









Installation of the Attachment Indexing plugins

Table of Contents

- [List of Supported File Types](#)
- [Installation of plugins to search attached files on Windows Server](#)
- [Installation of plugins to search attached files on Linux/Unix Server](#)
- [What to do after the installation of indexing plug-ins?](#)
 - [How to enable automatic indexing of file attachments when they are uploaded?](#)
 - [How to index the attached files manually?](#)
 - [How does PHPKB search the content of the attached PDF Files?](#)
 - [How to Disable Indexing for Specific File Types?](#)

PHPKB knowledge base software is able to index text content of the files attached to knowledge articles in order to make them searchable. Attachment indexing is supported in all editions of PHPKB knowledge base software. Some document types can be searched without any additional tools, others need PHP modules enabled or third-party tools (plugins) installed. All these modules (plugins) and third-party tools are free.

List of Supported File Types

File Type	Supported Format	Required Tool (Plug-in)
 MS Office 2003 Word Document	.DOC	AntiWord (free) is required
 MS Office 2003 Excel Workbook	.XLS	xlhtml (free) is required
 MS Office 2003 PowerPoint Presentation	.PPT	ppthtml (free) is required
 MS Office 2007 Word Document	.DOCX	PHP ZIP library is required
 MS Office 2007 Excel Workbook	.XLSX	PHP ZIP library is required
 MS Office 2007 PowerPoint Presentation	.PPTX	PHP ZIP library is required
 Adobe PDF Documents	.PDF	pdftotext (free) is required
 Plain-text Documents	.TXT, .HTM, .HTML, .CSV, .XML	No Plugin Required

Installation of plugins to search attached files on Windows Server

We strongly recommend you use the latest version of PHP. These plugins work under PHP 5.3+ correctly. Earlier versions of PHP have [bugs](#) and may freeze when launching external programs (e.g. attachment indexation plugins) using the Windows command line.

[Download the latest PHP package for Windows](#) (VC9 x86 Non Thread Safe is recommended).

PHP 5.3 doesn't support ISAPI anymore. So you need to use FastCGI instead.

1. Enabling Required PHP Extensions (Modules) on Windows Server

You need to enable certain PHP modules in order to index the content of MS Office 2007 documents.

1. Find the "ext" subdirectory of your PHP installation (it is C:\PHP\ext\ by default).
2. Check if the following files exist in that folder: php_mbstring.dll, php_zip.dll.
3. If any of these files do not exist, you should run PHP installation and install appropriate modules (Mbstring or PHP ZIP respectively).

Note: If you have PHP 5.3 or higher, you need not enable PHP ZIP extension as it is already built-in to the PHP engine.

4. Open the php.ini configuration file of your PHP engine in any text editor such as notepad.
5. Search for the "extension=" (without quotes).
6. You'll find the section with the list of PHP extensions. Some of them are commented with the # symbol.
7. You should enable these modules by removing the comment symbol (#).
8. Save the php.ini file.
9. Restart the web server for changes to take effect.

2. Attachment Indexing Plugins Installation on Windows Server

1. The "phpkb/admin/" folder contains the following plugins:

xlhtml (required for Excel 2003 files)

pphtml (required for PowerPoint 2003 files)

pdftotext (required for PDF files)


- If your server runs under IIS, you should grant the Internet Guest Account with the read and execute permissions to the "admin" folder. If it is an Apache server, you should check which user runs the Apache process and grant read and execute permissions for that user.
- There is a folder called "antiword" under the "admin" folder of your PHPKB installation. You need to copy that folder to "C:\\" of your server so that its path becomes "C:\antiword\".
- Add read and execute permissions for the web-server to this folder and its contents as well.

Installation of plugins to search attached files on Linux/Unix Server

Please follow the instructions below to install/enable the attachment indexation plugins and required PHP extensions.

1. Enabling Required PHP Extensions (Modules) on Linux Server


1. Open the php.ini configuration file.
2. Find the "extension_dir" parameter. It indicates the path to the PHP extensions directory. Go to that directory and check that zip.so file exists there.
3. Add reference for the PHP ZIP extension to the php.ini:

```
 extension=zip.so
```


4. Restart the webserver for changes made in php.ini to take effect.

2. Attachment Indexing Plugins Installation on Linux Server

If your system (RedHat / Fedora / CentOS) supports [Yum Package Manager](#) you can run this command instead to install necessary modules:

```
 # yum install poppler-utils
```


OR run the following command to install necessary modules on a system that has an APT library (e.g. Ubuntu, Debian):

```
 $ sudo apt-get install poppler-utils
```


How to Install the "xlhtml" and "pphtml" Package on Ubuntu?

See below for quick step-by-step instructions of SSH commands, Copy/Paste to avoid miss-spelling or installing a different package by mistake.

1. Run the update command to update package repositories and get the latest package information.

```
 $ sudo apt-get update -y
```

2. Run the install command with -y flag to quickly install the **xlhtml package** and dependencies.

```
 $ sudo apt-get install -y xlhtml
```

3. Run the install command with -y flag to quickly install the **pphtml** package and dependencies.

```
$ sudo apt-get install -y pphtml
```

4. Check the system logs to confirm that there are no related errors.

Note: -y flag means to assume yes and silently install, without asking you questions in most cases.

