# Analyze Results of Word Search

# sys.dm\_fts\_parser (Transact-SQL)

Returns the final tokenization result after applying a given [word breaker](http://msdn.microsoft.com/en-us/library/ms142509.aspx), [thesaurus](http://msdn.microsoft.com/en-us/library/ms142491.aspx), and [stoplist](http://msdn.microsoft.com/en-us/library/ms142551.aspx) combination to a query string input. The tokenization result is equivalent to the output of the Full-Text Engine for the specified query string.

**sys.dm\_fts\_parser** is a dynamic management function.

http://i.msdn.microsoft.com/Global/Images/clear.gif Syntax

sys.dm\_fts\_parser('query\_string', lcid, stoplist\_id, accent\_sensitivity)

http://i.msdn.microsoft.com/Global/Images/clear.gif Arguments

*query\_string*

The query that you want to parse. *query\_string* can be a string chain that [CONTAINS](http://msdn.microsoft.com/en-us/library/ms187787.aspx) syntax support. For example, you can include inflectional forms, a thesaurus, and logical operators.

*lcid*

Locale identifier (LCID) of the word breaker to be used for parsing *query\_string*.

*stoplist\_id*

ID of the stoplist, if any, to be used by the word breaker identified by *lcid*. *stoplist\_id* is **int**. If you specify 'NULL', no stoplist is used. If you specify 0, the system STOPLIST is used.

A stoplist ID is unique within a database. To obtain the stoplist ID for a full-text index on a given table use the [sys.fulltext\_indexes](http://msdn.microsoft.com/en-us/library/ms186903.aspx) catalog view.

*accent\_sensitivity*

Boolean value that controls whether full-text search is sensitive or insensitive to diacritics. *accent\_sensitivity* is **bit**, with one of the following values:

|  |  |
| --- | --- |
| **Value** | **Accent sensitivity is…** |
| 0 | Insensitive  Words such as "café" and "cafe" are treated identically. |
| 1 | Sensitive  Words such as "café" and "cafe" are treated differently. |

|  |
| --- |
| **Cc280463.note(en-us,SQL.100).gifNote:** |
| To view the current setting of this value for a full-text catalog, run the following Transact-SQL statement: SELECT fulltextcatalogproperty('*catalog\_name*', 'AccentSensitivity');. |

http://i.msdn.microsoft.com/Global/Images/clear.gif Table Returned

|  |  |  |
| --- | --- | --- |
| **Column name** | **Data type** | **Description** |
| **keyword** | **varbinary(128)** | The hexadecimal representation of a given keyword returned by a word breaker. This representation is used to store the keyword in the full-text index. This value is not human-readable, but it is useful for relating a given keyword to output returned by other dynamic management views that return the content of a full-text index, such as [sys.dm\_fts\_index\_keywords](http://msdn.microsoft.com/en-us/library/cc280900.aspx) and [sys.dm\_fts\_index\_keywords\_by\_document](http://msdn.microsoft.com/en-us/library/cc280607.aspx).  Cc280463.note(en-us,SQL.100).gifNote:  OxFF represents the special character that indicates the end of a file or dataset. |
| **group\_id** | **int** | Contain an integer value that is useful for differentiating the logical group from which a given term was generated. For example, 'Server AND DB OR FORMSOF(THESAURUS, DB)"' produces the following **group\_id** values in English:  group\_id display\_term  1 Server  2 DB  3 DB |
| **phrase\_id** | **int** | Contains an integer value that is useful for differentiating the cases in which alternative forms of compound words, such as full-text, are issued by the word breaker. Sometimes, with presence of compound words ('multi-million'), alternative forms are issued by the word breaker. These alternative forms (phrases) need to be differentiated sometimes.  For example, 'multi-million' produces the following **phrase\_id** values in English:  phrase\_id display\_term  1 multi  1 million  2 multimillion |
| **occurrence** | **int** | Indicates the order of each term in the parsing result. For example, for the phrase "SQL Server query processor" **occurrence** would contain the following **occurrence** values for the terms in the phrase, in English:  occurrence display\_term  1 SQL  2 Server  3 query  4 processor |
| **special\_term** | **nvarchar(8000)** | Contains information about the characteristics of the term that is being issued by the word breaker, one of:  Exact match  Noise word  End of Sentence  End of paragraph  End of Chapter |
| **display\_term** | **nvarchar(8000)** | Contains the human-readable form of the keyword. As with the functions designed to access the content of the full-text index, this displayed term might not be identical to the original term due to the denormalization limitation. However, it should be precise enough to help you identify it from the original input. |
| **expansion\_type** | **int** | Contains information about the nature of the expansion of a given term, one of:  0 =Single word case  2=Inflectional expansion  4=Thesaurus expansion/replacement  For example, consider a case in which the thesaurus defines run as an expansion of jog:  <expansion>  <sub>run</sub>  <sub>jog</sub>  </expansion>  The term FORMSOF (FREETEXT, run) generates the following output:  run with **expansion\_type**=0  runs with **expansion\_type**=2  running with **expansion\_type**=2  ran with **expansion\_type**=2  jog with **expansion\_type**=4 |
| **source\_term** | **nvarchar(8000)** | The term or phrase from which a given term was generated or parsed. For example, a query on the '"word breakers" AND stemmers' produces the following **source\_term** values in English:  source\_term display\_term  word breakers word  word breakers breakers  stemmers stemmers |

http://i.msdn.microsoft.com/Global/Images/clear.gif Remarks

**sys.dm\_fts\_parser** supports the syntax and features of full-text predicates, such as [CONTAINS](http://msdn.microsoft.com/en-us/library/ms187787.aspx) and [FREETEXT](http://msdn.microsoft.com/en-us/library/ms176078.aspx), and functions, such as [CONTAINSTABLE](http://msdn.microsoft.com/en-us/library/ms189760.aspx) and [FREETEXTTABLE](http://msdn.microsoft.com/en-us/library/ms177652.aspx).

