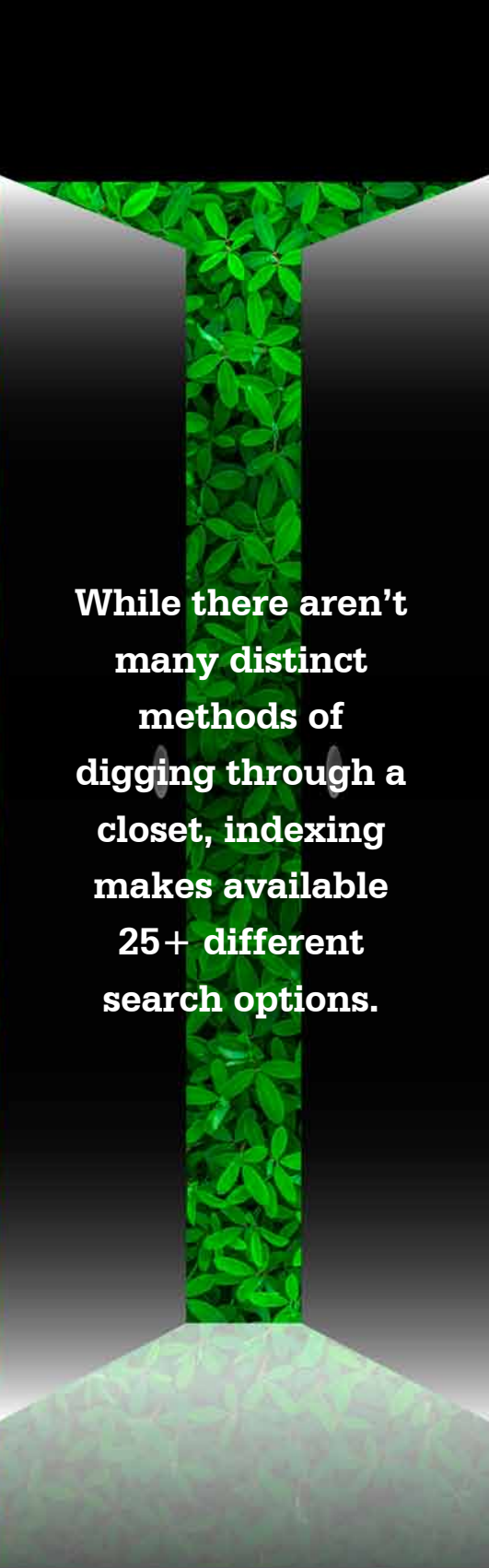
The key is a search engine. This is not an “across the Internet” search engine like Google; rather this is an enterprise search engine like dtSearch®. Such a search engine allows one individual or multiple people concurrently to instantly locate anything in the full-text or metadata across terabytes of organizational content. The data itself can span multiple different repositories and consist of mixed “Office” documents, PDFs, emails along with nested attachments, web-ready data, etc.

A search engine works by first indexing all content. When complete, an index stores each unique word and number in the data along with information on where each resides in the data. But isn't indexing a lot of work, you may ask? (Might as well clean my closet!) Indexing is a lot of work, but exclusively for the search engine. Just point to the folders and the like to index, and the search engine does everything else.

Note that while a baseball cap may all but disappear under a pile of coats, it is very hard to hide data from a search engine. In order to parse each file, a search engine has to correctly identify the applicable file format. But the search engine does this by looking inside each binary file. A mismatched file extension, such as a PDF document saved with a .DOCX file extension, has no effect on the process.

Multilevel nested data is also not a problem. For example, the indexer can parse an email with a ZIP or RAR attachment holding a PowerPoint file with an embedded Excel spreadsheet inside. Additionally, white on white or black on black text is just straight up text to the indexer. Indexing covers not only the main

Article contributed
by [dtSearch®](#)



**While there aren't
many distinct
methods of
digging through a
closet, indexing
makes available
25+ different
search options.**

text but also all scraps of metadata, even metadata that may be quite hard to spot when looking at a file in its associated application. The files (or even the same file) can include not only English, but also other European text, double-byte Chinese, Japanese and Korean text, as well as right-to-left Hebrew and Arabic text.

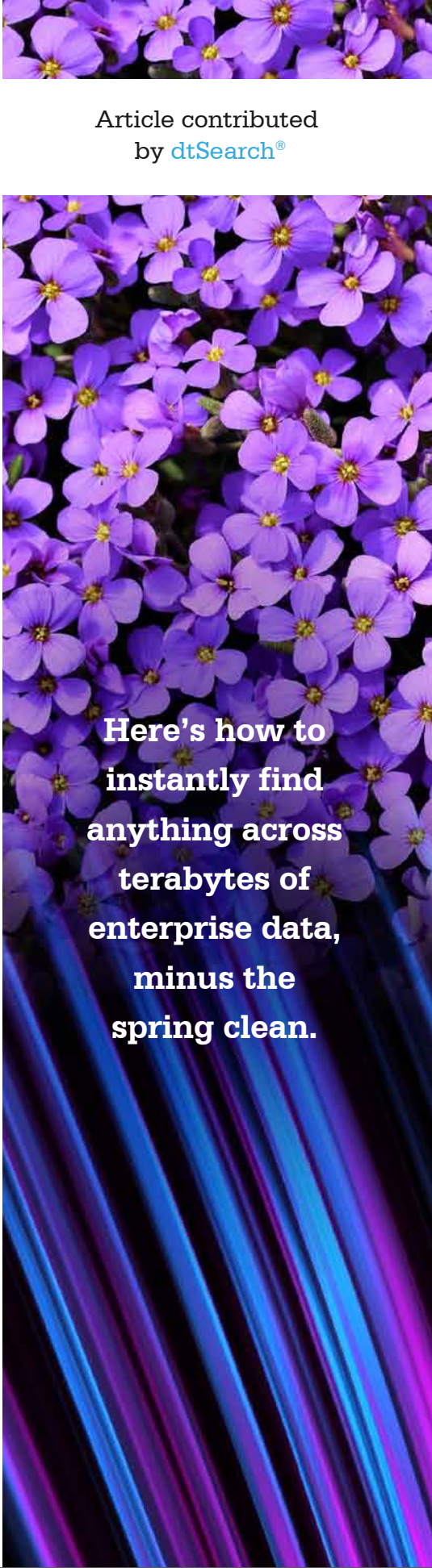
After indexing, search can run not only on an individual basis but also on a multiuser basis from a Windows network, a local web server or a remote web server such as on Azure or AWS. (Online search can run in a stateless manner, so there are no limits in the search engine itself on the number of simultaneous search threads that can instantly execute.) The search engine can automatically update indexes as often as you want to accommodate new data without affecting individual or concurrent searching.

While there aren't many distinct methods of digging through a closet, indexing makes available 25+ different search options. Search types range from natural language unstructured search requests to highly structured search requests encompassing and/or/not, proximity operators, etc. Concept searching finds synonyms. Fuzzy searching adjusts from 0 to 10 to sort through typographical and OCR misspellings.

Beyond words, a search engine can also locate numbers, numeric ranges, dates and date ranges, even sifting through mixed date formats. A search engine can further identify credit card numbers residing in data. Searching includes multiple options for relevancy ranking and can display the full text of retrieved items with highlighted hits.

So go immediately find whatever you need in your enterprise data, no spring cleaning required. What are you going to do with all the extra time?

Article contributed
by [dtSearch®](#)



**Here's how to
instantly find
anything across
terabytes of
enterprise data,
minus the
spring clean.**