

Foxit PDF IFilter 1.0

For Microsoft SQL Server 2005

User Manual

© 2009 Foxit Corporation.

All rights reserved.

Foxit PDF IFilter 1.0 for Microsoft SQL Server 2005

Copyright © 2009 Foxit Corporation. All Rights Reserved.

No part of this document can be reproduced, transferred, distributed or stored in any format without the prior written permission of Foxit.

Anti-Grain Geometry - Version 2.3

Copyright (C) 2002-2005 Maxim Shemanarev (<http://www.antigrain.com>)

Permission to copy, use, modify, sell and distribute this software is granted provided this copyright notice appears in all copies. This software is provided "as is" without express or implied warranty, and with no claim as to its suitability for any purpose.

Contents

License Agreement for Foxit PDF IFilter	4
Overview	5
Why PDF IFilter?.....	5
Why Foxit PDF IFilter?	5
How Foxit PDF IFilter works?	6
Key benefits	6
Getting Started	7
System requirements.....	7
Installing Foxit PDF IFilter	7
Updating Foxit PDF IFilter.....	9
Uninstalling Foxit PDF IFilter	9
Working with Foxit PDF IFilter	11
Setting up full-text search	11
Uploading PDFs.....	12
Starting a full population.....	13
Searching PDFs	14
More help.....	15
Additional Information	16
Microsoft Developer Support	16
Language Support	16
FAQ	17
Contact Us	19

License Agreement for Foxit PDF IFilter

By using software of Foxit Corporation or its subsidiaries ("Foxit"); you agree to the following terms and conditions. If you do not agree with such terms and conditions; do not use the software. The terms of an end user license agreement accompanying a particular software file upon installation or download of the software shall supersede the terms presented below.

If the software is designed for use with an application software product (the "Host Application") published by Foxit; Foxit grants you a non-exclusive license to use such software with the Host Application only; provided you possess a valid license from Foxit for the Host Application. Except as set forth below; such software is licensed to you subject to the terms and conditions of the End User License Agreement from Foxit governing your use of the Host Application.

Foxit PDF IFilter can be used on desktop environment for free. However, users need to acquire additional license when they use Foxit PDF IFilter on server environment, e.g. Microsoft SQL Server.

DISCLAIMER OF WARRANTIES: YOU AGREE THAT FOXIT HAS MADE NO EXPRESS WARRANTIES TO YOU REGARDING THE SOFTWARE AND THAT THE SOFTWARE IS BEING PROVIDED TO YOU "AS IS" WITHOUT WARRANTY OF ANY KIND. FOXIT DISCLAIMS ALL WARRANTIES WITH REGARD TO THE SOFTWARE; EXPRESS OR IMPLIED; INCLUDING; WITHOUT LIMITATION; ANY IMPLIED WARRANTIES OF FITNESS FOR A PARTICULAR PURPOSE; MERCHANTABILITY; MERCHANTABLE QUALITY OR NONINFRINGEMENT OF THIRD PARTY RIGHTS. Some states or jurisdictions do not allow the exclusion of implied warranties; so the above limitations may not apply to you.

LIMIT OF LIABILITY: IN NO EVENT WILL FOXIT BE LIABLE TO YOU FOR ANY LOSS OF USE; INTERRUPTION OF BUSINESS; OR ANY DIRECT; INDIRECT; SPECIAL; INCIDENTAL; OR CONSEQUENTIAL DAMAGES OF ANY KIND (INCLUDING LOST PROFITS) REGARDLESS OF THE FORM OF ACTION WHETHER IN CONTRACT; TORT (INCLUDING NEGLIGENCE); STRICT PRODUCT LIABILITY OR OTHERWISE; EVEN IF FOXIT HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. Some states or jurisdictions do not allow the exclusion or limitation of incidental or consequential damages; so the above limitation or exclusion may not apply to you. Foxit's liability is limited to replacement or refund of only the virus-infected software downloaded from Foxit website.

Overview

Welcome to Foxit PDF IFilter 1.0 -- a professional tool that helps you to index PDF files.

Foxit PDF IFilter is an application designed to help users index a large amount of PDF documents and then quickly find text within these documents. The PDF documents can not only be files, but also email attachments or database records.