Note: In case **pphtml** & **xlhtml** do not get installed using the above commands then you can search them [here](#) and install accordingly.

What to do after the installation of indexing plug-ins?

1. Go to the PHPKB admin control panel.
2. Go to the **Tools » Manage Settings » Miscellaneous** tab » **Search Settings** section.
3. Enable the checkbox on the " **Search Attached Files** " and on each document type as shown in the image below.

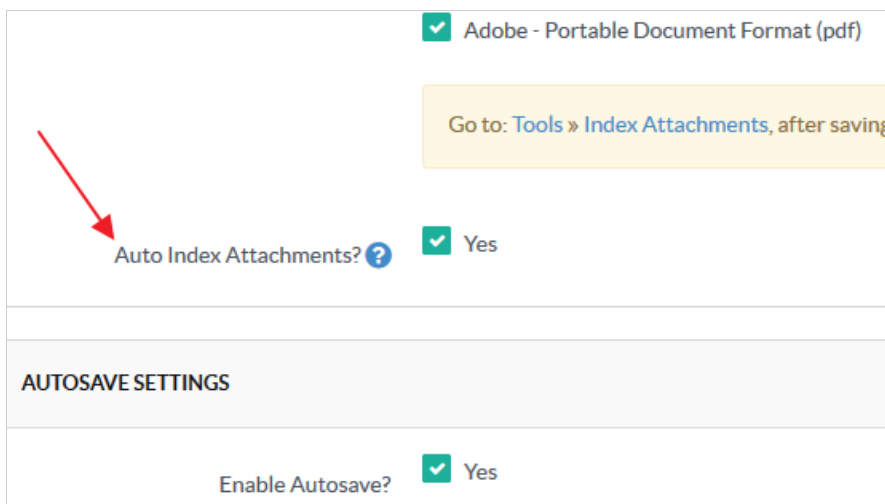
- ☒ Default Plugin (htm, html, txt, csv, xml)
- ☒ MS Word Documents 95-2003 (doc)
- ☒ MS Word Documents 2007 and onwards (docx)
- ☒ MS PowerPoint Presentations 95-2003 (ppt)
- ☒ MS PowerPoint Presentations 2007 and onwards (pptx)
- ☒ MS Excel Workbooks 95-2003 (xls)
- ☒ MS Excel Workbooks 2007 and onwards (xlsx)
- ☒ Adobe - Portable Document Format (pdf)

Go to: [Tools » Index Attachments](#), after saving the selected file attachments.

4. Click on the " **Save Changes** " button to save the settings.

How to enable automatic indexing of file attachments when they are uploaded?

If you would like to auto-index the attached files as soon as they are uploaded, then there is another setting under the " **File Upload Settings** " section of the " **Miscellaneous** " tab. You can set the checkbox for "Index Attachments" as shown in the image below.



☒ Adobe - Portable Document Format (pdf)

Go to: [Tools » Index Attachments](#), after saving

Auto Index Attachments? ☒ Yes

AUTOSAVE SETTINGS

Enable Autosave? ☒ Yes

Now you can upload attachments and they will be automatically indexed for search.

How to index the attached files manually?

You can also manually run indexation for existing file attachments whenever required from the "Tools" » "Index Attachments" section of the admin control panel as shown below.

phpkb ENTERPRISE v9.0 ML

Search here...

KB Menu

YOU ARE HERE: Tools / Index Attachments

EN

NAVIGATION

- Dashboard
- Categories
- Articles
- Downloads
- Comments
- Tickets
- Glossary
- News
- Languages
- Users
- Groups
- API
- Statistics
- Tools
- Manage Settings
- Optimize Database
- KB Backup
- Import Data
- Sitemap Generator
- Index Attachments**
- Email Harvester

Index Attachments

This section allows you to index text content of files attached to the knowledge base articles to make them searchable on front-end area.

Note: Please perform file indexing one after another to prevent script timeout. Maximum script timeout on this server is 600 seconds and generally, a total of 25 MB files takes about 1 minute to be indexed. However, it depends on various factors, like server load, cache etc.

Index Files

- PDF (1 Files)**
Index: 10 Files **Start**
- Word (4 Files)**
Index: 10 Files **Start**
- Excel (1 Files)**
Index: 10 Files **Start**
- PowerPoint (2 Files)**
Index: 10 Files **Start**

Powered by PHPKB Knowledge Base Software © 2005-2020 Chadha Software Technologies

How does PHPKB search the content of the attached PDF Files?

PHPKB knowledge base software is able to index the text content of PDF documents and make them searchable. It converts a PDF file to text file format to search its contents. It uses "pdftotext utility" to convert Portable Document Format (PDF) files to plain text. It reads the PDF file and writes a text file thus making itself able to search within the contents of PDF documents.

How to Disable Indexing for Specific File Types?

It may be necessary to disable attachment indexing for specific file types if the conversion is unable to process the attachments due to size or other errors. When disabled, the conversion will be bypassed for only specified file types. Attachment indexing can also be disabled for all types if desired.

1. Login to the PHPKB admin control panel.
2. Go to the **Tools » Manage Settings » Miscellaneous Settings**.
3. Remove the checkbox on the " **Search File Attachments** " if you would like to disable indexing for all file types.
4. Remove the checkbox for specific file types if you would disable indexing for those file types.

