

Seasonal Time Crunch? Try a Search Engine

Holiday season is a lot of fun, but it can also be a big time commitment with family gatherings, travel and present shopping. And the same holiday season is often crunch time at work. That's where a search engine comes in, enabling instant retrieval of files and emails at the office and at home, giving you an extra time cushion.

The relevant type of search engine is a precision "search your own data" search engine like dtSearch. dtSearch can instantly sift through terabytes of web-ready data, "Office" files, PDFs, emails plus multilayer nested attachments, etc., taking you right to whatever you need. The product works by first building one or more search indexes cataloging each unique word and each unique number in the data, and the locations of each word and number in the data.

But isn't building a search index a lot of effort? Only for the search engine. For the end-user, the whole process is automatic – just point to the folders and the like you want the search engine to cover in the index and the search engine does this rest. No need to even identify the kinds of data the search engine is processing. The search engine can figure out for itself what types of "Office" documents, web-ready data, emails and the like it is working with. If you have an email with a ZIP attachment and inside is a OneNote file with an embedded PowerPoint, the search engine will figure out all that for itself.

After indexing, the search engine can instantly process over 25 different types of search requests like looking for specific words, phrases and numbers; Boolean (and/or/not) search formulations; proximity search looking for a word or phrase within X words of another word or phrase; metadata-specific search; numeric range search; date and date range searching; synonym or concept searching; and even fuzzy searching adjustable from 1 to 10 to sift through minor typographical errors that often pop up in OCR'ed text and in emails.

For example, the search engine could find *gold tinsel within 37 words of LED lights* in any file that doesn't also contain *Fourth of July or Summer Solstice*. Even if *tinsel* is misspelled *tinsol*, the search engine can still find it with fuzzy searching. After a search, the search engine can display matching files with highlighted hits, so you can instantly navigate to what you are looking for.

Relevant for holiday shopping, you can further ask dtSearch to find any credit card numbers that may be lurking in indexed data. What dtSearch does under-the-hood is take any X digits that could represent a credit card number and run those by a credit card confirmation algorithm. That way, you know everywhere a credit card number appears, giving you the option to edit the file to delete them.

Article contributed
by **dtSearch®**

**The product works
by first building
one or more search
indexes cataloging
each unique word
and each unique
number in the
data, and the
locations of each
word and number
in the data.**

Now if a search retrieves a small number of files, you can quickly browse through all of them. But sometimes a search will retrieve a huge number of matching files, and you want options to further sift through those. One option is to do a search within a search.

Take *gold tinsel within 37 words of LED lights and not Fourth of July or Summer Solstice* and add on an additional element, further limiting search results to only those files that also contain *holiday ornaments or talking snowmen*. The search engine can, at your election, limit the search to files that contain *holiday ornaments or talking snowmen* in the full-text, or look for *holiday ornaments or talking snowmen* only in specific metadata, like the subject line of email threads.

Another option is relevancy ranking. By default, a search engine like dtSearch will rank retrieved items by so-called vector space hit term density and rarity. For example, if there are tons of *ornaments* in the indexed data, but only a handful of *snowmen*, then files with *snowmen* would get a higher relevancy ranking. And files with denser *snowmen* mentions would get an even higher relevancy ranking.

Beyond the default vector space relevancy ranking, you can apply your own custom relevancy ranking. You could give *snowmen* a relevancy ranking of 6 and *tinsel* a relevancy ranking of 9 and *ornaments* a negative relevancy ranking of 7. Or you can apply custom relevancy ranking to specific metadata only. And you can always sort or instantly re-sort by other criteria like file data, file location, etc.

In a shared office environment, concurrent instant searching can run across a standard Windows network. Or it can run from an online environment, either from an “on premises” web server or in a cloud environment like Azure or AWS. In a concurrent user environment, each search thread operates independently without impacting other search requests. When a data collection changes, dtSearch can update an index or multiple indexes, adjusting only data that has been added, deleted or modified since the last index update, without affecting immediate concurrent searching.

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 precision search options, many dtSearch customers are Fortune 100 companies and government agencies. But anyone with lots of data to search is welcome to download a fully-functional 30-day evaluation version from dtSearch.com to instantly search through terabytes and save some extra time this holiday season.

Article contributed
by [dtSearch®](http://dtSearch.com)

**But isn't
building a
search index a
lot of effort?
Only for the
search engine.**