

11.11 How to Include Additional Document Types in the Search Process (SSP)

Although SharePoint will allow you to upload any document type not specifically blocked into any document library or alternatively allow you to expose additional storage locations through SharePoint (e.g. a web part that exposes a Shared Network Folder or Exchange Public Folder), if these locations contain additional document types unknown to SharePoint the search engine will at most scan the titles by default. In order to expose the internal textual content and meta-data contained in the document type several pieces of configuration need to be done: Adding the appropriate Icon, File Type, i-Filter, and re-crawling the content.

11.11.1 Default File Types and associated iFilter availability in SharePoint

NOTE: For a list of the default File Types and iFilter Availability in SharePoint see APPENDIX C.

11.11.2 Step1: Add the Icon for New Document Types (1x per Front End Web Server)

One of the first steps you might take to enabling a new content type in SharePoint is to locate and add a new Icon to be displayed when this document type is stored in SharePoint. It is interesting to note that EVEN IF you have installed this document type on your operating system, SharePoint also requires you to re-identify it as SharePoint stores these content types internally and does NOT make use of the base system's file type registrations.

IMPORTANT: All Icons that you want to use with SharePoint should be the standard icon size of 16x16 pixels. Also, your icon must be stored in the GIF format. If you do not follow these rules you may have difficulties with the icon being displayed correctly.

IMPORTANT: These steps MUST be repeated on **EACH FRONT END WEB SERVER** or your results will vary depending on which server services a particular request.

1. Once you have located the appropriate 16x16 pixel GIF based icon file, you need to make a copy of it and name the original IC***.GIF and the copy to ***16.GIF where the "****" represent the file extension.

EXAMPLE: assume that we would like to store Word Perfect documents in SharePoint: We would take the icon file for Word Perfect (wpd.gif) and copy it. Re-name the original ICWPD.gif. Re-name the copy to WPD16.gif.

2. Next you will move both of these icon files to the following directory location "C:\Program Files\Common Files\Microsoft Shared\web server extensions\12\TEMPLATE\IMAGES". Once the files are located here they are universally available to all SharePoint requests on that server.
3. Next you must **edit** the **DOCICON.XML** file to add the appropriate entry to expose this image to the appropriate program. Open the C:\Program Files\Common Files\Microsoft Shared\web server extensions\12\TEMPLATE\XML\DOCICON.XML file and add an entry at the end for your new document type and then be sure to **save** the file.

EXAMPLE: to add the ICWPD16.GIF to the appropriate file in the DOCICON.XML file you must add it to the <ByExtension> section (you can add it just before the closing </ByExtension> tag or put it into alphabetical order like the existing entries) in a <Mapping > tag:
 <ByExtension><Mapping Key="wpd" Value="icwpd.gif" /> </ByExtension>

4. In order for this change to be picked up by SharePoint you must run an **IISRESET**.
Start menu > **Run...** - type **IISRESET** and press **Enter**.
5. Repeat these steps 1x on each Front End Web Server
6. Now you should see the appropriate icon next to documents of this type in SharePoint

11.11.3 Step 2: Adding New File Types for SharePoint to Index (1x per SSP)

The next step to searching a new document type correctly is to tell SharePoint to attempt to search into this new file type when it locates an instance of it by adding this new type to the Manage File Types page of your SSP.

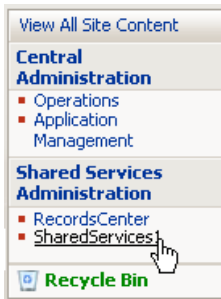
Manage file types to Include a new type

You can specify file types (by using their file name extensions) to include in the content index when crawling content sources. For example, you might want to include files with .abc and .def extensions in the index.

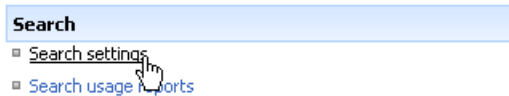
NOTE: You must also add the appropriate iFilter for any proprietary file types see section 11.11.4

To manage file types, you must first open the **Manage File Types** page:

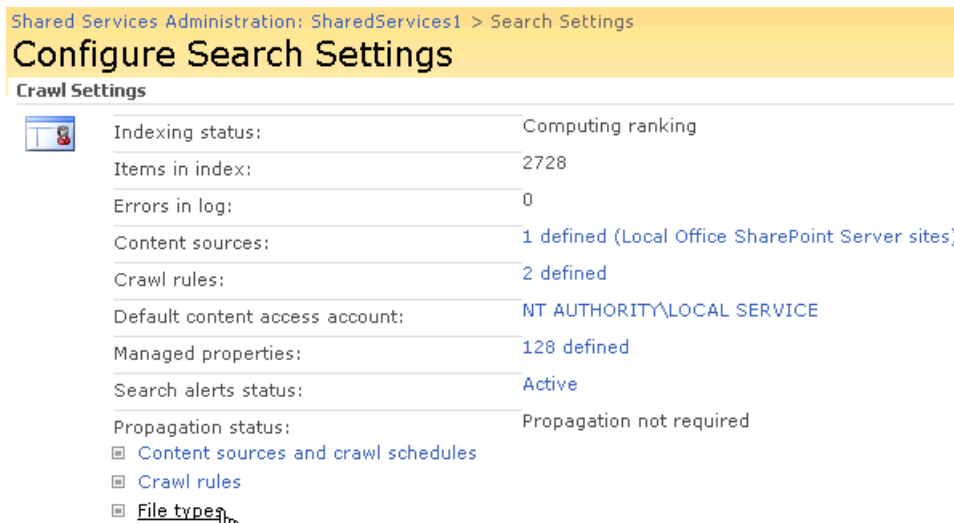
1. Open the administration page for the Shared Services Provider (SSP).
 - a. On the Quick Launch bar, under **Shared Services Administration**, click **SharedServices1** (or whatever your Shared Services is named).



