SQL Server 2005 Books Online (September 2007)

Loading Licensed Third Party Wordbreakers

New: **5 December 2005**

Microsoft SQL Server 2005 includes licensed third-party wordbreakers for the following languages:

- Danish
- Polish
- Portuguese-Brazilian
- Portuguese-Iberian
- Russian
- Turkish

These wordbreakers are available but are not installed by default, and must be manually registered.

> **Note:**
>
> We recommend having the Microsoft Full-Text Engine for SQL Server (MSFTESQL) service set to run under a low-privileged account.

Registering the Wordbreakers

To register a wordbreaker, you must do the following:

1. Add the COM ClassID(s) for the wordbreaker and stemmer interfaces for the language being registered as a key to the <InstanceRoot>\MSSearch\CLSID node of the Microsoft Windows Registry.
2. Add a key to the <InstanceRoot>\MSSearch\Language node for the language.
3. Add configuration values that specify the location of the lexicon, noise word, and thesaurus files for the language.

> ⚠ **Caution:**
>
> Incorrectly editing the registry can severely damage your system. Before making changes to the registry, you should back up any valued data on the computer.

You must also have the following information:

- Instance IDs for each instance of SQL Server on which you want to register the wordbreakers.
- The FTDATA path for each instance.
- Retrieving Instance IDs for Multiple Instances of SQL Server

The registry paths listed in the following instructions are for the first instance of SQL Server 2005, which has instance ID MSSQL.1. If you have multiple instances of SQL Server, you must modify the registry paths by substituting the instance ID for that instance instead of MSSQL.1. To obtain the instance ID for an instance:

1. Click **Start**, and click **Run**.
2. In the Run dialog box, in the **Open** box, type Regedit.
3. Click **OK**. This opens the Registry Editor.
4. In the Registry Editor, select the following registry key for the first instance of SQL Server:
   ```
   HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Microsoft SQL Server\Instance Names\SQL
   ```

The right pane displays the instance names and their corresponding instance IDs.

Retrieving Instance-Specific FTDATA Path

After obtaining the instance IDs, you must retrieve the appropriate instance-specific path to the FTData folder. You will use this path when adding configuration values that specify the lexicon, noise word, and thesaurus files for a language. To obtain the instance-specific FTData folder path:

1. Click **Start**, and click **Run**.
2. In the **Run** dialog box, in the **Open** box, type Regedit.
3. Click **OK**.
4. In the Registry Editor, select the following registry key for an instance of SQL Server:
   `HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Microsoft SQL Server\<Instance ID>\MSSQLServer` where <Instance ID> is MSSQL.1 for the first instance of SQL Server. The registry key value will therefore be:
   `HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Microsoft SQL Server\MSSQL.1\MSSQLServer`

The right pane displays the **FullTextDefaultPath** value, which contains the instance specific path to the FTData folder. For example, for the first instance this might be C:\Program.

Installing the Wordbreakers

The installation procedure for third-party wordbreakers licensed by Microsoft consists of three stages.

> **Note:**
>
> The Danish wordbreaker is used as an example in the steps below. Values to install other language wordbreakers are provided in the tables later in this topic.

Stage 1: Add the COM ClassID(s) for the wordbreaker and stemmer interfaces for the language being registered.

To add COM Class ID(s) for these components:

1. Open the Registry Editor, by:
   a. Clicking **Start**, and clicking **Run.**
   b. In the **Run** dialog box, in the **Open** box, type Regedit.
2. In Registry Editor, select the following registry key for the first instance of SQL Server:
   HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Microsoft SQL Server\MSSQL.1\MSSearch\CLSID
3. On the menu bar, click **Edit**, click **New**, and click **Key**.
4. Type **{16BC5CE4-2C78-4CB9-80D5-386A68CC2B2D}**.
5. Press ENTER.
6. In the right pane, right-click the **Default** registry value, and then click **Modify**.
7. In the **Edit String** dialog box, in the **Value data** box, type **danlr.dll**, and then click **OK**.
8. Repeat steps 3 through 7, replacing the value in step 4 with **{83BC7EF7-D27B-4950-A743-0F8E5CA928F8}**.

