

در قسمت‌های قبل، نحوه‌ی کار با فیلترهای FTS آفیس را بررسی کردیم. شرکت Adobe نیز برای جستجوی Full-Text بر روی فایل‌های PDF، یک iFilter خاص را طراحی کرده‌است که نسخه‌ی آخر آن را از آدرس ذیل می‌توانید دریافت کنید:

<http://www.adobe.com/support/downloads/product.jsp?product=1&platform=Windows>

یک تجربه‌ی مهم: نگارش 11 آن را با SQL Server X64 تست کردم کار نکرد. اما نگارش 9 کار می‌کند.

مستندات کامل

پس از نصب ابتدایی آن، مراحل ذیل را برای فعال سازی آن باید طی کرد:

1) تنظیم مسیر پوشه bin نصب فیلتر (مهم!)

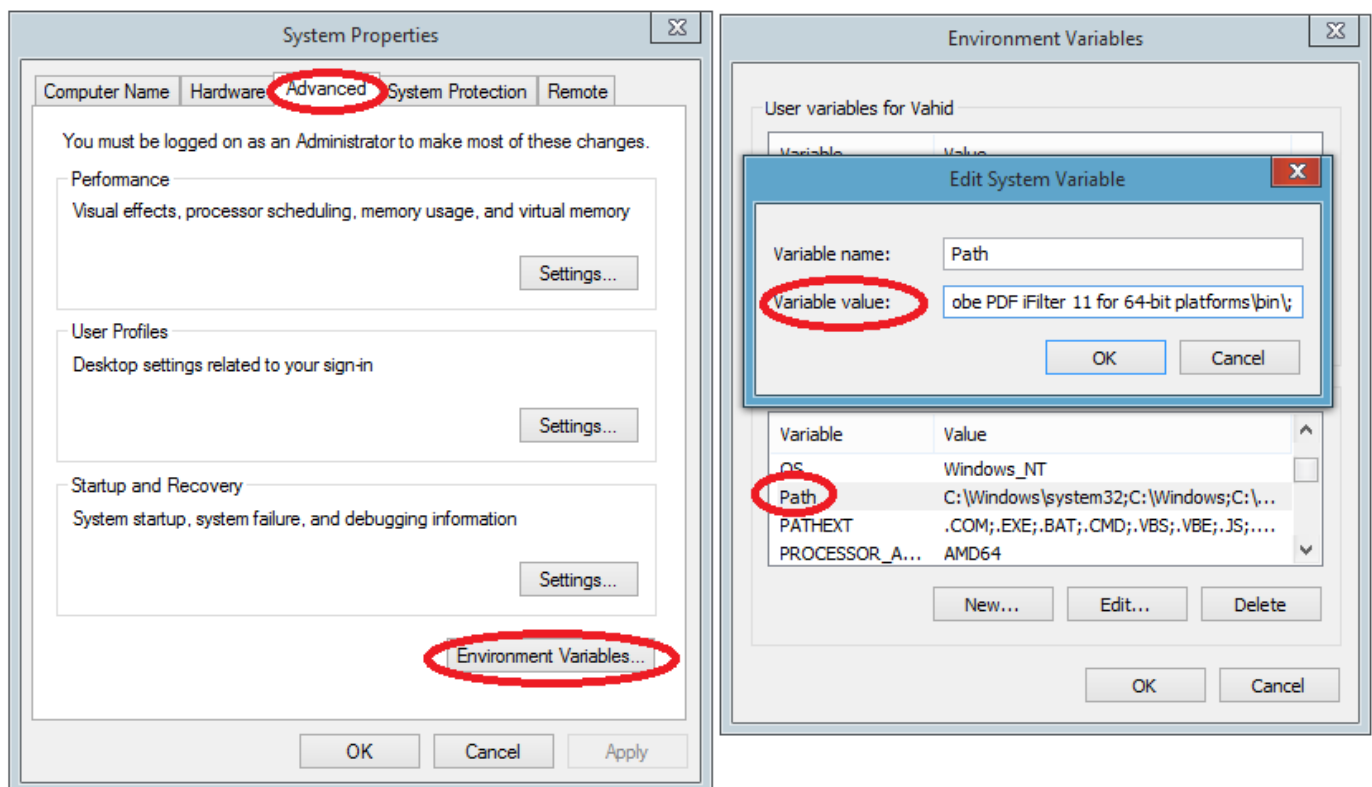
Start > Control Panel > System > Advanced
Environment Variables -> System Variables -> find PATH

مسیر فوق را در تنظیمات ویندوز یافته و سپس به انتهای Path، آدرس پوشه bin فیلتر نصب شده را اضافه کنید:

C:\Program Files\Adobe\Adobe PDF iFilter 9 for 64-bit platforms\bin\

دقت داشته باشید که این مسیر باید به \ ختم شود.

سپس کل سیستم را ری استارت کنید.



2) ثبت آن در وهله‌ی جاری SQL Server

برای این منظور ابتدا دستورات ذیل را اجرا کنید:

```
exec sys.sp_fulltext_service 'load_os_resources', 1;
EXEC sp_fulltext_service 'verify_signature', 0
EXEC sp_fulltext_service 'update_languages'; -- update language list
EXEC sp_fulltext_service 'restart_all_fdhosts'; -- restart daemon
reconfigure with override
```

گزینه‌ی verify_signature مربوط به فایل‌های iFilter ایی است که امضای دیجیتال ندارند. سپس در management studio یکبار بر روی وهله‌ی جاری کلیک راست کرده و گزینه‌ی Restart را انتخاب کنید (مهم).