Foxit PDF IFilter supports PDF 1.7 specification and can be used in many Microsoft search engines, such as Windows Indexing Service, MSN Desktop Search, Internet Information Server, SharePoint Portal Server, Windows SharePoint Services (WSS), Site Server, Exchange Server, SQL Server and all other products based on Microsoft Search technology. With Foxit PDF IFilter, all these search engines can find PDF files with ease.

Why PDF IFilter?

PDF IFilter is necessary for several reasons listed below:

- Microsoft creates and maintains office file formats, such as .doc, .xls, .ppt, etc. However, many other file formats are created and maintained by other companies. For example, PDF was created by Adobe.
- With the advantages of security, reliability and accessibility, PDF is now being used by more and more people all over the world.
- PDF IFilter is a plug-in that allows Microsoft search engines to index PDF files so that they become searchable. Without an appropriate PDF IFilter, contents of a PDF file cannot be indexed.

Why Foxit PDF IFilter?

Foxit PDF IFilter is favored by a large number of government organizations and well-known companies because of the following reasons:

- Foxit PDF IFilter helps to do a two-stage search – Index and Search, which makes finding information much easier.
- Foxit PDF IFilter supports multi-threading and makes full use of multi-CPU/core in a server.
- Foxit PDF IFilter supports native 64-bit code.
- Smaller and faster than other PDF IFilters.

- Supports Chinese/Japanese/Korean PDF documents.

How Foxit PDF IFilter works?

Foxit PDF IFilter acts as a plug-in for full-text search engines. A search engine usually works in two steps:

Step 1, the search engine goes through a designated place, e.g. a file folder or a database, and indexes all documents or newly modified documents, including PDF documents, in the background and creates internal data to store indexing result.

Step 2, a user specifies some keywords he/she would like to search and the search engine answers the query immediately by looking up the indexing result and responds to the user with all the documents that contains the keywords.

During Step 1, the search engine itself doesn't understand format of a PDF document. Therefore, it looks in windows registry for an appropriate PDF IFilter and finds the Foxit PDF IFilter. Foxit PDF IFilter understands PDF format. It filters out embedded formatting and extracts text from the document and return text back to the search engine. Further information about IFilter specification can be found on the [Microsoft Web site](#).

Key benefits

Here's a list of benefits that Foxit PDF IFilter brings to you:

- Integrates with existing operating systems and tools within your company.
- Provides an easy solution to search within PDF documents located on local computer, local network and intranet.
- Greatly increases your ability to accurately locate information.
- Much smaller and faster than IFilter offered by other vendors.
- Supports Chinese/Japanese/Korean PDF documents.
- Supports native 64-bit code.
- Supports multi-threading and takes full use of multiple-CPU in a server.

Getting Started

This section provides everything you'll need to know about getting started with Foxit PDF IFilter, including its system requirements, installation and uninstallation.

System requirements

Foxit PDF IFilter 1.0 requires the same environment as that of SQL Server 2005. For the information on SQL Server 2005 system requirements, please follow the link here: <http://www.microsoft.com/sql/prodinfo/sysreqs/default.msp>

At a minimum, each environment must also contain Microsoft SQL Server 2005.

Installing Foxit PDF IFilter

Foxit PDF IFilter has two versions respectively for 32-bit Windows and 64-bit Windows. You can choose one of them to install into your machine according to your system requirements. Before installing Foxit PDF IFilter, please note that Microsoft SQL Server 2005 has been installed successfully.

To install Foxit PDF IFilter 1.0, please do the following:

Step 1, create a new functioning database to store PDF files.

- A. Run SQL Server Management Studio.
- B. Create a new database: right-click the Database folder in the Object Explorer and select New Database. Type your Database name and click OK.
- C. Create a table: right-click the Table folder in the new database and select New Table.
- D. Complete the details in the "Column Name" column, the "Data Type" column, "Length" column, and "Allow Nulls" column. Foxit PDF IFilter indexes PDF files in SQL Server based on your specified column names. For example, you may add some fields for PDF files as below:

Table - dbo.Table_1*		Summary
Column Name	Data Type	Allow Nulls
PdfText	varbinary(MAX)	<input checked="" type="checkbox"/>
Ext	nvarchar(50)	<input checked="" type="checkbox"/>
PdfID	int	<input type="checkbox"/>
		<input type="checkbox"/>

PdfText: stores binary PDF files, whose data type should be varbinary(MAX).