For other languages, follow the steps above, replacing the key values in steps 4 and 8 with the key values for the language you want. These values are listed below. In step 7, replace danlr.dll with the .dll name for the language you want.

Language	Key value for step 4	.DLL name for step 7	Key value for step 8
Polish	**{B8713269-2D9D-4BF5-BF40-2615D75723D8}**	**lrpolish.dll**	**{CA665B09-4642-4C84-A9B7-9B8F3CD7C3F6}**
Portuguese-Brazilian	**{25B7FD48-5404-4BEB-9D80-B6982AF404FD}**	**ptblr.dll**	**{D5FCDD7E-DBFF-473F-BCCD-3AFD1890EA85}**
Portuguese-Iberian	**{5D5F3A69-620C-4952-B067-4D0126BB6086}**	**ptslr.dll**	**{D4171BC4-90BE-4F70-8610-DAB1C17F063C}**
Russian	**{20036404-F1AF-11D2-A57F-006052076F32}**	**ruslr.dll**	**{20036414-F1AF-11D2-A57F-006052076F32}**

Turkish	{23A9C1C3-3C7A-4D2C-B894-4F286459DAD6}	trklr.dll	{8DF412D1-62C7-4667-BBEC-38756576C21B}

Stage 2. Add a key to the <InstanceRoot>\MSSearch\Language node for the language.

To add a key to this node for the Danish language:

1. Select the following registry key for the first instance of SQL Server:

   `HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Microsoft SQL Server\MSSQL.1\MSSearch\Language`

2. Repeat steps 3 through 5 in the preceding procedure, replacing the key name in step 4 with **dan**.

For other languages, follow the preceding steps, replacing the key name in step 4 with the value listed below for the specific language.

Language	Key name for step 4
Polish	**plk**
Portuguese-Brazilian	**ptb**
Portuguese-Iberian	**pts**
Russian	**rus**
Turkish	**trk**

Stage 3. Add configuration values that give the location of the lexicon, noise word, and thesaurus files for the language.

To add configuration values for these components for the Danish language:

1. Select the registry key you entered in Stage 2 above. For the first instance of SQL Server this would be:

   `HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Microsoft SQL Server\MSSQL.1\MSSearch\Language\dan`

2. On the menu bar, click **Edit**, click **New**, and click **String Value**.
3. Type **NoiseFile.**
4. Press ENTER.
5. Right-click the NoiseFile registry value you just added, and then click **Modify**.
6. In the **Edit String** dialog box, in the **Value data** box, type **<Instance_specific_FTData_path>\noisedan.txt**, where <Instance_specific_FTData_path> is the path retrieved in the section described earlier ("Retrieving the instance specific FTData folder path").
7. Click **OK**.

Repeat steps 2 through 7 to add the values listed below, replacing the value type (step 2), value name (steps 3 and 5), and value data (step 6) for each value.

Note:
These values are for Danish.

Value type for step 2	Value names for steps 3 and 5	Value type for step 6
String value	TsaurusFile	<Instance_specific_FTData_path>\tsdan.xml
DWORD value	Locale	00000406
String value	WBreakerClass	{16BC5CE4-2C78-4CB9-80D5-386A68CC2B2D}
string value	StemmerClass	{83BC7EF7-D27B-4950-A743-0F8E5CA928F8}

For the Polish language, follow the steps outlined above, using the values listed below. Select the registry key you entered for Polish in Stage 2 above. For the first instance of the SQL Server this would be:
`HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Microsoft SQL Server\MSSQL.1\MSSearch\Language\plk`