2. On the **Shared Services Administration Home** page, in the **Search** section, click **Search settings**.

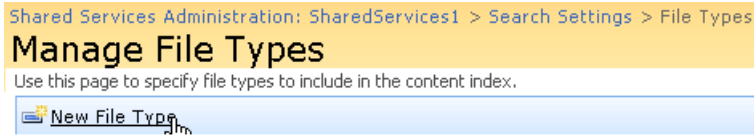


3. On the **Configure Search Settings** page, in the **Crawl Settings** section, click **File types**.

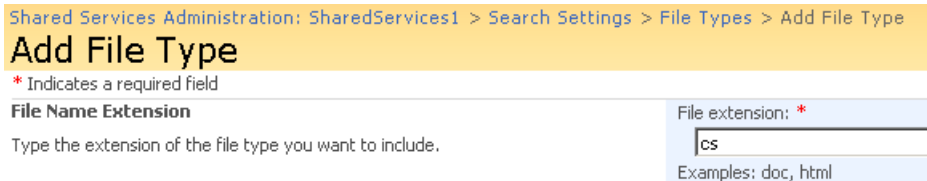


Add a file type





1. On the **Manage File Types** page, click **New File Type**.



2. On the **Add File Type** page, in the **File extension** box in the **File Name Extension** section, type the file name extension for the file type that you want to add (for example, type **doc**).

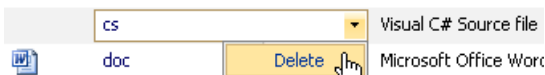


3. Click **OK**.

Icon	File name extension	File type
	ascx	ASP.NET User Control
	asp	Active Server Document
	aspx	ASP.NET Server Page
	cs	Visual C# Source file

Delete a file type

1. On the **Manage File Types** page, in the list of file name extensions, click **Delete** on the menu for the extension of the file type that you want to delete.



2. In the message box, click **OK** to confirm that you want to delete the file type.

11.11.4 Step 3: Installing the iFilter (1x per Index Server)

The next step in the process of adding a new document type into Search is to locate and install the document's iFilter. The iFilter is utilized in extracting the text and metadata from the document type in a uniform manner so that the indexing process may capture and categorize this information.

IMPORTANT: The actual installation of the iFilter will vary according to the way it is packaged up. Some iFilter's will only require running the Setup.exe file, others will require you to extract and install the iFilter using regsrv32.exe, while others may require something altogether different. The main Point to keep in mind is that iFilter technology incorporate into MS Search and MS Search incorporates into SharePoint, so utilize the instructions from the iFilter provider as to how to install their technology correctly.

Installation requiring regsrv32.exe

1. Download and extract out the dll files to somewhere convenient: For the purpose of this example we will create a new subdirectory called "%systemroot%\system32\SharePoint iFilter".
2. Copy your dll's into this subdirectory.
3. Next we need to register this dll for use on our system: **Start** menu > **Run...** - type **regsrv32.exe** "%systemroot%\system32\SharePoint iFilter\NameofDll.dll" and press **Enter**.
4. In order for this change to be picked up by SharePoint you must run an **IISRESET**. **Start** menu > **Run...** - type **IISRESET** and press **Enter**.
5. you must now add this file type to those indexed by SharePoint see section 11.11.3

Installing Adobe's iFilter Example

IMPORTANT: Adobe stopped creating a new iFilter as a separate component after version 6.0 (called pdffilt.dll). Many locations on the internet still link to this as the "ONE" to download and install with SharePoint (found at this location Adobe iFilter 6: <http://www.adobe.com/support/downloads/detail.jsp?ftpID=2611>). On this site Adobe advises you that the iFilter is now incorporated into, among other things, Acrobat Reader. Currently there is no possibility of extracting and utilizing the current pdf iFilter outside of installing one of Adobe's Acrobat products on your server (i.e. **AcroRdIF.dll** and located at **C:\Program Files\Adobe\Reader 8.0\Reader\AcroRdIF.dll**) The issue is that this iFilter dll is not independent, it requires several others to run correctly which cannot be installed independently of one of Adobe's Acrobat products.

IMPORTANT: Simply defaulting to Adobe's older pdffilt.dll is also NOT an option as this will be unable to crawl many documents created with a newer version of Acrobat.