3) پس از ری‌استارت SQL Server، اطمینان حاصل کنید که این فیلتر جدید نصب شده‌است:

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

Results Messages

	componenttype	componentname	clsid	fullpath	version
94	filter	.pdf	E8978DA6-047F-4E3D-9C78-CDBE46041603	C:\Program Files\Adobe\Adobe PDF iFilter 11 for 64-bit ...	11.0.1.36

و یا کوئری ذیل نیز برای این منظور مفید است:

```
SELECT document_type, path from sys.fulltext_document_types where document_type = '.pdf'
```

4) در ادامه برای استفاده از آن آزمایش ذیل را ترتیب خواهیم داد.

ایجاد یک جدول جدید که فایل‌های باینری PDF را در خود ذخیره می‌کند:

```
CREATE TABLE PdfDocuments
(
id INT IDENTITY(1,1) NOT NULL,
doctype NCHAR(4) NOT NULL,
doccontent VARBINARY(MAX) NOT NULL,
CONSTRAINT PK_PdfDocuments
PRIMARY KEY CLUSTERED(id)
);
```

ستون‌های doctype معرف نوع سند و doccontent ذخیره کننده‌ی محتوای کامل فایل‌های PDF خواهند بود. سپس چند رکورد را در آن ثبت می‌کنیم. برای نمونه دو مقاله‌ی خروجی PDF سایت جاری را در این جدول ثبت خواهیم کرد:

```
INSERT INTO PdfDocuments(doctype, doccontent)
SELECT
N'PDF',
bulkcolumn FROM OPENROWSET (BULK 'C:\Users\Vahid\Downloads\1732-DotNetTips.pdf', SINGLE_BLOB) AS doc;

INSERT INTO PdfDocuments(doctype, doccontent)
SELECT
N'PDF',
bulkcolumn FROM OPENROWSET (BULK 'C:\Users\Vahid\Downloads\1733-DotNetTips.pdf', SINGLE_BLOB) AS doc;
```

در ادامه علاقمندیم تا بر روی خواص و متادیتای فایل‌های PDF نیز بتوانیم جستجوی FTS انجام دهیم. به همین منظور search property list را نیز تعریف خواهیم کرد. همانطور که در قسمت‌های قبل عنوان شد، نیاز است GUID هر خاصیت را برای

تعریف از سازنده‌ی iFilter دریافت کرد. این اطلاعات در سند ذیل مستند شده‌اند:

<http://www.ifiltershop.com/downloads/pdfplusfilter/readme.html>

<http://msdn.microsoft.com/en-us/library/ms692560%28v=vs.85%29.aspx>

```
-- Search property list
CREATE SEARCH PROPERTY LIST PdfSearchPropertyList;
GO
ALTER SEARCH PROPERTY LIST PdfSearchPropertyList
ADD 'Author'
WITH (PROPERTY_SET_GUID = 'F29F85E0-4FF9-1068-AB91-08002B27B3D9',
PROPERTY_INT_ID = 4,
PROPERTY_DESCRIPTION = 'Author - author of a given item.');
```

در اینجا اگر علاقمند بودید، list stop معرفی شده در [قسمت‌های قبل](#) را نیز می‌توان افزود.

```
CREATE FULLTEXT STOPLIST SQLStopList;
GO
-- Add a stopwords
ALTER FULLTEXT STOPLIST SQLStopList ADD 'به' LANGUAGE 'English';
ALTER FULLTEXT STOPLIST SQLStopList ADD 'با' LANGUAGE 'English';
--.....
```

سپس یک کاتالوگ FTS و ایندکس Full-Text ایی را بر روی این جدول ایجاد می‌کنیم:

```
-- Full-text catalog
CREATE FULLTEXT CATALOG PdfDocumentsFtCatalog;
GO

-- Full-text index
CREATE FULLTEXT INDEX ON PdfDocuments
(
    [doccontent] TYPE COLUMN [doctype]
    Language 1033
    STATISTICAL_SEMANTICS
)
KEY INDEX PK_PdfDocuments
ON PdfDocumentsFtCatalog
WITH STOPLIST = SQLStopList,
    SEARCH PROPERTY LIST = PdfSearchPropertyList,
    CHANGE_TRACKING AUTO;
GO
```

[آیا کار می‌کند ؟ چیزی ایندکس شده‌است؟](#)

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

	document_id	display_term	occurrence_count
1...	2	عمل	2
1...	2	عملیة	2
1...	1	عنوان	4
1...	2	عنوان	1
1...	2	فایل	1

انجام دو کوئری بر روی آن . یکی برای یافتن متنی ساده و دیگری برای یافتن خواص

```
SELECT *
FROM PdfDocuments
WHERE CONTAINS(doccontent, N'است')

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

	id	doctype	doccontent
1	1	.pdf	0x255044462D312E340A25E2E3CFD30A31352030206F626A...
2	2	.pdf	0x255044462D312E340A25E2E3CFD30A31382030206F626A...

	document_id	display_term	property_id
1	1	مسعود	1
2	2	مسعود	1
3	1	پاکدل	1
4	2	پاکدل	1