A Search Index is a Digital Treasure Map to Your Data

Article contributed
by dtSearch®

If you are sailing the ocean looking for treasure, you could navigate randomly from one island cove to the next looking for a spot that may hold buried treasure. Or you could follow a treasure map.

Comparing these approaches to digital text search, the random search would be like an unindexed search digging through individual files, emails and the like to find X marks the spot. The treasure map would be more like an indexed search, where you would first build a search index that functions as a meta-guide and then use that meta-guide treasure map to instantly locate X marks the spot across the entire dataset.
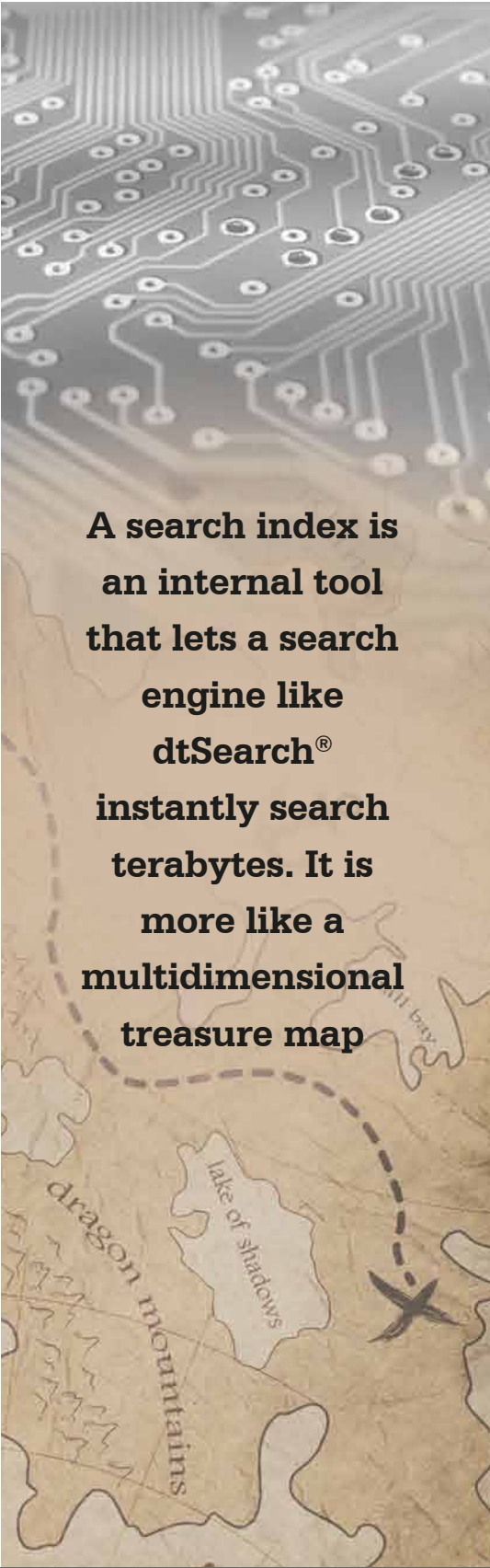
A search index is not like an index in the back of a book. Rather, a search index is an internal tool that lets a search engine like dtSearch® instantly search terabytes. It is more like a multidimensional treasure map to all indexed full-text and metadata content than a flat back-of-a-book type index. A single dtSearch index can cover up to a terabyte of data and there is no limit on the number of terabyte-size indexes that dtSearch can build and instantly search.

The same index can support not only individual searching but also enterprise-wide multithreaded concurrent searching, allowing any number of people to simultaneously search with instant hit-highlighted search results. In plain English, multiple people can use the same treasure map at the same time and each immediately arrive at their own X marks the spot.

Indexing is easy. Just point dtSearch at the relevant folders and other data and dtSearch will go off on its own and build an index. In fact, dtSearch will even figure out for itself what mix of data you have such as Word documents, Excel spreadsheets, Access databases, PowerPoints, XML, HTML or PDF files, or even compressed archives like ZIP or RAR. For emails, dtSearch will comb through not only text and metadata, but also multilayer nested email attachments.

dtSearch can update its indexes to account for files that have been added, deleted or changed without affecting current searching. That way, users can still effectively search using a search index treasure map even at the multithreaded enterprise search level at the same time an index updates.

> A search index is an internal tool that lets a search engine like dtSearch® instantly search terabytes. It is more like a multidimensional treasure map

Returning to the treasure map analogy, a search index enables instant X marks the spot basic single word and phrase searching. But in the case of dtSearch, a search index goes way beyond that to also enable instant Boolean and/or/not searching, proximity searching, wildcard searching, synonym or concept searching, fuzzy searching which looks for slight letter deviations such as can result from typing in emails or OCR'ed text, numeric range searching, numerous options for relevancy ranking and multicolor hit-highlighting.

A dtSearch index further allows for advanced search options like the ability to find any credit cards in data and developer-oriented features such as faceted search leveraging metadata as well as granular data classification for security purposes. Indexing, searching and display of documents, emails and the like does not alter original data. Each file has a unique hash value which can be generated for that document. dtSearch can generate hash values and then let you search on those, all without affecting the original data.

dtSearch enterprise and developer products instantly search terabytes of "Office" files, PDFs, emails along with attachments, databases and web-based data, including running on online platforms like Azure and AWS. Because dtSearch can instantly search terabytes, many dtSearch customers are large enterprises like Fortune 100 companies and federal, state and international government agencies.

But, in addition to enterprise-level search, dtSearch offers dtSearch Desktop to instantly search your own documents, emails and the like. You can immediately go to dtSearch.com and download a fully-functional 30-day evaluation version to find X marks the spot in terabytes of your own data.

Indexing is easy. Just point dtSearch at the relevant folders and other data and dtSearch will go off on its own and build an index