

Deploying Enterprise Search To Maximize Productivity: Some Next-Level Tips


You've probably heard the basics on how enterprise search can enhance productivity by enabling multiple end-users to instantly and concurrently search across terabytes. This article aims to cover some next-level enterprise search tips.

Tip #1: Consider the type of enterprise search deployment. Enterprise search can run across a classic Windows network or in a web-access capacity. While both implementations work for large-volume concurrent search access, network deployment can include features like the ability to instantly launch a retrieved file in its associated application that, for security reasons, would be problematic in web-access environments. Network installation also makes it easier for end-users to index and search their own files and emails in addition to common data repositories.

A web-based implementation is great for facilitating access from anywhere on the planet through any browser-running device with Internet/Intranet connectivity. For security, online enterprise search can sit behind any level of firewalls and log-ins. Web-based search can run from an "on premises" server or through a cloud hosting platform like Azure or AWS.

Tip #2: Use indexed search. While dtSearch®, for example, has both indexed and unindexed search options, indexed searching is vastly more efficient for concurrent searching. Indexing is easy: just point to the document folders, email archives and the like to cover and the software will do the rest. The files themselves can be local or remote—SharePoint attachments, OneDrive / Office 365 files, etc.—as long as these appear as part of the Windows folder system. A single index can hold up to a terabyte of content and the software can build as many terabyte indexes as needed, with

Article contributed
by dtSearch®



A web-based implementation is great for facilitating access from anywhere on the planet through any browser-running device with Internet/Intranet connectivity. For security, online enterprise search can sit behind any level of firewalls and log-ins.

end-users simply checking off the indexes to cover in a search.

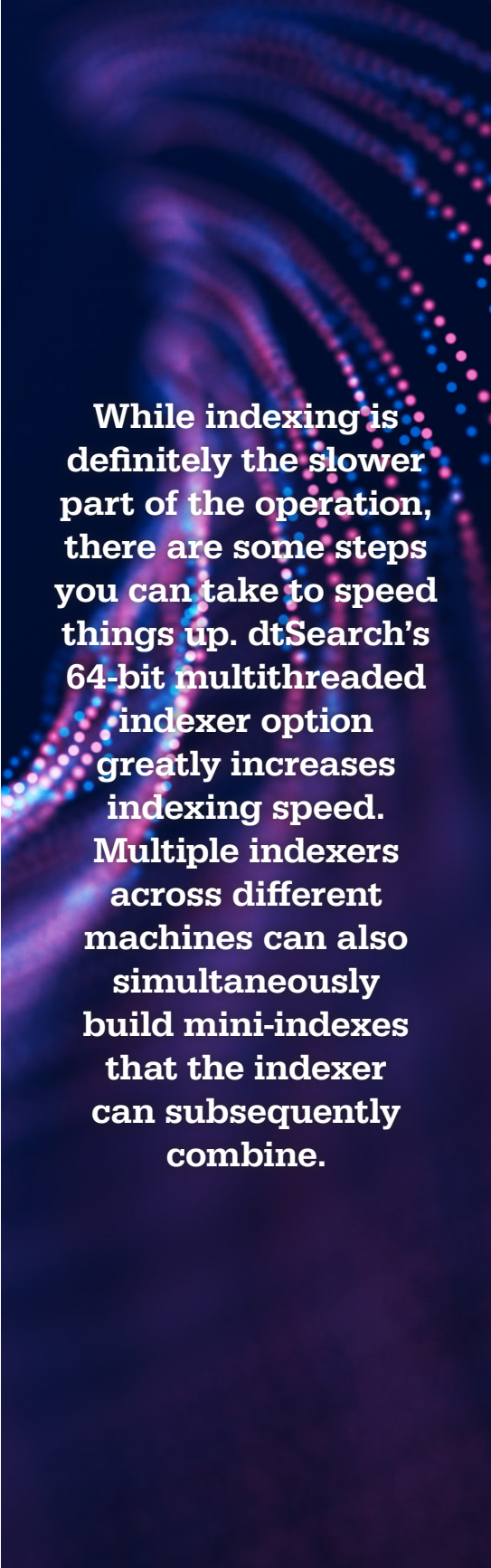
Although data parsing requires identifying the format of each individual item, the indexer can figure this out on its own for both local and remote files. The indexer looks inside the binary format of each item to ascertain whether the file type is PDF, Microsoft Word, Access, Excel, PowerPoint, OneNote, Outlook, Exchange, etc. Because the indexer uses the contents of the binary format for this determination, a Word document with a .PDF extension or a PDF with a .DOCX extension will not impede the process. And the indexer can automatically dig through multilevel nested files like an email with a ZIP or RAR attachment with a Word document with an Excel spreadsheet embedded inside.

Tip #3: While indexing is definitely the slower part of the operation, there are some steps you can take to speed things up. dtSearch's 64-bit multithreaded indexer option greatly increases indexing speed. Multiple indexers across different machines can also simultaneously build mini-indexes that the indexer can subsequently combine. For greater efficiency, merge indexes into an empty shell index rather than merging one content-containing index into another content-containing index.

Also, use a location as near as possible to the indexing engine to hold the index files as they build. Adding content to an index requires a huge amount of read/write activity. While it is fine if the folders and other target data sit on lower-speed read/write storage, the generating index should reside on a high-speed read/write environment like SSD.

Tip #4: After indexing, review the index logs. Look for externally encrypted files. For example, while dtSearch can automatically handle internally-encrypted PDFs where the password is inside the PDF, externally-encrypted files require separate decryption. If the log flags a file as encrypted, run it through a separate unencryption process and then bring it back to the indexer.

Article contributed
by dtSearch®



While indexing is definitely the slower part of the operation, there are some steps you can take to speed things up. dtSearch's 64-bit multithreaded indexer option greatly increases indexing speed. Multiple indexers across different machines can also simultaneously build mini-indexes that the indexer can subsequently combine.

Also, check the logs for “image only” PDFs. You know when you are viewing a PDF and try to copy and paste some text and nothing copies out? That is likely an “image only” PDF. In an “image only” PDF, the indexer will not have text to work with apart from the filename and metadata. “Image only” PDFs can be tough to identify in mixed data collections. But the index logs identify these so you can run them through an OCR product like Adobe Acrobat and then bring them back to the indexer.

Tip #5: Set the Windows Task Scheduler for frequent and automatic index updates to accommodate data additions, modifications and deletions. Updating an index will look for changes in the data, and then just re-index the new items. Index updates do not block out searching so there is no reason not to keep indexes current.

Tip #6: As you consider which of the over 25 different indexed search options to deploy, also bear in mind the scope of enterprise search. I noted that dtSearch indexes recursively nested file formats. Indexing will also cover all metadata—including obscure metadata that is easy to miss viewing a file in its native application. The software will also find camouflaged text like white writing against a white background. Some editing programs show text as redacted when it is really just underneath a black rectangle. dtSearch can find that text too. And if a file still contains “tracked” changes, those “tracked” changes will be available for search as well.

Ready to maximize productivity across your organization? Get started at [dtSearch.com](#) with fully-functional 30-day evaluation downloads.

Indexing will also cover all metadata—including obscure metadata that is easy to miss viewing a file in its native application. The software will also find camouflaged text like white writing against a white background.