Value type for step 2	Value names for steps 3 and 5	Value data for step 6
String value	NoiseFile	<Instance_specific_FTData_path>\noiseplk.txt
String value	TsaurusFile	<Instance_specific_FTData_path>\tsplk.xml
DWORD value	Locale	00000415
String value	WBreakerClass	{CA665B09-4642-4C84-A9B7-9B8F3CD7C3F6}
String value	StemmerClass	{B8713269-2D9D-4BF5-BF40-2615D75723D8}

For the Portuguese-Brazilian language, follow the steps outlined above, using the values listed below. Select the registry key you entered for Portuguese-Brazilian in Stage 2 above. For the first instance of the SQL Server this would be: `HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Microsoft SQL Server\MSSQL.1\MSSearch\Language\ptb`

Value type for step 2	Value names for steps 3 and 5	Value data for step 6
String value	NoiseFile	<Instance_specific_FTData_path>\noiseptb.txt
String value	TsaurusFile	<Instance_specific_FTData_path>\tsptb.xml
DWORD value	Locale	00000416
String value	WBreakerClass	{25B7FD48-5404-4BEB-9D80-B6982AF404FD}
String value	StemmerClass	{D5FCDD7E-DBFF-473F-BCCD-3AFD1890EA85}

For the Portuguese-Iberian language, follow the steps outlined above, using the values listed below. Select the registry key you entered for Portuguese-Iberian in Stage 2 above. For the first instance of SQL Server, this would be: `HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Microsoft SQL Server\MSSQL.1\MSSearch\Language\pts`

Value type for step 2	Value names for steps 3 and 5	Value data for step 6
String value	NoiseFile	<Instance_specific_FTData_path>\noisepts.txt
String value	TsaurusFile	<Instance_specific_FTData_path>\tspts.xml
DWORD value	Locale	00000816
String value	WBreakerClass	{5D5F3A69-620C-4952-B067-4D0126BB6086}
String value	StemmerClass	{D4171BC4-90BE-4F70-8610-DAB1C17F063C}

For the Russian language, follow the steps outlined above, using the values listed below. Select the registry key you entered for Russian in Stage 2 above. For the first instance of SQL Server, this would be:
`HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Microsoft SQL Server\MSSQL.1\MSSearch\Language\rus`

Value type for step 2	Value names for steps 3 and 5	Value data for step 6
String value	NoiseFile	<Instance_specific_FTData_path>\noiserus.txt
String value	TsaurusFile	<Instance_specific_FTData_path>\tsrus.xml

DWORD value	Locale	00000419
String value	WBreakerClass	{20036404-F1AF-11D2-A57F-006052076F32}
String value	StemmerClass	{20036414-F1AF-11D2-A57F-006052076F32}

For the Turkish language, follow the steps outlined above, using the values listed below. Select the registry key you entered for Turkish in Stage 2 above. For the first instance of SQL Server this would be:

`HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Microsoft SQL Server\MSSQL.1\MSSearch\Language\trk`

Value type for step 2	Value names for steps 3 and 5	Value data for step 6
String value	NoiseFile	<Instance_specific_FTData_path>\noisetrk.txt
String value	TsaurusFile	<Instance_specific_FTData_path>\tstrk.xml
DWORD value	Locale	0000041f
String value	WBreakerClass	{8DF412D1-62C7-4667-BBEC-38756576C21B}
String value	StemmerClass	{23A9C1C3-3C7A-4D2C-B894-4F286459DAD6}

See Also
Concepts

Word Breakers and Stemmers [http://msdn2.microsoft.com/en-us/library/ms142509(printer).aspx]
Loading Licensed Third Party Wordbreakers [http://msdn2.microsoft.com/en-us/library/ms345188(printer).aspx]
Full-Text Search [http://msdn2.microsoft.com/en-us/library/ms142571(printer).aspx]

Help and Information

Getting SQL Server 2005 Assistance [http://msdn2.microsoft.com/en-us/library/ms166016(printer).aspx]

Community Content