

Re: relevance sorting with multiple search terms?

## Re: relevance sorting with multiple search terms?

---

*Source:*

<http://www.tech-archive.net/Archive/SQL-Server/microsoft.public.sqlserver.fulltext/2006-04/msg00110.html>

---

- *From:* "Hilary Cotter" <[hilary.cotter@xxxxxxxx](mailto:hilary.cotter@xxxxxxxx)>
  - *Date:* Mon, 24 Apr 2006 10:53:51 -0400
- 

I was talking a building a thesaurus. For people might search on chocolate bars, chocco bars, bars, candy bars, and they mean the same thing. So someone searching on chocolate bars, would be best served by expanding their search to all synonyms of the phrase chocolate bars.

--

Hilary Cotter  
Director of Text Mining and Database Strategy  
RelevantNOISE.Com – Dedicated to mining blogs for business intelligence.

This posting is my own and doesn't necessarily represent RelevantNoise's positions, strategies or opinions.

Looking for a SQL Server replication book?  
<http://www.nwsu.com/0974973602.html>

Looking for a FAQ on Indexing Services/SQL FTS  
<http://www.indexserverfaq.com>

"msft-sql" <[noone@xxxxxxxx](mailto:noone@xxxxxxxx)> wrote in message  
[news:er\\$62iXZGHA.1192@xxxxxxxxxxxxxxxxxxxxxxxxxxxx](mailto:news:er$62iXZGHA.1192@xxxxxxxxxxxxxxxxxxxxxxxxxxxx)

--

"Hilary Cotter" <[hilary.cotter@xxxxxxxx](mailto:hilary.cotter@xxxxxxxx)> wrote in message  
[news:%2306K4BCZGHA.4916@xxxxxxxxxxxxxxxxxxxxxxxxxxxx](mailto:news:%2306K4BCZGHA.4916@xxxxxxxxxxxxxxxxxxxxxxxxxxxx)

If you know in advance all the possibilities of the search terms, I would use the expansion type of the thesaurus option.

Hi Hilary: I'm not sure I understand what you mean by "all the possibilities"...This is a user-controlled search tool...they're just

Re: relevance sorting with multiple search terms?

typing in stuff they're looking for.

I can compile a sort-of "actionable words" list for words that should be "demoted" in relevance, like:

bulk  
candy  
bar

...but I don't know if that's what you mean?

Or do you mean creating my own custom dictionary of search "keywords" with weightings?

Or are you talking about something other than weighting altogether?

--

Hilary Cotter  
Director of Text Mining and Database Strategy  
RelevantNOISE.Com – Dedicated to mining blogs for business intelligence.

This posting is my own and doesn't necessarily represent RelevantNoise's positions, strategies or opinions.

Looking for a SQL Server replication book?  
<http://www.nwsu.com/0974973602.html>

Looking for a FAQ on Indexing Services/SQL FTS  
<http://www.indexserverfaq.com>

"msft-sql" <aklist@xxxxxxxxxxxxxx> wrote in message  
[news:%23VUics\\$YGHA.1196@xxxxxxxxxxxxxxxxxxxxxxxxxxxx](news:%23VUics$YGHA.1196@xxxxxxxxxxxxxxxxxxxxxxxxxxxx)

Hi: I'm trying to get a handle on the best way to approach this issue.

I have a product database of candy with perhaps 5000 products, and I'm indexing the product name and description fields (both varchars).

People will search for "easter candy" for example.

Splitting the string and using a contains query on both terms will usually produce too broad a result, returning every product with "candy"

Re: relevance sorting with multiple search terms?

in the name. A freetext query is also too broad.

Searching with an "AND" is too limiting, since I want to return "chocolate easter egg" even if "candy" is nowhere in the name or description.

A proximity search can be too limiting as well, because the words could be completely separate in the description, e.g. "These chocolate easter eggs are the perfect type of candy for ..."

I've tried a weighted search, like:  
select \* from products

where contains (name, 'isabout(easter Weight(1.0), candy Weight(0.0))')

...but that produces the same results even if I reverse the weighting for the two terms. Maybe I'm not writing the query correctly?

I'm wondering if other people have dealt with this before? I could add words like "candy", "bar", "bulk", "package", etc. to the noise list, but I don't want to exclude them all the time; e.g., if someone searches for "bulk chocolate" I don't want to drop the word "bulk" and then return every instance of "chocolate". Similarly I don't want to return hits on "bulk lemon drops" when I'm searching for "bulk chocolate"

It seems like weighting is the way to go, and I can even maintain an array of low-weight words and dynamically assign the weight to them when building the query, but it doesn't seem to work properly.

Any suggestions would be appreciated!

Re: relevance sorting with multiple search terms?