IMPORTANT: Microsoft has several kb articles related to the pdf iFilter (927675, 832809, 928619). As of September 2007 **NONE** of these will truly fix the problems with the version 8.1 of Adobe's iFilter, so do **NOT** bother with them.

11.11.5 The Correct way to set up Adobe's pdf iFilter is as follows:

1. First you must **download** and **install** the Acrobat product of choice on your servers. For the purpose of this example we will install **Acrobat Reader 8.1**. located at <http://www.adobe.com/products/acrobat/readstep2.html>
2. Add the **pdf** file extension to the **File types** crawled by SharePoint:
 - a. From **SharePoint 3.0 Central Administration** select **<Name of SharedService Provider> > Search Settings > File Types > New File Type** (Add extension **pdf** here)
3. **Modify the following Registry** keys by changing their "Default" value to the new **CLSID of the Adobe iFilter: {E8978DA6-047F-4E3D-9C78-CDBE46041603}**
 - a. **HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Office server\12.0\Search\Setup\ContentIndexCommon\Filters\Extension\.pdf**
set the Default value to **{E8978DA6-047F-4E3D-9C78-CDBE46041603}**
 - b. **HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Shared Tools\Web Server Extensions\12.0\Search\Setup\ContentIndexCommon\Filters\Extension\.pdf**
set the Default value to **{E8978DA6-047F-4E3D-9C78-CDBE46041603}**
4. Add the **Installation directory** of the Adobe Reader v.8.1 to the **System Path**.

NOTE: This is necessary because the Adobe filter .dll file (**AcroRdIF.dll**) depends on the **AcroRd32.dll** and **AGM.dll** files. These files are located in the **Drive_Letter:\Program Files\Adobe\Reader 8.0\Reader folder**. Because the **Drive_Letter:\Program Files\Adobe\Reader 8.0\Reader** path is **not** included in the system path on the computer that is running SharePoint Server 2007, the SharePoint Server 2007 Search service cannot properly scan the .pdf files.

 - a. Click **Start**, click **Control Panel**, and then double-click **System**.
 - b. In the **System Properties** dialog box, click the **Advanced** tab, and then click **Environment Variables**.
 - c. In the **Environment Variables** dialog box, double-click **Path** in the **System Variables** area.
 - d. In the **Edit System Variable** dialog box, put the **cursor** at the **end** of the text in the Variable value box, and then type **;Drive_Letter:\Program Files\Adobe\Reader 8.0\Reader**
 - e. Click **OK** three times.
5. **Recycle the search service:**
 - a. **net stop oSearch**
 - b. **net start oSearch**
6. Now you need to conduct a Full Crawl of your content sources and you should be finished.

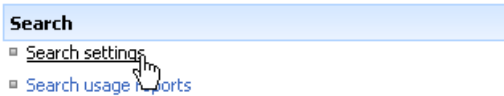
11.11.6 Step 4: Re-crawl your content (1x per Content Source)

whenever you add a new file type into SharePoint you MUST implement a FULL CRAWL of your Indexes or these new types will not be picked up correctly.

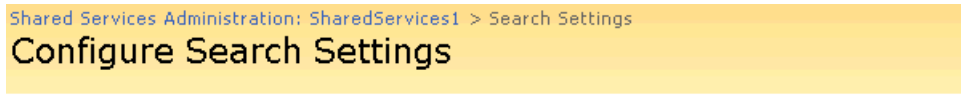
1. On the Quick Launch bar, under **Shared Services Administration**, click **SharedServices1** (or whatever your Shared Services is named).



2. On the **Shared Services Administration Home** page, in the **Search** section, click **Search settings**.



3. On the Search page, Select **Content sources and crawl schedules**



Crawl Settings

A screenshot of the 'Crawl Settings' page. It features a table of settings and a list of expandable options. The table includes: Indexing status (Idle), Items in index (2728), Errors in log (0), Content sources (1 defined (Local Office SharePoint Server sites)), Crawl rules (0 defined), Default content access account (NT AUTHORITY\LOCAL SERVICE), Managed properties (128 defined), Search alerts status (Active), and Propagation status (Propagation not required). Below the table is a list of expandable options: Content sources and crawl schedules, Crawl rules, File types, Crawl logs, Default content access account, Metadata property mappings, Server name mappings, Search-based alerts, Search result removal, and Reset all crawled content.

Indexing status:	Idle
Items in index:	2728
Errors in log:	0
Content sources:	1 defined (Local Office SharePoint Server sites)
Crawl rules:	0 defined
Default content access account:	NT AUTHORITY\LOCAL SERVICE
Managed properties:	128 defined
Search alerts status:	Active
Propagation status:	Propagation not required

- [Content sources and crawl schedules](#)
- [Crawl rules](#)
- [File types](#)
- [Crawl logs](#)
- [Default content access account](#)
- [Metadata property mappings](#)
- [Server name mappings](#)
- [Search-based alerts](#)
- [Search result removal](#)
- [Reset all crawled content](#)

4. Select the drop down off of each **content source** and select **Start Full Crawl**.
5. You have completed the iFilter addition.