

Fuzzy and Other Methods of Broadening a Search

For those who aren't familiar with dtSearch, what does dtSearch do?

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 different search features, many dtSearch customers are Fortune 100 companies and government agencies. But anyone with lots of data can download a fully-functional 30-day evaluation copy from dtSearch.com

Can you describe how dtSearch works?

dtSearch instantly searches terabytes by first building a search index containing each unique number and word and its location in the data. Indexing is easy. No need to even tell dtSearch what file formats and the like dtSearch is working with. The software will figure that out for itself. Once dtSearch finishes indexing, it can instantly search through terabytes, displaying retrieved items with highlighted hits.

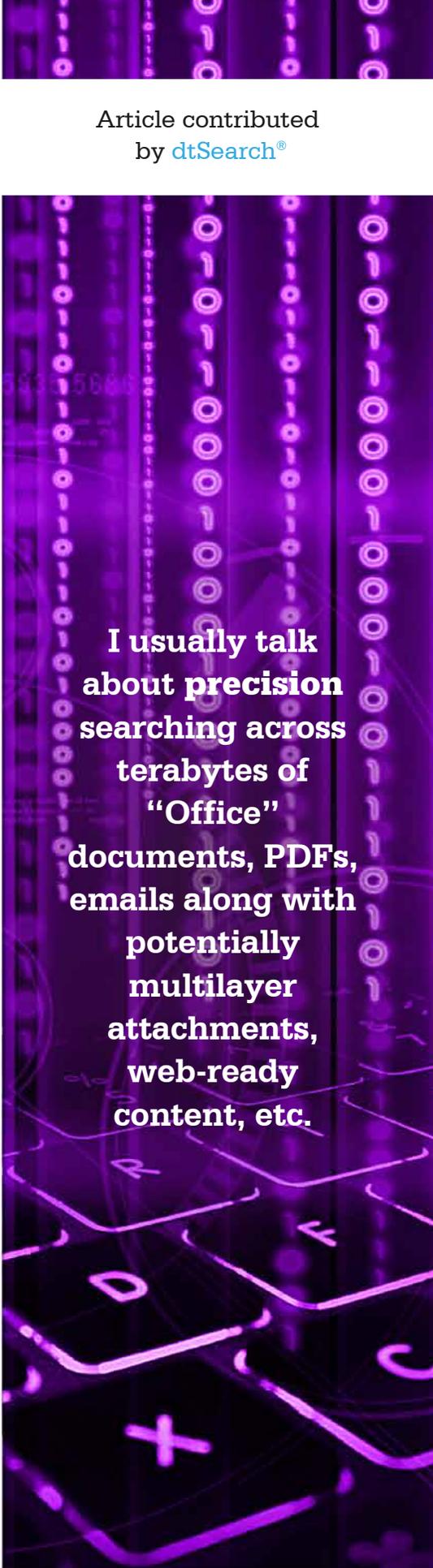
Can multiple people search at once?

dtSearch supports concurrent searching on a network or online. Whether running "on premises" or in a cloud environment like Azure or AWS, each search proceeds independently for multithreaded searching. And updating an index to add new content does not affect ongoing concurrent searching.

What is today's topic?

I usually talk about **precision** searching across terabytes of "Office" documents, PDFs, emails along with potentially multilayer attachments, web-ready content, etc. Today I wanted to focus on methods of **broadening** a search request. Metaphorically, this would look at a search request as less of a straight arrow and more like a "fishing expedition."

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



**I usually talk
about **precision**
searching across
terabytes of
"Office"
documents, PDFs,
emails along with
potentially
multilayer
attachments,
web-ready
content, etc.**

What do you mean by that?

Let's start with precision searching. You can combine Boolean (and/or/not) and proximity operators to make your search as precise as you want. For example, let's say I was searching government documents for extraterrestrial life sightings. I might search for: *area 51* within 75 words before *little green men* with no mention of *weather balloons* or *satellites*. That precision search would hone in on documents that only contain the phrase *little green men* appearing within 75 words after *area 51*, and exclude any documents also containing the phrase *weather balloons* or *satellites*. Now suppose I'm searching not official government documents, but emails. Misspellings in emails are rampant. So I want to make sure that if GREEN is mistyped GREAN, the search would still find the text.

How would you do that?

By applying the search broadening technique that dtSearch calls fuzzy searching. Fuzzy searching adjusts from 1 to 10 to look at varying degrees of typographical deviations. That makes it a good option for searching not only emails where misspellings are common but also OCR'ed text. When you have text that is OCR'ed, especially if the original is a little blurry, misspellings like GREEM for GREEN can occur that you may want fuzzy searching to sift through.

What are some other options for broadening a search?

Another option would be concept searching. A related word to *green* might be *teal*. If I activate concept searching, every time I look for *green*, I could also find *teal*.

