

MediaWiki Importer

To use, point your browser to `tiki-importer.php`

Attention: the importer can handle the conversion of most of the MediaWiki syntax to Tiki syntax but not all. If you are planning to use it, it is very likely that you will have to handle some corner cases manually.

Since Tiki4, you can import content from MediaWiki installations to Tiki. The importer is based on the PEAR package Text_Wiki that support a different wiki syntaxes. So you can use the same code base to add support to different softwares (like DokuWiki or others).

Basically you need to generate a MediaWiki XML file and upload this file through the importer interface. Please read carefully all the documentation before using the importer to understand its features and limitations. Also look the comparative table between MediaWiki and Tiki syntax for a list of supported and not supported syntax.

Warning

Before trying importer, it is highly recommended that you perform a full backup of your Tiki installation.

Requirements

- DOMDocument (should be enabled by default on recent PHP installations)

Supported versions

The script was initially made with Tiki 4.0 and MediaWiki 1.14, and has evolved ever since. You should use the latest stable version of Tiki to import data from a recent version of MediaWiki.

What can be imported?

- Wiki page content
- Wiki page history
- Wiki page attachments
- Wiki page categories (Tiki \geq 7.0 only). If the page contains `[[Category:xxx]]`, a category XXX is created if needed and the page is assigned to this category. The link is taken away from the page source. Only the last revision is assigned to the category.
- MediaWiki users (see more information below)

Importing a MediaWiki site step-by-step

1.1.1. The MediaWiki XML file

There are many ways to generate the MediaWiki XML file but only two have been tested with this importer (although it is very likely that all ways work).

The easiest one is to use the MediaWiki Special:Export page. Beware that using the special page has two significant limitations: there is no easy way to export all pages (the easiest one is to copy and paste the list of pages from Special:AllPages) and it is not possible to include a reference to the wiki page attachments (images or files).

Due to those two limitations with Special:Export page, the recommended way to generate the XML file is by using the PHP script dumpBackup.php. This script is included in any MediaWiki installation since version 1.6 and is located in the directory maintenance/. With this script is possible to export all the revisions of each wiki page or just the last revision.

Also you can include a reference for the wiki page attachments. Its official documentation is [here](#).

For information on how to generate the XML file see the official MediaWiki documentation.

Tips: test always your xml is well formatted before importing it - otherwise you will have a seg fault. Open the xml file with firefox for instance. It will tell you if error.

1.1.2. Using the importer interface

Some users have reported that MediaWiki sometimes generate invalid XML files. So first of all check if you have created a valid XML by opening it on Firefox (or another program that can open XML files). If your file has XML errors Firefox will tell in which lines, whereas Tiki importer will just generate an error without further information. After verifying that you have a valid XML file from your MediaWiki site, go to your Tiki installation and under the Admin menu click on the option "Tiki Importer". Read carefully the messages and select the software from which you are going to import (MediaWiki is the only option until Tiki7, where Wordpress is also offered).

Tiki importer

Warning:

If you are NOT running a new Tiki installation, make a backup of your database before using this importer!

Note:

Depending on the size of the file from the source software, the import process may take a while to complete. This might be a problem according to your PHP and web server settings. This script tries to change the relevant settings but there are some settings that the script cannot change. So, if you are having problems with the script, please try to increase the value of the following PHP settings: `max_input_time`, `max_execution_time` (this setting is limited by the web server setting, if you are running Apache also change its Timeout setting), `post_max_size`, `upload_max_filesize`, `memory_limit`. It is recommended that you run this script on a server where you can change the values of those settings (if needed).

Please choose the software to import from:

 

On the next screen you have three different options to configure the importer behavior:

- *Import images and attachments*: mark this option if you want to import wiki page attachments. Note that your XML file will only have information about the wiki page attachments if you have created it using the `dumpBackup.php` script with the option --

uploads. Also in the XML file there is only the URL for the attachment, the importer will try to download the file from this URL so make sure that you have access to the source URL from the computer where the importer is running. Note that importing attachments will NOT work if you generate the XML file using MediaWiki interface as the generated file has no information about the attachments. If for any reason you are not able to generate the XML file using `dumpBackup.php` you can check this workaround for importing attachments.

- *Number of page revisions to import*: the default one meaning the last page revision. You can define any number of page revisions to import (use zero to import all page revisions).
- *What to do with page names that already exists in Tiki*: this option only make sense if you are importing MediaWiki data to a pre-existing Tiki installation. In this case you can decide what to do in the case of a page name collision. The options are to not import the conflicting pages, to override the pre-existing pages or to append the prefix "MediaWiki" to the conflicting page names that are going to be imported.

Below the options you will find the upload form to select your XML file from your computer. Click on it, select the file and click on the import button to start the importation.

