

با استفاده از امکانات ابتدایی T-SQL مانند like می‌توان جستجوهای را برای یافتن موارد مشابه با عبارتی خاص انجام داد، اما این جستجوها بسیار هزینه‌بر و کند هستند. در SQL Server برای مدیریت جستجوهای سریع و پیشرفته بر روی متون، افزونه‌های توکاری مانند Full text search, Semantic search, Term lookup و Term extraction دیده شده‌اند. Semantic search از نگارش 2012 آن افزوده شده‌است و مابقی در نگارش‌های پیشین آن نیز وجود داشته‌اند.

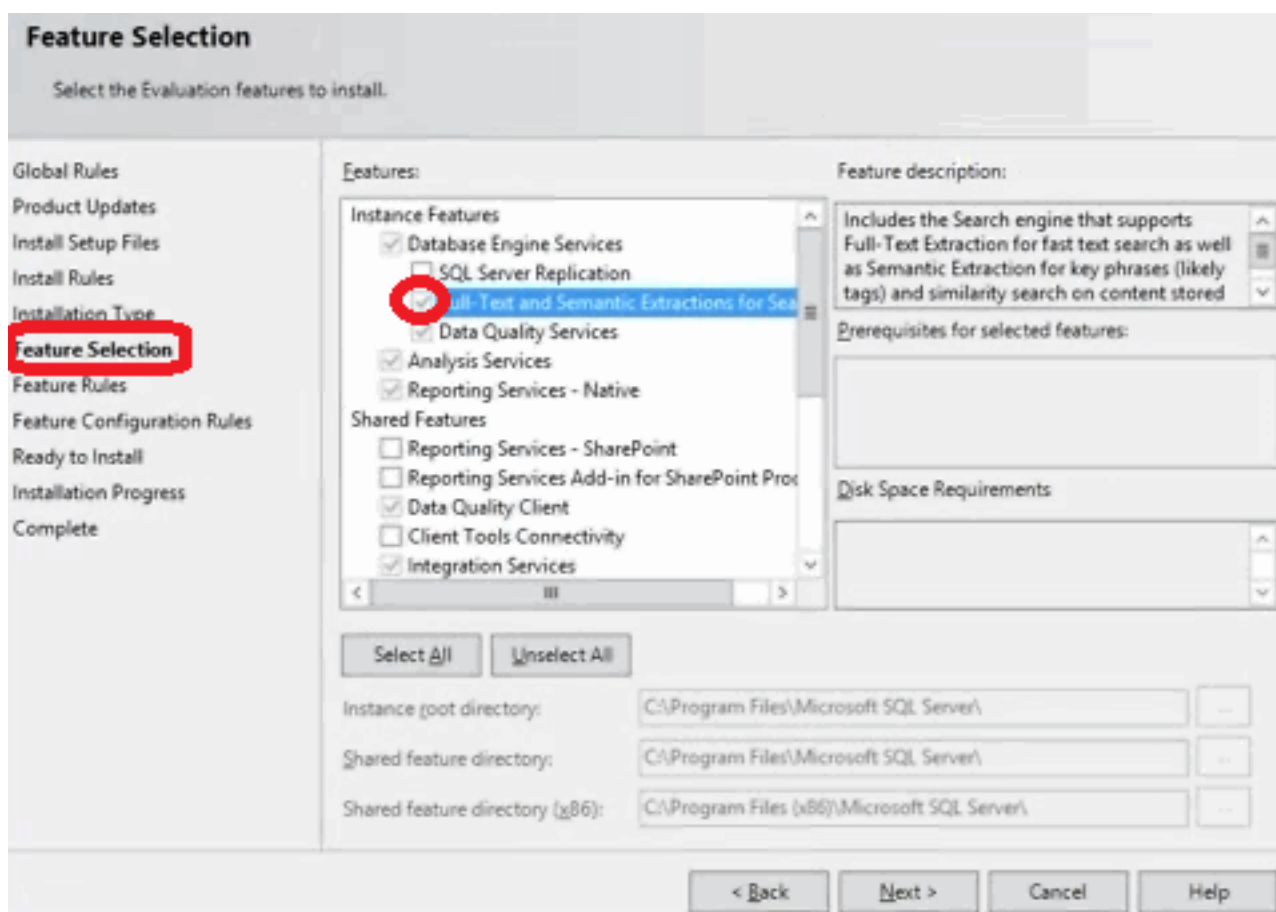
بررسی‌های مقدماتی

ابتدای کار نیاز است بررسی کنیم آیا افزونه‌ی Full Text Search، به همراه SQL Server نصب شده‌است یا خیر. برای این منظور کوئری ذیل را اجرا کنید:

```
select SERVERPROPERTY('IsFullTextInstalled');
```

