

Kofax PDF iFilter for SharePoint

Installation Guide

Version: 4.0.0

Date: 2020-07-09

The logo for KOFAX, consisting of the word "KOFAX" in a bold, blue, sans-serif font.

© 2020 Kofax. All rights reserved.

Kofax is a trademark of Kofax, Inc., registered in the U.S. and/or other countries. All other trademarks are the property of their respective owners. No part of this publication may be reproduced, stored, or transmitted in any form without the prior written permission of Kofax.

Table of Contents

- Document purpose 2**
- Target Audience 2**
- Notes 2**
- Levels of access..... 2**
- How to use iFilter 2**
 - Access text layer 3
 - Use OCR 4
 - Steps to create the wsdi.ini file manually..... 5
 - Test the indexing service..... 6

Document purpose

This guide provides assistance for the installation and configuration of the PDF iFilter designed for Microsoft SharePoint 2013, 2016, or 2019. This application dynamically indexes PDF files placed in SharePoint so that its search facility can rapidly find text strings inside PDF files.

Target Audience

System administrations for network environments using the Microsoft SharePoint repository for file storage and sharing.

Notes

This guide is written for SharePoint 2010. Differences exist for SharePoint 2013 and 2016; these are detailed below.

iFilter respects PDF security and does not extract text if document security settings prohibit it.

Levels of access

There are two levels of access to texts within PDF files:

- Access PDF files, pages or areas that already have a text layer.
- Access image-only PDF files, pages or areas that contain text.

SharePoint 2013, 2016, and 2019 introduce a built-in indexer application for PDF files, but it can handle only files with an accessible text layer. It is possible to replace the built-in filter with the Kofax product that gives you the ability to also index and search texts in image-only PDF files, pages or areas.

How to use iFilter

There are three typical tasks related to iFilter.

[Access text layer](#). – This task provides access to the existing text layers within PDF files. It is valid for all SharePoint versions (see the footnote).

[Use OCR](#). – This task works for all SharePoint versions. Proceed with this task to grant access to the text within image-only PDF page areas. iFilter can automatically detect image-only page areas deemed to contain text, and performs OCR (Optical Character Recognition) to generate the text layer. Perform this task also for replacing the built-in PDF indexer with the Kofax product.

[Test the indexing service](#). – This task describes how to stop a running indexer and start indexing with the Kofax product, allowing the results to be searched to verify successful installation.

PDF files with a security prohibition against text extraction cannot be handled by the Kofax iFilter, nor by its OCR services.

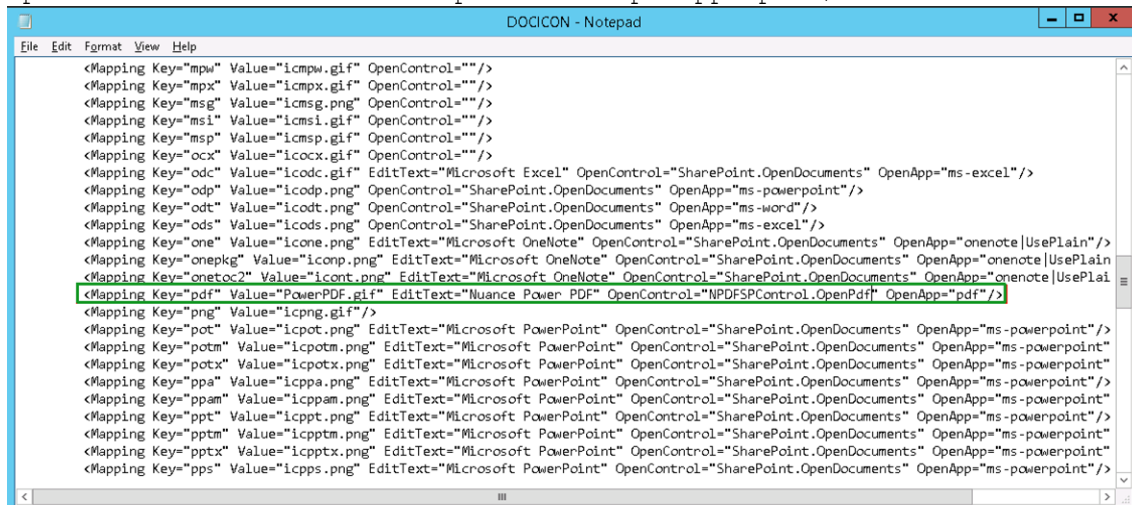
Access text layer

The installation is done on the SharePoint server, so one instance is enough for multiple users.

Important! On SharePoint 2013: Please ensure before starting that you have installed the hotfix KB2883000. This hotfix allows 3rd party iFilters to replace the built-in PDF filter for SharePoint.