**Using Unicode for Parsing Special Characters**

When you parse a query string, **sys.dm\_fts\_parser** uses the [collation](http://msdn.microsoft.com/en-us/library/ms187582.aspx) of the database to which you are connected, unless you specify the query string as [Unicode](http://msdn.microsoft.com/en-us/library/ms187828.aspx). Therefore, for a non-Unicode string that contains special characters, such as ü or ç, the output might be unexpected, depending on the collation of the database. To process a query string independently of the database collation, prefix the string with N, that is, **N'***query\_string***'**.

For more information, see "C. Displaying the Output of a String that Contains Special Characters," later in this topic.

**When to Use sys.dm\_fts\_parser**

**sys.dm\_fts\_parser** can be very powerful for debugging purposes. Some major usage scenarios include:

* To understand how a given word breaker treats a given input  
  When a query returns unexpected results, a likely cause is the way that the word breaker is parsing and breaking the data. By using **sys.dm\_fts\_parser**, you discover the result that a word breaker passes to the full-text index. In addition, you can see which terms are stopwords, which are not searched in the full-text index. Whether a term is a stopword for a given language depends on whether it is in the stoplist specified by the *stoplist\_id* value that is declared in the function.   
  Note as well the accent sensitivity flag, which will allow the user to see how the word breaker will parse the input having in mind its accent sensitivity information.
* To understand how the stemmer works on a given input  
  You can find out how the word breaker and the stemmer parse a query term and its stemming forms, by specifying a CONTAINS or CONTAINSTABLE query containing the following FORMSOF clause:

[[http://i.msdn.microsoft.com/Global/Images/clear.gif](javascript:CopyCode('ctl00_mainContentContainer_ctl51other');)Copy Code](javascript:CopyCode('ctl00_mainContentContainer_ctl51other');)

FORMSOF( INFLECTIONAL, query\_term )

The results tell you what terms are being passed to the full-text index.

* To understand how the thesaurus expands or replaces all or part of the input  
  You can also specify:

[[http://i.msdn.microsoft.com/Global/Images/clear.gif](javascript:CopyCode('ctl00_mainContentContainer_ctl52other');)Copy Code](javascript:CopyCode('ctl00_mainContentContainer_ctl52other');)

FORMSOF( THESAURUS, query\_term )

The results of this query show how the word breaker and thesaurus interact for the query term. you can see the expansion or replacements from the thesaurus and identify the resulting query that is actually being issued against the full-text index.   
  
Note that if the user issues:

[[http://i.msdn.microsoft.com/Global/Images/clear.gif](javascript:CopyCode('ctl00_mainContentContainer_ctl53other');)Copy Code](javascript:CopyCode('ctl00_mainContentContainer_ctl53other');)

FORMSOF( FREETEXT, query\_term )

The inflectional and Thesaurus capabilities will take place automatically.

In addition to the preceding usage scenarios, **sys.dm\_fts\_parser** can help significantly to understand and troubleshoot many other issues with full-text query.

http://i.msdn.microsoft.com/Global/Images/clear.gif Permissions

Requires membership in the **sysadmin** fixed server role and access rights to the specified stoplist.

http://i.msdn.microsoft.com/Global/Images/clear.gif Examples

**A. Displaying the output of a given word breaker for a keyword or phrase**

The following example returns the output from using the English word breaker, whose LCID is 1033, and no stoplist on the following query string:

The Microsoft business analysis

Accent sensitivity is disabled.

[[http://i.msdn.microsoft.com/Global/Images/clear.gif](javascript:CopyCode('ctl00_mainContentContainer_ctl61other');)Copy Code](javascript:CopyCode('ctl00_mainContentContainer_ctl61other');)

SELECT \* FROM sys.dm\_fts\_parser (' "The Microsoft business analysis" ', 1033, 0, 0)

**B. Displaying the output of a given word breaker in the context of stoplist filtering**

The following example returns the output from using the English word breaker, whose LCID is 1033, and an English stoplist, whose ID is 77, on the following query string:

"The Microsoft business analysis" OR "MS revenue"

Accent sensitivity is disabled.

[[http://i.msdn.microsoft.com/Global/Images/clear.gif](javascript:CopyCode('ctl00_mainContentContainer_ctl62other');)Copy Code](javascript:CopyCode('ctl00_mainContentContainer_ctl62other');)

SELECT \* FROM sys.dm\_fts\_parser (' "The Microsoft business analysis" OR " MS revenue" ', 1033, 77, 0)

**C. Displaying the Output of a String that Contains Special Characters**

The following example uses Unicode to parse the following French string:

**français**

The example specifies the LCID for the French language, **1036**, and the ID of a user-defined stoplist, **5**. Accent sensitivity is enabled.

[[http://i.msdn.microsoft.com/Global/Images/clear.gif](javascript:CopyCode('ctl00_mainContentContainer_ctl63other');)Copy Code](javascript:CopyCode('ctl00_mainContentContainer_ctl63other');)

SELECT \* FROM sys.dm\_fts\_parser(N'français', 1036, 5, 1);