اگر خروجی این کوئری عدد 1 بود، یعنی FTS نصب شده‌است؛ اگر خیر، مجدداً برنامه‌ی نصاب SQL Server را اجرا کرده و زمانیکه به قسمت feature selection رسیدید، گزینه‌ی ذیل را باید انتخاب کنید:

instance features -> database engine services -> Full Text



راه اندازی سرویس Full Text Search

پیش از ادامه‌ی بحث، به کنسول سرویس‌های ویندوز مراجعه کرده و مطمئن شوید که سرویس SQL Full-text Filter Daemon Launcher MSSQLSERVER نیز در حال اجرا است. در غیراینصورت با خطای ذیل مواجه خواهید شد:

SQL Server encountered error 0x80070422 while communicating with full-text filter daemon host (FDHost) process.

اگر این سرویس در حال اجرا است و باز هم خطای فوق ظاهر شد، مجدداً به کنسول سرویس‌های ویندوز مراجعه کرد، در برگه‌ی خواص سرویس SQL Full-text Filter Daemon Launcher MSSQLSERVER، گزینه‌ی logon را یافته و آن را به local system account تغییر دهید. سپس سرویس را ری استارت کنید. پس از آن نیاز است دستور ذیل را نیز اجرا کنید:

```
sp_fulltext_service 'restart_all_fdhosts'
```

Services (Local)			
SQL Full-text Filter Daemon Launcher (MSSQLSERVER)		Name	Description Status
Stop the service Restart the service Description: Service to launch full-text filter daemon process which will perform document filtering and word breaking for SQL Server full-text search. Disabling this service will make full-text search features of SQL Server unavailable.		SQL Full-text Filter Daemon Launcher (MSSQLSERVER)	Service to la... Running
		SQL Server (MSSQLSERVER)	Provides sto... Running
		SQL Server Agent (MSSQLSERVER)	Executes jo... Running
		SQL Server Analysis Services (MSSQLSERVER)	Supplies onl... Running
		SQL Server Browser	Provides SQ... Running
		SQL Server Distributed Replay Client	One or mor... Running
		SQL Server Distributed Replay Controller	Provides tra... Running
		SQL Server Integration Services 11.0	Provides m... Running
		SQL Server Reporting Services (MSSQLSERVER)	Manages, e... Running
		SQL Server VSS Writer	Provides th... Running
		SSDP Discovery	Discovers n... Running

چه نوع داده‌هایی را می‌توان توسط FTS ایندکس کرد؟

با استفاده از امکانات FTS می‌توان کلیه ستون‌هایی را که دارای نوع‌های ذیل باشند، ایندکس کرد:

```
char, nchar, varchar, nvarchar, text, ntext, image, xml, varbinary(max)
```

البته نوع باینری را که ملاحظه می‌کنید مانند image و varbinary max، نیاز به یک ستون اضافی، برای ذخیره سازی پسوند فایل‌های ذخیره شده در آن‌ها مانند docx، pdf، xlsx و امثال آن نیز دارند. برای مثال ابتدا یک فایل word را در ستونی از نوع varbinary max ذخیره می‌کنید و سپس نیاز است در همانجا در ستونی دیگر، پسوند این فایل را نیز قید نمایید. همچنین FTS برای پردازش این فایل‌های باینری و ایندکس کردن اطلاعات آن‌ها، نیاز به افزونه‌هایی به نام IFilters دارد. کار این فیلترها استخراج متن بدون فرمت، از فایل‌های باینری مرتبط و ارائه‌ی آن‌ها به موتور FTS می‌باشد.

