

SQL Server به همراه تعدادی تابع سیستمی است که امکان مشاهده‌ی ریز جزئیات تشکیل دهنده‌ی ایندکس‌های FTS را فراهم می‌کند. در ادامه قصد داریم این موارد را بررسی کنیم.

### متد sys.dm\_fts\_index\_keywords

این متد محتوای full-text index یک جدول را باز می‌گرداند. از آن می‌توان برای موارد ذیل استفاده کرد:

- آیا واژه کلیدی خاصی جزو full-text index است؟
- چه تعداد رکورد دارای واژه‌ی کلیدی خاصی هستند؟
- متداولترین واژه‌های کلیدی موجود در ایندکس کدامند؟
- کدام واژه را می‌توان به عنوان stop word تشخیص داد؟ شاید پس از بررسی، تشخیص داده شود که بهتر است متداولترین واژه‌ی کلیدی ایندکس شده، به stop list اضافه شود.

```
SELECT *
FROM sys.dm_fts_index_keywords(DB_ID(DB_NAME()), OBJECT_ID(N'dbo.Documents'));
```

Results		Messages		
	keyword	display_term	column_id	document_count
80	0x00610067006700720065006700610074006...	aggregated	5	1
81	0x00610067006700720065006700610074006...	aggregating	4	1
82	0x00610067006700720065006700610074006...	aggregating	5	1
83	0x00610067006700720065006700610074006...	aggregation	5	2
84	0x00610067006700720065006700610074006...	aggregations	5	1
85	0x00610067006F	ago	5	1
86	0x0061006C0067006F0072006900740068006D	algorithm	5	2
87	0x0061006C00690067006E00650064	aligned	5	1
88	0x0061006C006C	all	4	2

### متد sys.dm\_fts\_index\_keywords\_by\_document

این متد اطلاعاتی را در سطح اسناد باز می‌گرداند. کاربردهای آن می‌توانند شامل موارد زیر باشند:

- یافتن جمع تعداد واژه‌های کلیدی که یک full-text index دارا است.
- آیا واژه‌ی کلیدی مورد نظر، در ردیف در حال بررسی وجود دارد؟
- یک واژه‌ی کلیدی چندبار در کل ایندکس ظاهر شده‌است؟
- یک واژه‌ی کلیدی در یک ردیف یا سند مشخص، چندبار تکرار شده‌است؟
- یک ردیف یا سند، از چند واژه‌ی کلیدی تشکیل شده‌است؟

```
SELECT
    I.document_id,
    D.title,
    I.display_term,
    I.occurrence_count
FROM sys.dm_fts_index_keywords_by_document(DB_ID(DB_NAME()), OBJECT_ID(N'dbo.Documents')) AS I
INNER JOIN dbo.Documents D
ON D.id = I.document_id;
```

	document_id	title	display_term	occurrence_count
1	4	Additivity of Measures	\$100.00	1
2	4	Additivity of Measures	\$130.00	1
3	4	Additivity of Measures	\$150.00	1
4	4	Additivity of Measures	\$2,000.00	2
5	4	Additivity of Measures	\$200.00	1
6	4	Additivity of Measures	\$230.00	1
7	4	Additivity of Measures	\$3,000.00	1
8	4	Additivity of Measures	\$350.00	1

#### متد sys.dm\_fts\_index\_keywords\_by\_property

در قسمت‌های قبل، خواص و متادیتای اسناد آفیس را نیز ایندکس کردیم. این متد، اطلاعات مرتبط با خواص اسناد موجود در full-text index را باز می‌گرداند.

کاربردهای آن:

- چه محتوایی، در خاصیتی مشخص از سندی معلوم، ذخیره شده‌است؟
- خاصیت مورد نظر چه اندازه بکار رفته و تکرار شده‌است؟
- چه اسنادی دارای خاصیتی مشخص هستند؟

```
SELECT
    I.document_id,
    D.title,
    I.display_term,
    I.property_id
FROM sys.dm_fts_index_keywords_by_property(DB_ID(DB_NAME()), OBJECT_ID(N'dbo.Documents')) AS I
INNER JOIN dbo.Documents D
ON D.id = I.document_id;
```

Results		Messages		
	document_id	title	display_term	property_id
1	1	Columnstore Indices and Batch Processing	test	1
2	2	Introduction to Data Mining	test	1
3	3	Why Is Bleeding Edge a Different Conference	test	1
4	4	Additivity of Measures	test	1

### متد sys.dm\_fts\_parser

متدهای قبلی که بررسی کردیم، نیاز به یک جدول و وجود full-text index بر روی آن دارند؛ اما متد dm\_fts\_parser خیر. این متد یک ورودی را گرفته و سپس تمام مراحل تهیه‌ی یک full-text index را به صورت پویا انجام می‌دهد. کاربردهای آن:

- درک اینکه موتور FTS با یک ورودی رشته‌ای چگونه رفتار می‌کند.
- استخراج ایندکس‌های یک متن و ذخیره‌ی دستی آن در یک جدول.
- استخراج واژه‌های کلیدی یک رشته.
- آنالیز پویای INFLECTIONAL (مانند مثال زیر)

```
SELECT
    display_term,
    keyword
FROM sys.dm_fts_parser(N'"Mycustom string"', 1033, NULL, 0);
```

Results		Messages	
	display_term	keyword	
1	mycustom	0x006D00790063007500730074006F006D	
2	string	0x0073007400720069006E0067	

```
SELECT *
FROM sys.dm_fts_parser('FORMSOF(INFLECTIONAL, ' + 'term' + ')', 1033, NULL, 0);
```

Results		Messages						
	keyword	group_id	phrase_id	occurrence	special_term	display_term	expansion_type	source_term
1	0x007400650072006D00270073	1	0	1	Exact Match	term's	2	term
2	0x007400650072006D00650064	1	0	1	Exact Match	termed	2	term
3	0x007400650072006D0069006E0067	1	0	1	Exact Match	teming	2	term
4	0x007400650072006D0073	1	0	1	Exact Match	terms	2	term
5	0x007400650072006D00730027	1	0	1	Exact Match	terms'	2	term
6	0x007400650072006D	1	0	1	Exact Match	term	0	term

در اینجا پارامتر دوم آن شماره زبان مورد استفاده است. پارامتر سوم مشخص کننده‌ی stop list می‌تواند باشد و پارامتر سوم حساسیت به لهجه را مشخص می‌کند.

## نظرات خوانندگان

نویسنده: امیرحسین ابراهیمیان  
تاریخ: ۱۹:۴۳ ۱۳۹۴/۰۴/۱۳

ممنون! میشه semantic search را با entityframework شبیه سازی کرد ؟

نویسنده: وحید نصیری  
تاریخ: ۱۹:۵۹ ۱۳۹۴/۰۴/۱۳

« [استفاده از Full text search توسط Entity Framework](#) »