Import options:

☒ Import images and attachments (see documentation for more information)

Number of page revisions to import (0 for all revisions):

What to do with page names that already exists in TikiWiki?



While the script runs it will output information about its execution. After it finishes you have to click on the link located on the bottom of the page. The last screen will show a summary of the importation. If there are a error box pay attention to the error messages, they might give you some tips of what went wrong.

Page Draw sucessfully imported
Page Talk:Words beginning with 'D' sucessfully imported
Page Writing pad:Searching sucessfully imported
Page User:Maziar sucessfully imported
Page Holding Pattern sucessfully imported
Page Holding sucessfully imported

Importation completed! [Click here](#) to finish the import process

1.1.3. Handling MediaWiki users

Unfortunately, MediaWiki XML file does not has enough information about the users to allow the importer to automatically create new Tiki users. It only has the username and to be able to properly create the user we also need the user e-mail or password. So you have to separately import the users.

To make this easy we created a MediaWiki extension to generate a CSV file with the username and email of all the MediaWiki users. For more information about how to install it and create the CSV file see the extension page. Once you have created the file you need to edit it and add a new column called password between the columns

username and e-mail. Fill this column with random created passwords for each user.

Then use the Tiki built-in feature to batch import users. This feature has an option to force the new users to change their password at first login.

1.1.4. Video tutorials

- Exporting MediaWiki users and importing them to Tiki
- Creating a MediaWiki XML file
- Importing MediaWiki content to Tiki using MediaWiki XML file

Known issues

The following is a list of known issues with the current version of the importer. If you have technical knowledge please feel free to help.

- The first letter of file attachments on MediaWiki is case insensitive and on Tiki it is case sensitive.
- Wiki page name on MediaWiki is case sensitive and on Tiki is case insensitive, so two different pages for MediaWiki can be the same page for Tiki. So if in a MediaWiki installation there are a page called "Test" and other called "TeSt", only one of those pages will be imported.
- MediaWiki syntax does not differentiate between images and files so the importer consider them all to be images. This means that both images and files in wiki pages will be imported but only images will be properly rendered on Tiki pages.
- In some cases there might be missing or unexpected new lines before or after headings after parsing the syntax from MediaWiki to Tiki. It is not clear yet the reason. For more information see the [Text_Wiki](#) bug report about this issue.
- The importer use Tiki's built-in functions `create_page()` and `update_page()` to insert the data. The `update_page()` method has serious performance issues when called multiple times (which happen if you are going to import more than one revision per page). To improve the performance of the importer you can change the two lines of this method. See the example below (the commented lines are the original ones):

```
//$bytes = diff2($data , $edit_data, 'bytes'); $bytes = 0; //$diff = diff2($old["data"] , $edit_data, "unidiff"); $diff = '';
```

- The importer uses the PHP class DOMDocument to handle the MediaWiki XML file. Apparently some Linux distributions (like Fedora) compile PHP with the option --disable-dom. If this is your case you will see the following error when trying to use the importer "Fatal error: Class 'DOMDocument' not found in lib/importer/tikiimporter_wiki_mediawiki.php on line 69". To solve this you need to follow your distribution specific instruction to enable PHP DOM. For more info see this bug report. A new version of the importer should check if the DOMDocument class exists and if not produce a more user friendly message.
- Single quotes inside links are valid MediaWiki syntax but they are not support by the current version of Text_Wiki. See this bug report for more details.
- html tags in the text are not very well handled. If your wiki pages have html code, you will have to allow html in the wiki page. If you have html instruction in your table, it will probably break the import. Better to clean before
 - a | in a table cell is not escaped
 - a link like [[aa*aa|aa]] is not rendered properly
 - a line with a space creates a pre box
 - an image name with comma is not excaped
 - a return in a table cell is not repalce with
 - inappropriate link are done on text like common.1.2.jar
 - a

- !title does not clean the p tags
- definition list used to indent
- [[https://...]] is transformed into https://..." class="wiki wikinew text-danger tips">https://...
- [[xxx(y)]] is translated correctly (due to a parsing bug for xxx(y). It should be translated to xxx(y) until the parsing is fixed

TODO:

- parse the html tags
... that MediaWiki accepts into tiki syntax. For now , the page needs to be html, that gives a lot of problems

Support

- MediaWiki Importer support forum

Related

- [MediaWiki importer development page](#) (page used to document the progress of the project during the Google Summer of Code 2009)
- [Tiki to MediaWiki importer](#)
- [Import-Export](#)
- [Dictionary](#)

Modifications in the lib/pear/Text library

This library is used in the importer but is no more maintained

List of commit that must be integrated if

- 39250: redirect param name
- 39263: prefs for alternative syntax to simplebox and centered

alias

- MediaWiki to TikiWiki converter
- MediaWiki
- Mediawiki to Tiki importer
- MediaWiki to TikiWiki