نصب فیلترهای مخصوص FTS آفیس

اگر علاقمند هستید که بدانید در حال حاضر چه تعداد فیلترهای FTS بر روی سیستم شما نصب شده‌است، کوئری ذیل را اجرا

نمائید:

```
exec sys.sp_help_fulltext_system_components 'filter';
```

برای نمونه اگر آفیس بر روی سیستم شما نصب باشد، در حاصل کوئری فوق، فیلتری مانند offfilt.dll را نیز مشاهده خواهید کرد که به پسوندی مانند doc, ppt, xls و امثال آن انتساب داده شده است. فیلترهای آفیس را جداگانه نیز می‌توانید دریافت و نصب کنید (بدون نیاز به نصب کامل آفیس بر روی سرور):

[Microsoft Office 2010 Filter Packs](#)

این فیلترها تا نگارش 2013 آفیس را نیز پشتیبانی می‌کنند و اگر آپدیت ویندوز نیز روشن باشد، [سرویس پک 2](#) آن را نیز دریافت خواهید کرد.

پس از اینکه فیلترها را نصب کردید، باید آن‌ها را در وهله‌ی جاری SQL Server ثبت کرد:

```
exec sys.sp_fulltext_service 'load_os_resources', 1;
EXEC sp_fulltext_service 'update_languages';
EXEC sp_fulltext_service 'restart_all_fdhosts';
```

اکنون اگر مجدداً کوئری sys.sp_help_fulltext_system_components یاد شده را اجرا کنید. خروجی آن حدوداً 50 سطر خواهد بود؛ این اطلاعات را از کوئری ذیل نیز می‌توان بدست آورد:

```
select * from sys.fulltext_document_types;
```

اگر پس از نصب و همچنین ثبت و معرفی فیلترهای آفیس 2010 به بعد، هنوز تعداد 50 ردیف را ملاحظه می‌کنید (اکنون باید بیشتر از 160 مورد باشند)، نیاز است یکبار وهله‌ی جاری SQL Server را ری استارت کنید. برای اینکار در management studio بر روی وهله‌ی جاری، کلیک راست کرده و گزینه‌ی Restart را انتخاب کنید.

فیلترهای فوق علاوه بر اینکه امکان FTS را بر روی کلیه فایل‌های مجموعه آفیس میسر می‌کنند، امکان جستجو FTS را بر روی خواص ویژه اضافی آن‌ها، مانند نام نویسنده، واژه‌های کلیدی، تاریخ ایجاد و امثال آن نیز به همراه دارند.

FTS چگونه کار می‌کند؟

زبان‌های پشتیبانی شده توسط FTS را توسط کوئری ذیل می‌توانید مشاهده کنید:

```
select lcid, name from sys.fulltext_languages order by name;
```

کار FTS با word-breakers و stemmers شروع می‌شود. این‌ها کار آنالیز متن را بر اساس زبانی مشخص انجام می‌دهند. اگر زبان مدنظر توسط FTS پشتیبانی نمی‌شود، می‌توان از زبان انگلیسی و یا همچنین Neutral نیز برای آنالیز آن استفاده کرد. زبان Neutral جزو خروجی کوئری فوق با شماره آی دی صفر است.

word-breakers تک تک کلمات را (که به آن‌ها token نیز گفته می‌شود) تشخیص داده و سپس FTS آن‌ها را با فرمتی فشرده شده، درون ایندکس‌های مخصوص خود ذخیره می‌کند. کار stemmers تولید حالات inflectional (صرفی) یک کلمه بر اساس دستور زبانی مشخص است.

اهمیت آنالیز inflectional، در اینجا است که برای مثال اگر در متنی واژه‌ی jumps وجود داشت و کاربر در حین جستجو، jumped را وارد کرد، FTS بر اساس دستور زبان مورد استفاده، بیشتر، حالات مختلف صرفی jump را ذخیره کرده است و امکان انجام یک چنین کوئری پیشرفته‌ای را پیدا می‌کند.