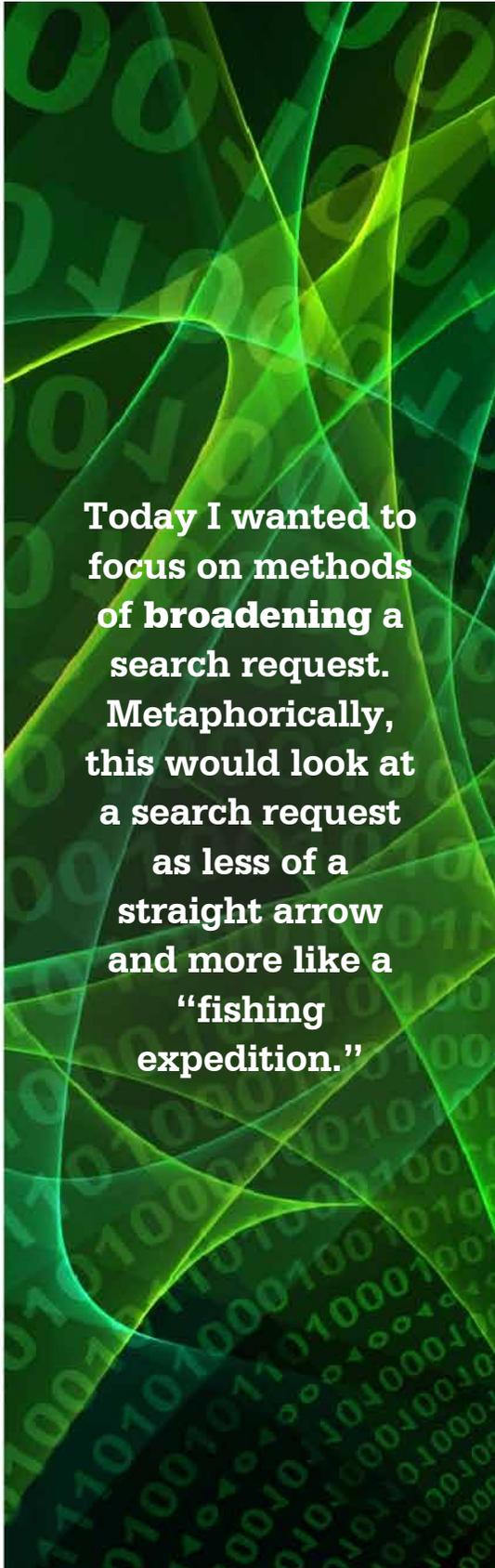
How does dtSearch come up with synonyms?

There is a built-in thesaurus in dtSearch to find synonyms or related words. And you can also enter your own custom synonym rings. That way, even if *drone* and *satellite* might not be standard English language synonyms, I can still make them synonyms for purpose of my search if I want to.

What are some other options for broadening a search?

My original search referenced *Area 51*. But suppose I also wanted to extend that to *Areas 49 through 65*. I could do that with numeric range searching. I didn't search for a specific date in the initial search request. But if I had a date for a possible little green man sighting, I could add that to my search request, and then maybe broaden that to cover all dates in a surrounding date range.

Article contributed
by [dtSearch®](#)



Today I wanted to focus on methods of broadening a search request. Metaphorically, this would look at a search request as less of a straight arrow and more like a "fishing expedition."

Any other ways to broaden a search?

Stemming looks for different versions of the same root word. With stemming on, I could extend *balloon* to *ballooning*. My original search also looked for an *area 51* mention within 75 words before *little green men*. I could extend that to look for either phrase within say 350 words of the other phrase before or after.

Are there circumstances in which you'd like to narrow a search request?

Narrowing a search request is a convenient option when a search request returns a vast number of items. One easy way to narrow a search is to do a "search within a search." With that, you take your previous search request and add on some additional items to winnow that down.

Like what?

I could take my *little green men* search and add on an additional requirement, such as only looking for results that also include *Top Secret* in specific metadata. Without narrowing my search results, I could also prioritize what I looked at by using relevancy ranking. Default relevancy ranking uses a vector-space algorithm, which looks at the number of times search terms appear in a data set and ranks them by density and rarity. The less common a search term in indexed data, the higher the ranking of that search term. Denser references to a less common search term would get an even higher relevancy ranking. And you can also add your own custom relevancy ranking, giving select words a custom positive or negative relevancy ranking. Or I could just instantly re-sort search results by some completely unrelated criterion, like ascending or descending file date, to provide a different window into search results.

Anything else you'd like to add?

Anyone is welcome to download a fully-functional 30-day evaluation version from dtSearch.com to come up with your own search requests across terabytes of data.

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

Narrowing a search request is a convenient option when a search request returns a vast number of items.