1. Launch `setup-iFilter-17277.100.x64.exe` to begin the installation process, using the Kofax iFilter Installation Wizard. Accept the EULA and other customer ID information, and enter a valid serial number: the number supplied with your Kofax Power PDF volume license should be used.
2. Continue installation by accepting or modifying the installation folder (SharePoint will find it anywhere on your system). In the following screen verify that the checkbox **SharePoint** is displayed and selected. This confirms that SharePoint is correctly installed and detected. Click **Start installation** and at the end click **Finish**.
3. Download the small (17×17 pixels) Kofax PDF icon image from the Kofax delivery media the path given in the introduction and save it as `PowerPDF.gif` (or with the name you prefer).
4. Copy the above icon file to `C:\Program Files\Common Files\Microsoft Shared\Web Server Extensions\15\Template\Images`¹
5. Edit the `DOCICON.XML` file located in `C:\Program Files\Common Files\Microsoft Shared\Web Server Extensions\15\Template\XML`¹
6. Under the section `<By Extension>`, determine whether or not there is a line for the PDF file extension. If so, replace it with the following. If not, add the following line anywhere within the section:

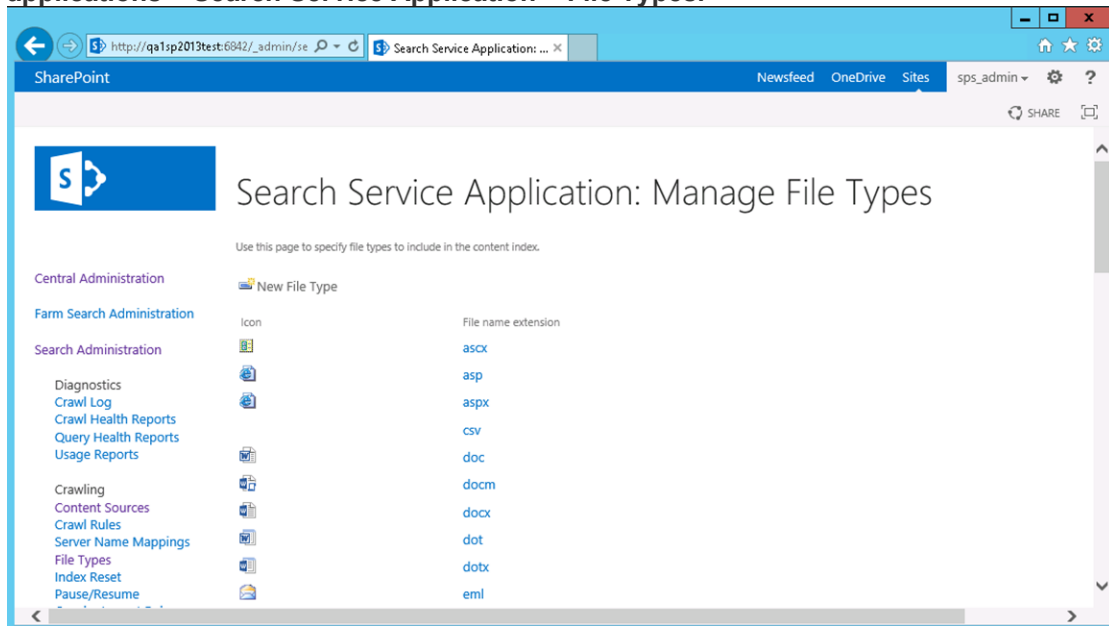
```
<Mapping Key="pdf" Value="PowerPDF.gif" EditText=" Kofax Power PDF"
OpenControl="NPDFSPControl.OpenPdf.3" OpenApp="pdf"/>
```



The icon file name in this line must be identical to the one used in step three.

¹ The example uses drive C, use the drive to which Web Server Extensions is installed on your server. The folder named "15" relates to SharePoint 2013 only. For SharePoint 2016 and 2019 use the name "16".

7. Go to **SharePoint's Central Admin > Application Management > Manage service applications > Search Service Application > File Types.**



Click **New File Type**, enter `pdf` in the edit box and click **OK**.

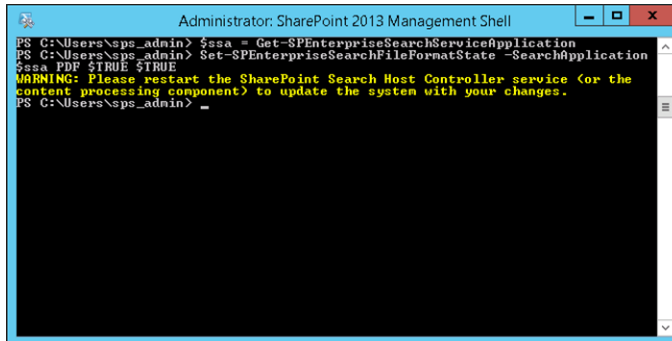
Use OCR

By default iFilter does not perform OCR on image-only PDF documents. This additional procedure is necessary to enable this feature.

