

The secret sauce to finding files, emails and other enterprise data

Finding the right file, email and other internal data without enterprise search is like grilling without barbecue sauce. It is theoretically possible, but who would even want to try?

While Internet search engines like Google specialize in directing you to the right website, enterprise search products do a deep dive into an organization's own data. The secret sauce to enterprise search is indexing. Indexing "pre-processes" Microsoft Office files, PDFs, emails plus attachments, compressed archives and other web-ready data.

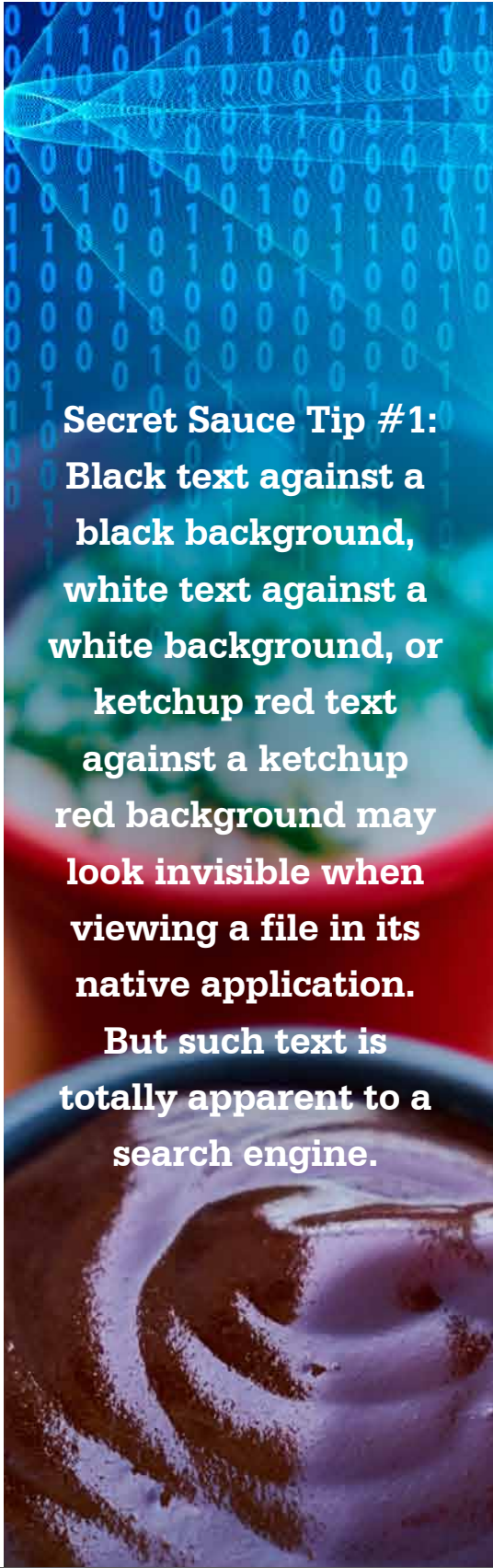
But isn't indexing a lot of work? (Might as well go into the barbecue business!) Indexing is a lot of work, but only for the search engine. All you have to do is point to the folders and the like to cover, and the indexer will do the rest. The indexer starts with figuring out the correct format of each item and applying the correct parsing specification.

After indexing, you can instantly search across terabytes for any references to *((grilling or cookout) w/12 secret sauce) and not ketchup*. Or pile on any of 25+ other full-text and metadata search options to hone search results. In a shared work environment, search can run across a network, from an "on premises" web server, or from a cloud server.

While indexing is resource intensive, online search can operate statelessly, so any number of search threads can proceed concurrently without slowing down other search threads. After a search runs, the search engine can display a full copy of retrieved data with highlighted hits. To reflect updated content, enterprise search lets you reindex automatically through the Windows Task Scheduler without blocking out individual or concurrent searching.

Now that you have the barbecue basics, some more advanced enterprise search tips follow.

Article contributed
by dtSearch®



Secret Sauce Tip #1:
Black text against a black background, white text against a white background, or ketchup red text against a ketchup red background may look invisible when viewing a file in its native application.
But such text is totally apparent to a search engine.

Secret Sauce Tip #1: Black text against a black background, white text against a white background, or ketchup red text against a ketchup red background may look invisible when viewing a file in its native application. But such text is totally apparent to a search engine.

Secret Sauce Tip #2: Some metadata can require tons of clicking around in a file's native application before you even know that it's there. But all metadata is easily accessible to a search engine.

Secret Sauce Tip #3: Mismatched file types like a PDF with a .DOCX extension will not impede the search engine, as search engines look inside a file's binary format to determine the applicable file type.

Secret Sauce Tip #4: Certain PDFs may look like regular text-based PDFs but in fact just consist of an image with no underlying text at all. Look for your search engine to flag such "image only" PDFs. That way, you can them run through an OCR program like Adobe Acrobat to turn them into full-text searchable PDFs.

Secret Sauce Tip #5: Caching will store the full text of the original files along with the index, enabling instant hit-highlighted display in the absence of a reliable network connection or if the original files are no longer present.

Secret Sauce Tip #6: Fuzzy searching can sift through typographical errors and other misspellings that may make their way into OCR'ed content or files like emails. You can add a fuzzy element (adjustable from 1 to 10 depending on the level of text errors) to any of the other search options.

Secret Sauce Tip #7: In addition to finding words, a search engine can also locate numbers and numeric ranges. The search engine can also find dates and date ranges, even across different date formats. Finally, a search engine can identify valid credit card numbers in text. That way, if the credit card that paid for the barbecue winds up in shared data, you can find it and take steps to delete it.

Article contributed
by [dtSearch®](#)



Secret Sauce Tip #2:
Some metadata can
require tons of
clicking around in
a file's native
application before
you even know that
it's there. But all
metadata is easily
accessible to a
search engine.