Ext: stores the file extension of PDF files. It's set as ".pdf" in this example. For more information, please refer to

<http://technet.microsoft.com/en-us/library/ms142499.aspx>

PdfID: designates a unique ID for each PDF file, which will act as the return value when you're trying to search PDFs.

Note: The example of the table above is applied in this manual.

E. Click the Save button and enter a name for the table.

Step 2, download and install Foxit PDF IFilter.

A. Download Foxit PDF IFilter from Foxit Website (<http://www.foxitsoftware.com>) onto the machine where you wish to install it.

B. Run the Foxit PDF IFilter 1.0 Setup program to install the filter on the server.

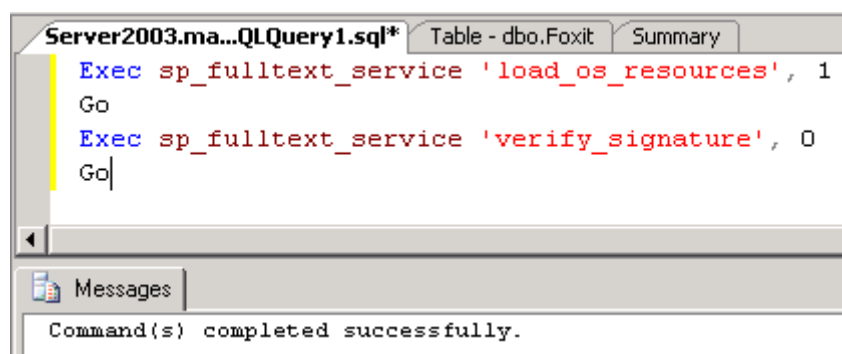
Notes:

1. Foxit PDF IFilter appears in the list of programs on the Programs Start Menu after installation.
2. Foxit PDF IFilter has no user interface.
3. The installation package will unzip a language file called fpdfcjk.bin into the installation folder, which will help you to index PDF documents containing Chinese/Japanese/Korean characters.

Step 3, after the installation of Foxit PDF IFilter, execute the following commands to register Foxit PDF IFilter.

Click the "Database Engine Query" button on the toolbar, and execute the command to register Foxit PDF IFilter:

```
Exec sp_fulltext_service 'load_os_resources', 1
Exec sp_fulltext_service 'verify_signature', 0
```



The message "Command(s) completed successfully." means you have registered Foxit PDF IFilter successfully.

Step 4, restart the server or stop and start SQL server and the Full-text Search engine.

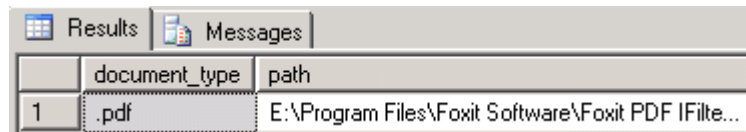
- A. Click Start > Programs > Microsoft SQL Server 2005 > Configuration Tools > SQL Server Surface Area Configuration.
- B. In the pop-up SQL Server 2005 Surface Area Configuration dialog box, please click to select Surface Area Configuration for Services and Connections.
- C. Select MSSQLSERVER > Service, and click the Stop button and then Start.
- D. Select Full-Text Search > Service, and click the Stop button and then Start.

Step 5, confirm the registration.

Check for success with use of these commands to confirm that the pdf file type has an association to Foxit PDF IFilter:

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

Check the results to confirm success visually. For example,



	document_type	path
1	.pdf	E:\Program Files\Foxit Software\Foxit PDF IFilte...

Updating Foxit PDF IFilter

When updating Foxit PDF IFilter, you need to uninstall the old version of Foxit PDF IFilter to install the new one. You do not have to reboot your machine after update.

Generally speaking, Foxit PDF IFilter consists of a core layer and an installation layer. When you upgrade the installation layer, you do not need to re-do a full crawl. But when you upgrade the core layer, you can decide if you need to rebuild index on your computer according to the importance of the updates by checking what's new details on Foxit Official Website.