The procedure is identical for all versions of SharePoint. It allows OCR to generate searchable results even on PDF page areas with image-only text, providing that it is not security protected. Be aware that OCR generated text is liable to be less accurate than extracting text directly. Quality strongly depends on image quality: a pixel density of at least 300 dpi is recommended. However, OCR in most cases can generate output adequate for indexing and searching. To enable the feature you must place a `wdsi.ini` file in the folder `%programdata%\Kofax\PDF\V1\` on the SharePoint server.

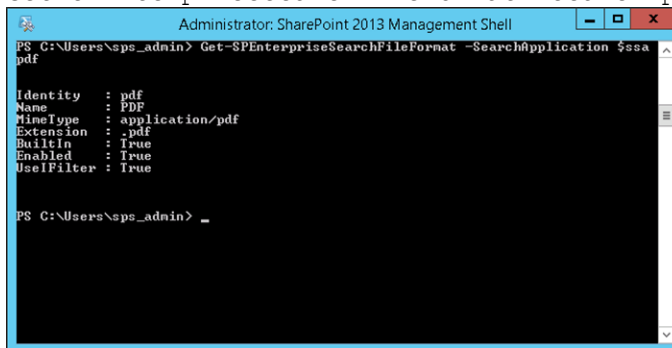
1. In Power PDF installed on a desktop computer go to **File > Options > General > Integrations**. Configure the Windows Desktop Search Integration according to the desired behavior then click **OK** in the Options dialog to have the changes take effect.
2. Then take the `wdsi.ini` file from the desktop computer's folder `%programdata%\Kofax\PDF\V1\` and copy it to the SharePoint server. As an alternative, the `wdsi.ini` file can be manually created on the server. See [Steps to create the wdsi.ini file manually](#) for details.
3. Navigate from the Windows Start menu and select **SharePoint Management Shell**. Please ensure that you run it as Administrator otherwise you may fail to perform the commands in the following two steps.
4. To replace the built-in PDF iFilter with a third-party one, perform the following two commands:
 - `$ssa = Get-SPEnterpriseSearchServiceApplication`

- `Set-SPEnterpriseSearchFileFormatState -SearchApplication $ssa PDF $TRUE $TRUE`

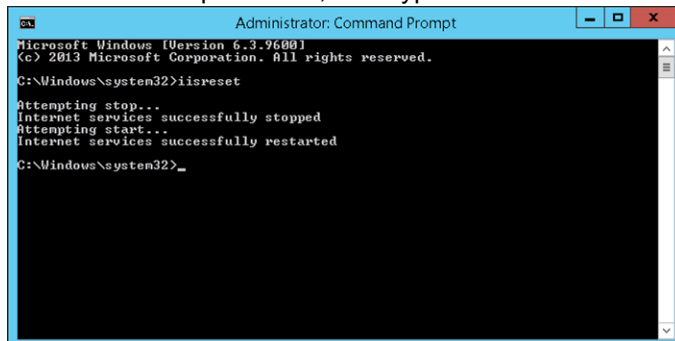


5. Run the following command to retrieve information about "pdf" in the Search service application.

- `Get-SPEnterpriseSearchFileFormat -SearchApplication $ssa pdf`



6. The returned value "UseFilter: True" means that you have successfully used a third-party PDF iFilter.
7. To make the changes take effect, you need to perform an iisreset operation. Launch a Command Prompt (cmd.exe) with Administrator rights (Run as Administrator). Type `cmd` to go into the Command Prompt window, then type `iisreset`.



8. Exit the Command Prompt window once IIS stops and restarts successfully.

Steps to create the wsd.ini file manually

As an alternative, the `wsdi.ini` file can be manually created on the server.

1. Use a plain text editor to edit the file, which should contain the following section:

```

[Main]
OCRPages=
OCRFirst=
    
```

OCRPageNum=
OCReDiscovery=

2. Set the values as you prefer:

- Set OCRPages to 1 if you would like to run OCR on pages that do not contain a text layer, otherwise set it to 0.
- If you would like to OCR images on pages that already contain a text layer to find additional text set OCReDiscovery=1. OCReDiscovery is effective only if OCRPages=1. When this is selected, an existing text layer is always retained and OCR is applied only to the remainder of the page.
- If you would like to set a limit to the number of pages that can be OCR-ed in a document set OCRFirst to 1 otherwise set it to 0. OCRFirst is effective only if OCRPages=1.
- You can set the OCR page limit in OCRPageNum. For example, OCRPageNum=3 means that on-ly the first three pages might be OCR-ed in a document. OCRPageNum is effective only if OC-RFirst=1. This is useful for long documents and reflects the fact that some PDF documents have image-only cover pages with the remainder having a text layer.

Test the indexing service

This procedure is valid for all supported SharePoint versions.

Restart the SharePoint Search Service with the following command lines:

- `net stop spsearchhostcontroller`
- `net start spsearchhostcontroller`
- `net stop osearch152`
- `net start osearch152`

² The process named “osearch15” relates to SharePoint 2013 only. For SharePoint 2016 and 2019 use the name “osearch16”.