

Does Your New Year's Resolution Include Reorganizing Your Data? Cross That Off The List Right Now!

Organizing your data so you can quickly locate what you are looking for is a great idea for the New Year, just as soon as there are 60 hours in each day instead of 24. In the real world, downloading a search engine is a more realistic option to finding data.

The Internet isn't organized, but you can instantly find what you are looking for using an online search engine. The same concept also works for your own data. A search engine can instantly search terabytes of your documents, emails, online data, etc. by running on an individual PC, across a network, over an "on premises" web server, or on a remote platform like Microsoft Azure or AWS.


A search engine lets you instantly search through vast amounts of full-text data and metadata because it first builds a search index. Building that index requires zero effort on your part. Just point to the relevant directories and the like that you want to index, and the application will do the rest. No need to even tell the search engine what format your data is in; it automatically recognizes popular file formats like PDF, HTML, XML, Microsoft Word, Excel, Access, PowerPoint, OneNote, etc. The application will automatically dig through ZIP and RAR archives as well.

The search engine can also index emails along with the full text of attachments, even nested attachments. For example, if you have an email with a ZIP attachment containing a Word file with an Excel spreadsheet embedded inside the Word file, the application will handle the whole thing. And a single index doesn't need to hold just one data source; it can hold multiple data sources. That way, an index can hold ordinary documents, email files and web-based data enabling integrated search.

So, what does this search index consist of? The index is just an internal roadmap that allows the search engine to instantly search through terabytes. The index stores each unique word or number in both full-text data and metadata and the location of each word or number in that data. Indexes can reside on your own PC or laptop, letting you instantly search your own personal data. They can also sit on a shared network drive or an online-access web site, enabling instant multiuser concurrent queries.

Once the search engine finds files that match a search request, the search engine can display the full text of the files with highlighted hits for easy review. What kind of searching can your search engine support? That's where you can be creative, with over 25 different search features to select individually or "mix and match."

Article contributed
by [dtSearch®](#)



**A search engine
lets you instantly
search through
vast amounts of
full-text data and
metadata because
it first builds a
search index.
Building that
index requires
zero effort on
your part.**

Natural language search lets you enter a “plain English” search request like Sam Smith memo and find pertinent files. Importantly, if the word memo appears in millions of files but Sam and Smith only appear in a few files, relevancy ranking will position files that contain Sam and Smith at the top of the list. That way, so you won’t have to wade through a kazillion non-relevant memo files first.

Phrase search lets you search for Sam Smith only if it appears as a phrase. You can also look for Sam Smith only if it appears at the top of a file, or only if it appears in specific metadata. Boolean search lets you enter more structured search requests using and/or/not operators, like (“New York” or “New Jersey”) and (“Sam Smith” or “Bob Jones”) and not “pumpkin pie.” Proximity search lets you find Sam only if it appears within X words of Smith. And you can further specify whether Sam and Smith can appear within X words of each other in either direction, or with one in front of the other.

Concept search expands your search to synonyms of any search terms you enter. Phonic search finds words that sound alike, like Smythe in a search for Smith. Stemming finds variations on endings, like runs, running, runner when you enter run. Wildcard search lets you insert a question mark to hold just a single letter space or an asterisk to hold multiple letter spaces.

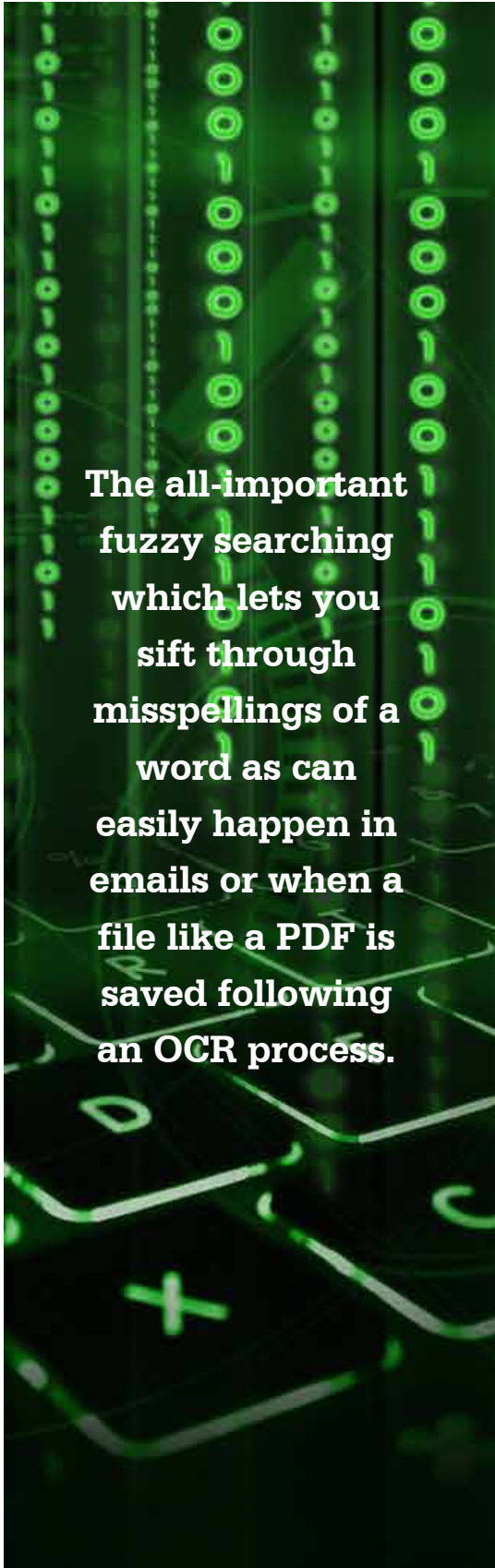
You can also search for numeric data finding specific numbers or dates, or numbers or dates in a certain range. Positive and negative term weighting lets you assign different weightings to search terms to “override” the default relevancy-ranking. You can specify a higher or lower ranking wherever the search term appears, or only if the search term appears in specific metadata or, say, near the top of a document.

Lastly, there is the all-important fuzzy searching which lets you sift through misspellings of a word as can easily happen in emails or when a file like a PDF is saved following an OCR process. Suppose that you were looking for pumpkin pie and it was misspelled pumkin pie in an email. A low level of fuzzy searching would still find that misspelling.

If you have documents in different languages, no problem. The application will automatically support any Unicode text from Russian to Chinese to Hebrew to Swedish. You can even tell the product to flag any credit card numbers. That way, in case a credit card may have accidentally made its way into a file on your network, the product can find that for you. For forensics-type searching, the application can also generate hash values for each file and then search for specific hash values.

Ready to apply a search engine to your own unruly data? Go to dtSearch.com and download a fully-functional 30-day evaluation version now. One New Year’s Resolution... done!

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



**The all-important
fuzzy searching
which lets you
sift through
misspellings of a
word as can
easily happen in
emails or when a
file like a PDF is
saved following
an OCR process.**