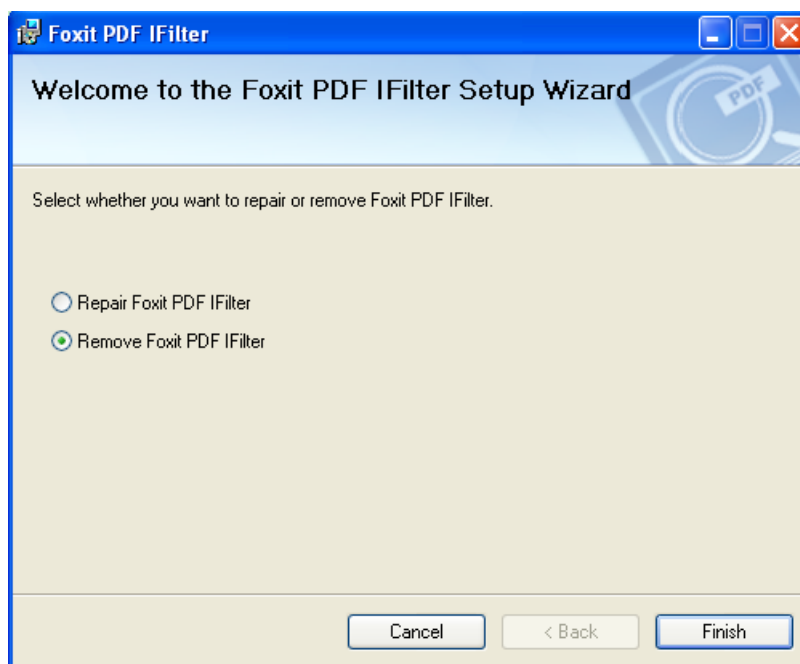
Uninstalling Foxit PDF IFilter

To uninstall Foxit PDF IFilter, please perform the following:

- Open the Start Menu > Settings > select "Control Panel".
- Click on the "Add or Remove Programs" tab.
- Click on the "Foxit PDF IFilter" record and then on the "Remove" button to uninstall it.

Or you can double-click the Foxit PDF IFilter.msi setup file that you have used to install Foxit PDF IFilter, and do the following:

- In the Foxit PDF IFilter dialog box, select Remove Foxit PDF IFilter.



- Click Finish button to uninstall it completely.

Working with Foxit PDF IFilter

Foxit PDF IFilter is able to search text in PDF files. After you install Foxit PDF IFilter, you should upload your PDF documents and perform a full population with Microsoft SQL Server 2005 before you start a searching.

You don't have to do full population every time you add some new PDF files. You may schedule the incremental population for all the new ones.

Setting up full-text search

To set up full-text search, you should take the following steps in order:

1. Confirming the Full-text search service.

Please relaunch Microsoft SQL Server Management Studio, and execute the following command:

```
print FULLTEXTSERVICEPROPERTY('IsFullTextInstalled')
```

Results:

1= Full-Text installed

0= Full-Text not installed

Null= invalid input, or error

2. Installing Full-Text component and service.

3. Enabling Full-Text indexing

In order to use Full-Text indexing function, you must first verify that Full-Text Indexing function has been enabled. To enable Full-text indexing, please do one of the following:

- Launch Microsoft SQL Server Management Studio. Right-click the database you have created > select Properties > choose Files tab > check "Use full-text indexing" > OK.
- Execute the following command:

```
exec sp_fulltext_database 'enable'
```

4. Creating a full-text catalog to store full-text indexes.

Please execute the following command:

```
create fulltext catalog pdfCatalog
```

5. Creating a unique index

Do the following:

- Open the table you have created > right-click one of the column names > select "Set Primary Key".