نصب و فعال سازی Semantic Language Database

کار TFS تنها به خرد کردن واژه‌ها و آنالیز صرفی آن‌ها خلاصه نمی‌شود. در مرحله‌ی بعد، انجام Statistical semantic search میسر می‌شود. در اینجا SQL Server بر اساس آمار واژه‌های کلیدی استخراج شده، توانایی یافتن متونی مشابه و یا مرتبط را پیدا می‌کند. Semantic Search جزو تازه‌های SQL Server 2012 است.

برای اینکار نیاز است بانک اطلاعاتی Semantic language statistics نیز نصب شود. برای اطمینان از نصب بودن آن، کوئری ذیل را اجرا کنید:

```
select * from sys.fulltext_semantic_language_statistics_database;
```

اگر حاصل آن خالی بود، نیاز است مستقلاً نصب شود. این بانک اطلاعاتی ویژه را در یکی از دو مسیر ذیل

```
x64\Setup\SemanticLanguageDatabase.msi  
x86\Setup\SemanticLanguageDatabase.msi
```

در DVD یا فایل ISO نصب SQL Server 2012 می‌توانید پیدا کنید. فایل نصاب msi آن را اجرا کنید، دو فایل mdf و ldf را در مسیری که مشخص می‌کنید، کپی می‌کند. پس از آن نیاز است این بانک اطلاعاتی را Attach و همچنین ثبت کرد:

```
CREATE DATABASE semanticcsdb  
ON ( FILENAME = 'D:\SQL_Data\SemanticLanguageDatabase\semanticcsdb.mdf' )  
LOG ON ( FILENAME = 'D:\SQL_Data\SemanticLanguageDatabase\semanticcsdb_log.ldf' )  
FOR ATTACH  
GO  
  
EXEC sp_fulltext_semantic_register_language_statistics_db @dbname = N'semanticcsdb'  
GO
```

زمانیکه این بانک اطلاعاتی کپی می‌شود، دسترسی Write کاربر وارد شده به سیستم را در برگه‌ی Security فایل‌های mdf و ldf آن ندارد. به همین جهت ممکن است در حین Attach، پیام عدم دسترسی را دریافت کنید که با مراجعه به خواص فایل‌ها و تنظیم دسترسی Write کاربر جاری، مشکل برطرف می‌شود. پس از مراحل فوق، اگر مجدداً کوئری یاد شده بر روی sys.fulltext_semantic_language_statistics_database را اجرا کنید، یک سطر خروجی خواهد داشت.

نظرات خوانندگان

نویسنده: ابوالفضل رجب پور
تاریخ: ۱۳۹۲/۱۲/۱۶ ۰:۴۹

در مقایسه با لوسین کدام قوی تره؟ و یا پیشنهاد میشه؟
و بنظرتون معیارهای انتخاب برای استفاده از هرکدوم چیه؟

نویسنده: وحید نصیری
تاریخ: ۱۳۹۲/۱۲/۱۶ ۱:۱۴

از [لوسین](#) برای بانکهای اطلاعاتی سبکی که قابلیت های Full text search ندارند، بهتر است استفاده شود. برای مثال اگر از SQLite استفاده می کنید یا حتی SQL Server CE (سبک ترین نسخه ی SQL Server که یک بانک اطلاعاتی embedded محسوب می شود)، لوسین بسیار مناسب است.

برای نمونه [در سایت جاری](#) از آن برای تهیه موتور جستجوی سایت استفاده شده و یا حتی برنامه ی [سبک Viewer بانک اطلاعاتی سایت](#) که با فرمت XML است، از لوسین استفاده می کند.

به صورت خلاصه برای کارهای سبک و یا بانکهای اطلاعاتی embedded، استفاده از لوسین فوق العاده است.

اما برای کار با SQL Server کامل، واقعا نیازی به لوسین نیست. یک زیرساخت کامل و توکار برای Full Text Search دارد که با زبان T-SQL آن یکپارچه است. نگهداری و به روز رسانی آن توسط برنامه نویس در حد صفر است و یکبار که تعریف شد، به خوبی کار می کند. نگهداری ایندکسهای لوسین خودکار نیست و باید توسط برنامه نویس به صورت مجزا هر بار که اطلاعات تغییر می کند انجام شود.