

Re: search phrase

Source:

<http://www.tech-archive.net/Archive/SQL-Server/microsoft.public.sqlserver.fulltext/2005-09/msg00034.html>

- *From:* "John Kane" <jt-kane@xxxxxxxxxxxx>
 - *Date:* Sun, 11 Sep 2005 16:06:01 -0700
-

Clarification...

I have directly confirmed with the SQL Dev team that this is functional using only FREETEXT or FREETEXTTABLE under SP4 (Service Pack 4), the most currently available service pack.

-- John

--

SQL Full Text Search Blog

<http://spaces.msn.com/members/jtkane/>

"John Kane" <JohnKane@xxxxxxxxxxxxxxxxxxxxxxxxxxxx> wrote in message news:D3BBB365-AAEE-429D-89BB-52873293F3E6@xxxxxxxxxxxxxxxxxxxx

- > Yair,
- > What you might want to try is using the SQL Server 2000 XML thesoursa
- > files
- > to include the company variations you are looking for. Note, that while
- > this
- > option is un-supported by Microsoft, I have directly confirmed with the
- > SQL
- > Dev team that this is functional using only FREETEXT or FREETEXTTABLE
- > under
- > SP3 as well.
- >
- > See KB article 309678 "FIX: Thesaurus Support is Not Available for
- > Full-Text
- > Search" at
- > <http://support.microsoft.com/default.aspx?scid=kb:en-us:309678&Product=sql2k>
- > for details. Below are the steps necessary to enable this functionality.
- >
- > use pubs
- > go
- >
- > CREATE TABLE test_07062001 (
- > PK int NOT NULL IDENTITY (1, 1),

Re: search phrase

```
> col char(50),
> PRIMARY KEY (PK)
> )
>
> insert test_07062001 (col) values('KW')
> insert test_07062001 (col) values('Karen Delaney')
> insert test_07062001 (col) values('Kalen Delaney')
> insert test_07062001 (col) values('Karen Waterson')
> insert test_07062001 (col) values('Karen Watterson')
> insert test_07062001 (col) values('KWatterson')
> insert test_07062001 (col) values('KarenW')
> insert test_07062001 (col) values('Sunset Boulevard')
> insert test_07062001 (col) values('Sunset Blvd')
> insert test_07062001 (col) values('Blvd of Broken Dreams')
> insert test_07062001 (col) values('Boulevard of Broken Dreams')
> insert test_07062001 (col) values('Stephen Dybing')
> insert test_07062001 (col) values('Steven Dybing')
>
> then build your FTS Index
> sp_fulltext_database 'enable'
> sp_fulltext_catalog 'Karen', 'create'
> sp_fulltext_table test_07062001, 'create', 'Karen', 'PK'
> sp_fulltext_column test_07062001, 'col', 'add', 1033
> sp_fulltext_table test_07062001, 'activate'
>
> After the Full Population is complete, run the following SQL FREETEXT
> queries
>
> select * from test_07062001 where freetext(col,'karen') -- 3 hits
> select * from test_07062001 where freetext(col,'steven') -- 1 hit
> select * from test_07062001 where freetext(col,'boulevard') -- 2 hits
>
> Navigate to c:\Program Files\Microsoft SQL
> Server\MSSQL\FTDATA\SQLServer\Config find a file named tseng.xml (note
> that
> this is for UK English and you need US English)
>
HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\ContentIndexCommon\LanguageResources\Override\SQLServer
> (United States) and changing TsaurusFile to read tsenu.xml – be sure to
> include the full path if you do).
>
> Modify your tseng.xml file and save it.
>
> <XML ID="Tahoe Thesaurus">
> <thesaurus xmlns="x-schema:tsSchema.xml">
> <replacement>
> <pat>steven</pat>
> <sub weight="1.0">Stephen</sub>
> </replacement>
> <replacement>
> <pat>blvd</pat>
```

Re: search phrase

Re: search phrase

```
> <sub weight="1.0">Boulevard</sub>
> </replacement>
> <replacement>
> <pat>KW</pat>
> <pat>KarenW</pat>
> <pat>KWatterson</pat>
> <pat>Karen Watterson</pat>
> <sub weight="1.0">Karen Watterson</sub>
> </replacement>
> <replacement>
> <pat>colour</pat>
> <sub weight="1.0">color</sub>
> </replacement>
> <expansion>
> <sub weight="0.5">run**</sub>
> <sub weight="0.5">jog**</sub>
> </expansion>
> </thesaurus>
> </XML>
> */
>
> Now, re-run the following SQL FREETEXT queries & compare the results from
> above:
>
> select * from test_07062001 where freetext(col,'karen')
> select * from test_07062001 where freetext(col,'steven')
> select * from test_07062001 where freetext(col,'boulevard')
>
> Hope that helps!
> John
> SQL Full Text Search Blog
> http://spaces.msn.com/members/jtkane/
>
>
> "Yair Nissan" wrote:
>> Hi,
>>
>> We are using SQL 2000 sp3. We have a Customers table in which the
>> customer's
>> name is FTS indexed.. Since the insertion into the DB is made manually
>> the
>> customers' names are often misspelled or written with small modifications
>> from the already existing names in our DB. The table is using
>> SQL_Latin1_General_CP1_CI_AS collation.
>>
>> We've removed all the content of the noise files since our customers are
>> from all over the world and we need to remove a different set of noise
>> words
>> for each country.
>>
```

Re: search phrase

Re: search phrase

>> I've been testing what the best search phrase for the customer name might
>> be, and currently we're using the following algorithm :
>>
>> 1. Remove all non alphanumeric characters from the customer name.
>> 2. Remove the country oriented "noise" words – such as Inc, LTD for US
>> customers , Gmbh of German customers etc .
>> 3. Split the customer name using the white space as the delimiter.
>> 4. Start querying the DB :
>> a. A FreeTextTable query with the full customer name (after stages
>> 1–3). If one of the results (after performing stages 1–3 on each of them)
>> is
>> the exact match it means we found the customer in the DB and we take the
>> appropriate actions.
>> b. If not, we remove the last word and go on until less than
>> ((number
>> of words in the original customer name) / 2) words are left. If we get to
>> that point it means that there is no record for the customer and we add a
>> new one.
>>
>> This query gives us the best results (performance is a different
>> issue...)
>> so far, however very often this algorithm doesn't work, especially when
>> the
>> customer name was entered with an extra white space ("Sun Microsystems"
>> and
>> "Sun Micro Systems") or using the plural version of the word ("Sun
>> Microsystem" AND "Sun Microsystems").
>>
>> I've been looking all over the net for good articles about that subject
>> but
>> found very few. I would appreciate any help you can provide.
>>
>> Thanks,
>> Yair
>

• **Follow-Ups:**

- ◆ **Re: search phrase**
◇ From: Yair Nissan

• **References:**

- ◆ **search phrase**
◇ From: Yair Nissan
- ◆ **RE: search phrase**
◇ From: John Kane

Re: search phrase

- Prev by Date: [**RE: search phrase**](#)
- Next by Date: [**Backing Up**](#)
- Previous by thread: [**RE: search phrase**](#)
- Next by thread: [**Re: search phrase**](#)
- Index(es):
 - ◆ [**Date**](#)
 - ◆ [**Thread**](#)