Column Name	Data Type	Allow Nulls
PdfText	varbinary(MAX)	<input checked="" type="checkbox"/>
PdfId	int	<input type="checkbox"/>
Ext	nchar(10)	<input checked="" type="checkbox"/>
		<input type="checkbox"/>

- Execute the following command:

```
create unique index ui_pdfIndex on dbo.pdfTable (PdfId)
```

6. Create a full-text index

Execute the following command:

```
create fulltext index on dbo. pdfTable
(
PdfText - Please assign the column name for full-text index.
)
key index ui_pdfIndex on pdfCatalog
with change_tracking auto
```

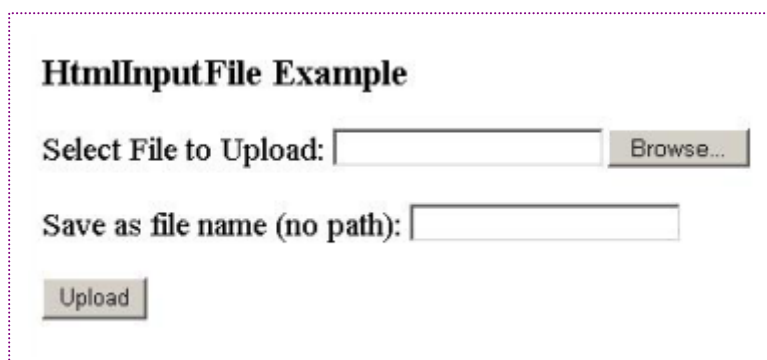
Uploading PDFs

To upload PDF files to Microsoft SQL Server, you can either write a segment of codes or copy the codes below:

```
using System.Web.UI.HtmlControls;
using System.Data.SqlClient;
using System.IO;
public void OnUpload(Object sender, EventArgs e)
{
// Create a byte[] from the input file
int len = File1.PostedFile.ContentLength;
byte[] PdfText = new byte[len];
File1.PostedFile.InputStream.Read(PdfText, 0, len);
// Insert the image and comment into the database
SqlConnection connection = new
```

```
SqlConnection(@"server=IP address or server
name;database=pdfindex;uid=username;pwd=password");
try
{
    connection.Open();
    SqlCommand cmd = new SqlCommand("insert into pdftable "
        + "(PdfText, PdfId, Ext) values (@PdfText, @PdfId, @Ext)", connection);
    // Insert values into the three columns.
    cmd.Parameters.Add("@PdfText", PdfText);
    cmd.Parameters.Add("@PdfId", 1); // You can only upload a pdf file at a time.
    cmd.Parameters.Add("@Ext", ".pdf");
    cmd.ExecuteNonQuery();
}
finally
{
    connection.Close();
}
}
```

When you're done, you will be prompted to upload your PDF files as below:



HtmlInputFile Example

Select File to Upload:

Save as file name (no path):

Note: The codes above uses ADO.NET as the data access mechanism, and C# as the development language.

Starting a full population

When you have finished the steps mentioned above, you should start a full population for the PDFs you uploaded. Actually, there are two ways under SQL Server 2005 to synchronize full-text indexes with their table data: Full Population and Incremental Population.

Full Population - A population when the index entries are built for all the rows in all the full-text catalog's tables.

Incremental Population - A population which only adjusts index entries for rows that have been added, deleted, or modified after the last population.

You may select to start a full population or an incremental population according to your needs.

- To start a full population, right-click the table you created > Full-Text index > Start Full Population.
- To start an incremental population, right-click the table you created > Full-Text index > Start Incremental Population.

Searching PDFs

After you finish indexing PDF files, you can start a search to find text within PDFs. Before you start a search, you have to verify if the full-text catalog or full-text index has been fully populated.

