

Search within files

In most cases, Tiki can search the textual content within files. For example, the text in a .docx file. For images, please see [OCR Indexing](#).

In most cases, this relies on utilities on the server. To check support on your server, use Tiki [Check](#).

It is possible, once you enabled "Automatic indexing of file content" (Control Panels, File Galleries, [Search Indexing tab](#)), to index the content of the files which are in the [File Gallery](#) so they can be retrieved by a [search](#). If you have a script that extracts the file content into a text, you can associate the script to the Mime type and the files content will be indexed.

If you want to search on files in the file galleries, you must provide handlers to extract the text for the file's MIME type (some may still work by default). The commands, such as *strings* or *pdftotext* must exist on your server. The type-command associations are defined in the [Indexing tab](#) of the [Admin: File Gallery](#) page.

System command

odt2txt %1

odt2txt %1

odt2txt %1

Ubuntu/Debian package with command

odt2txt

odt2txt

odt2txt

docx2txt.pl %1 -

xls2csv %1

catppt %1

catdoc %1

or

strings %1

pstotext %1

or

pdftotext %1 -

pstotext %1

pstotext %1

catdoc

catdoc

catdoc

poppler-utils or pstotext

pstotext

pstotext

catdoc %1

col -b %1

or

strings %1

xls2csv %1

catppt %1

xls2csv %1

pstotext %1

man -l %1

catdoc

bsdmainutils

catdoc

catdoc

catdoc

poppler-utils or pstotext

man-db

col -b %1

or

strings %1

elinks -dump -no-home %1

col -b %1

or

strings %1

col -b %1

or

strings %1

bsdmainutils

elinks

bsdmainutils

bsdmainutils

col -b %1

or

strings %1

col -b %1

or

strings %1

bsdmainutils

bsdmainutils

Several tools can be used to extract search strings; many Unix sites have "strings", which can detect things which appear to be text within files although without the accuracy of more specialized tools.

Ensure that the system command entered prints its output to the screen (standard output) and not to a file. Try the command on a console and check the manual. E.g. you have to add a trailing "-" to pdftotext.

It might be needed to clear the Tiki [Cache](#) after installing a new handler for the system to pick it up.

It's better if you have [fileinfo](#) installed to avoid misidentified mimetypes (install php-pear if you are using php < 5.3).

To install all required packages in a Debian-based server, you can use this command:

```
sudo apt-get install bsdmainutils catdoc elinks man-db odt2txt php-pear pstotext
```

If you use WikiSuite, everything is pre-installed.

Related:

- <http://stosberg.net/odt2txt/>