

# Enable searching in files

---

Text-based file searching/indexing (txt, html, etc.) is enabled by default. If you want to include searching PDF or Word documents then you need to do the following.

## Searching in PDFs

To enable search in PDFs you need to:

- Install a program called **xpdf**
- Ensure **Settings** points to where you installed it
- Ensure that PHP has access to your xpdf directory. (Check your `open_basedir` PHP setting in `php.ini`)
- Ensure that PHP can run the `system` function. (Check your `disable_functions`, `safe_mode_exec_dir` PHP settings in `php.ini`)

### Install xpdf

- xpdf is available from [www.foolabs.com/xpdf/download.html](http://www.foolabs.com/xpdf/download.html)
- Download and install it.

### Update the setting to point to xpdf

Once you have installed xpdf, you also need to set the correct path it in the Settings.

- You can find this under Settings menu, **Settings -> Admin -> XPDF installation path**
- Make sure that this points to the directory where you installed xpdf. For example: `/usr/local/bin/` or `c:/wwwroot/xpdf/`
- Set this to 'off' to de-activate this option.
- When you click "Save", test pdf file will be parsed/indexed and error occurs if it failed.

### Test xpdf from command line

Test to see if xpdf is working by running the following command from the command line:

```
$/path_to_xpdf/pdftotext -raw file_read.pdf file_write.txt;
```

Test xpdf from command line using PHP and included test file:

```
$ cd /path/to/kbp_directory  
$ php -r "system('/path_to_xpdf/pdftotext -raw admin/extra/file_extractors/extract_test.pdf file_write.txt');"
```

## Searching in Word 2007/2010, Excel 2007/2010 or Open Office document files

To enable search in .docx, .xlsx and .odt documents you need to:

- Install a [PHP Zip](#) extension if you do not have one
- You can see if you have it installed in *Home -> Setup Tests* tab in your KBPublisher installation

## Searching in Word 2003 and below files

To enable search in Word documents you need to:

- Install either **catdoc** or **Antiword**
- Ensure **Settings** points to where you installed it
- Ensure that PHP has access to your catdoc directory. Check your `open_basedir` PHP setting in `php.ini`.
- Ensure that PHP can run the `exec` function. Check your `disable_functions`, `safe_mode_exec_dir` PHP settings in `php.ini`.

### Install catdoc

- catdoc is available from <http://www.wagner.pp.ru/~vitus/software/catdoc/>
- Download and install it.

### Install Antiword

- **Antiword** is available from <http://www.winfield.demon.nl/>
- Download and install it.

### Update the setting to point to catdoc

Once you have installed catdoc, you also need to set the correct path it in the Settings.

- You can find this under Settings menu, **Settings -> Admin -> catdoc installation path**
- Make sure that this points to the directory where you installed catdoc. For example: `/usr/local/bin/` or `c:/wwwroot/catdoc/`
- When you click "Save", test pdf file will be parsed/indexed and error occurs if it failed.

### Test catdoc from command line

Test to see if catdoc is working by running the following command from the command line:

```
$/path_to_catdoc/catdoc -w file_read.doc;
```

Test catdoc from command line using PHP and included test file:

```
$ cd /path/to/kbp_directory
$ php -r "system('/path_to_catdoc/catdoc -w admin/extra/file_extractors/extract_test.doc');"
```

### Test antiword from command line

Test to see if antiword is working by running the following command from the command line:

```
$ /path_to_antiword/antiword -t file_read.doc;
```

Test antiword from command line using PHP and included test file:

```
$ cd /path/to/kbp_directory
$ php -r "system('/path_to_antiword/antiword -t admin/extra/file_extractors/extract_test.doc');"
```

## Turning PDF or Word search off

If you don't want to allow searching on PDF or Word documents, change the setting in *XPDF installation path* or *catdoc installation path* to **OFF**.

---

Article ID: 224

Last updated: 01 Mar, 2018

Updated by: Leontev E.

Revision: 9

User Manual v7.0 -> Files -> Enable searching in files

<https://www.kbpublisher.com/kb/entry/224/>