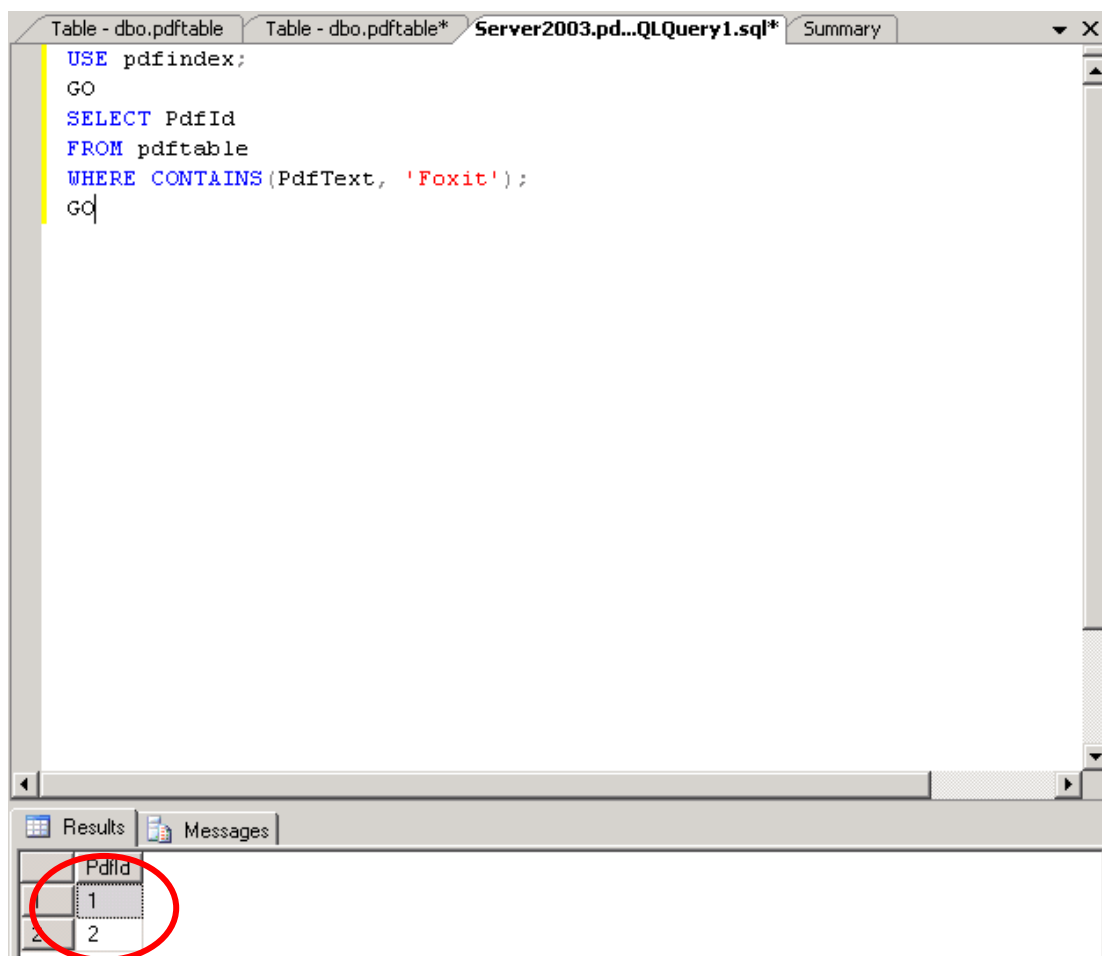
To check the population status, please execute the following command:

```
select fulltextcatalogproperty('PdfText', 'populatestatus')
```

You can start a search if the return value is 0.

To start a search in SQL Server 2005, please execute the following command:

```
USE pdfindex;  
GO  
SELECT PdfId  
FROM pdftable  
WHERE CONTAINS(PdfText, 'Foxit');  
GO
```



More help

For more information on how to work with Microsoft SQL Server 2005, please refer to Microsoft SQL Server 2005 Help included in the product.

Additional Information

Microsoft Developer Support

Foxit PDF IFilter is a Dynamic Link Library (DLL) file that provides a bridge between a Microsoft indexing client and a library of PDF files. It consists of code that understands the PDF Format as well as code that can interface with the indexing client. When an indexing client needs to index content from PDF files, it will look in its registry for an appropriate DLL and will find the Foxit PDF IFilter. Foxit PDF IFilter will return text to the indexing client. The indexing client will then index the results and return the appropriate results to the user.

Language Support

This section lists languages supported by Foxit PDF IFilter 1.0.

- German
- French
- Italian
- English
- Spanish
- Portuguese
- Russian
- Chinese (traditional and simplified)
- Korean
- Japanese
- Arab
- Hebrew

FAQ

1. What is Foxit PDF IFilter?

Foxit PDF IFilter is an application designed to help users index a large amount of PDF documents and then quickly find text within these documents. It acts as a plug-in for full-text search engines that scans documents for text and properties (also called attributes), extracts text from documents, filters out formatting and retaining information about text position, and extracts properties of an entire document.

2. Is Foxit PDF IFilter free?

Foxit PDF IFilter can be used for free on desktop environment. However, users need to purchase licenses when they use Foxit PDF IFilter on server environments.

3. I have a computer with one hyper-threading processor inside. Do I need to buy a license for 2 CPUs for the machine?

We will only count the "real" numbers of processors reported by your system. Please go to Start > Programs > Accessories > System Tools > System Information to check your CPU numbers.

4. What is the Foxit PDF IFilter's role in the PDF indexing process and PDF text search?

Foxit PDF IFilter acts as a plug-in for full-text search engines. A search engine usually works in two steps:

Step 1, the search engine goes through a designated place, e.g. a file folder or a database, and indexes all documents or newly modified documents, including PDF documents, in the background and creates internal data to store the indexing result.

Step 2, a user specifies some keywords he would like to search and the search engine answers the query immediately by looking up the indexing result and responds to the user with all the documents that contain the keywords.

During Step 1, the search engine itself doesn't understand the format of a PDF document. Therefore, it looks in the windows registry for an appropriate PDF IFilter and finds the Foxit PDF IFilter. Foxit PDF IFilter understands the PDF format. It filters out embedded formatting and extracts text from the document and returns text back to the search engine. Further information about IFilter specification can be found on the Microsoft Web site.

5. Do I have to remove Adobe PDF IFilter to install Foxit PDF IFilter?

Yes.

6. How do I verify the correct installation of Foxit PDF IFilter?

Firstly, please check if Foxit PDF IFilter appears in the list of programs on the Program Start Menu after installation.

Secondly, you can download IFilter detection tools to inspect IFilter installation information in the registry, e.g. the filtreg.exe program.

Thirdly, install Windows Desktop Search to verify if the text in PDF files can be found.

7. How do I upgrade Foxit PDF IFilter? Do I need to reboot the machine?

You need to uninstall the old version of Foxit PDF IFilter to install the new one. You do not have to reboot your machine after update.

8. Do I have to rebuild index after updating?

That depends. Foxit PDF IFilter consists of a core layer and an installation layer. When you upgrade the installation layer, you do not need to re-do a full index. But when you upgrade the core layer, you may need to rebuild index on your computer.

9. Why can't I find some PDF files?

There might be three reasons:

First, Foxit PDF IFilter can not index password-protected PDF files.

Second, the PDF file doesn't contain searchable text. You can check if there is searchable text in the PDF file by opening it in Foxit Reader and selecting the text with the Select Text tool. Or you can enter into the Text Viewer mode to select text. If neither method highlights any text, it is likely that the PDF contains a source image of text which Foxit PDF IFilter can not index.

Third, the PDF file doesn't contain text that is properly encoded. If the file contains searchable text, yet you cannot find the text with Find Tool in Foxit Reader, then the text may not be properly encoded.

10. Why can't Foxit PDF IFilter work after Microsoft updates? Do I have to re-install it?

If you installed the Foxit PDF IFilter on your search server, the IFilter does not work after you install a Microsoft update. It is because that the upgrade program automatically overwrites the registry value of Foxit PDF IFilter with that of Adobe PDF IFilter.

To make it work again, you must reinstall Foxit PDF IFilter. In this case, you do not have to uninstall IFilter. You can just run the Foxit IFilter MSI program again and select the repair function.

In addition, we will contact Microsoft to address this issue shortly.

Contact Us

Feel free to contact us should you need any information or have any problems with our products. We are always here, ready to serve you better.

- **Office Address:**

Foxit Corporation
39819 Paseo Padre Parkway
Fremont CA 94538
USA

- **Mailing Address:**

Foxit Corporation
39819 Paseo Padre Parkway
Fremont CA 94538
USA

- **Sales:**

1-866-MYFOXIT or 1-866-693-6948 (8AM-5PM PST Monday - Friday)
510-438-9090 (8AM-5PM PST Monday - Friday)
408-307-9358 (8AM-5PM PST Monday - Friday)

- **Support:**

1-866-MYFOXIT or 1-866-693-6948 (8AM-5PM PST Monday - Friday)
979-446-0280 (6AM-5PM PST Monday - Friday)

- **Fax:**

510-405-9288

- **Website:**

www.foxitsoftware.com

- **E-mail:**

Sales and Information - sales@foxitsoftware.com

Technical Support - support@foxitsoftware.com

Marketing Service - marketing@foxitsoftware.com