

## TikiWikifier

### Where did this feature go?

"On Admin->Wiki enable the option Allow HTML. It then appears on the tiki-editpage. From what I can find, its disabled with the Allow HTML to protect possible malicious HTML code from being imported. See the recent security work in 1.7 and 1.8 releases. - Damian

The feature is also missing when using moreneat.css and neat.css; this will be resolved in the next releases -Terence"

## Overview

As you may have guessed from the title 😊 **Tiki Wikifier** is a feature to convert remote files (is there a need to convert local files?) from HTML into wiki pages.

### What Has Been Implemented

- Look at the wiki edit form for 'Import HTML' field. Type or copy paste the URL (with http:// prefix) of the page you want to import.
- Set 'Try to convert HTML to wiki' as needed
  - If set **Tiki Wikifier** will parse HTML and generate wiki page with wiki syntaxes of corresponding HTML tags
  - If blank **Tiki Wikifier** will append pure HTML text to the end of current page
- Recognized wiki syntaxes
  1. bold
  2. italic
  3. monospaced
  4. underline
  5. centered text
  6. headers
  7. colored text (by HTML `<font>` tag)
  8. links
  9. images (including linked images)
  10. page title (HTML `<title>` tag) become `<H1>` header
  11. lists: ordered/unordered, mixed lists, depth
  12. tables (new style wiki tables generated)

### TODO and Future Plans

- Remove (or escape) wiki syntaxes in source text... i.e. need to change text in square brackets for example to avoid being interpreted as link by wiki parser.
- Add hostname to links w/o http prefix...
- This is rapid (sometimes ugly) implementation of wiki HTML import. I think it should be pluggable import/export converters with ability to choose from the edit form which converter to use for any given operation.
  - In case of import converter, it can try to guess what source file is (using mime type from remote server response).
- Of course converters may have its own configuration panel what should be plugged into wiki page edit form too... for example: HTML importer may have flags 'strip HTML tags' and 'try to convert HTML to wiki' 😊
  - At least one export filter for wiki already exists 😊— PDF exporter...
    - converter for swf in the future... could be very nice for slide shows...
    - converter for svg would also be very cool
    - converter for xforms
- Colored text also maybe done via `<span style="color: #XXXXXX">`
- Need to code converter HTML specials (like &nbsp; and so on) into wiki
  - it could be one `preg_replace()` with arrays of search/replace 😊

- Need to verify/fix HTML parser grammar file! As I see it, it lacks many tags/attributes! (so runtime PHP errors may occur on import) — **Somebody with knowledge of the HTML docs/specs can do this? Any volunteers?** 😊
- Maybe it is reasonable to introduce 'strict' parsing mode? (to reduce noise amount on converted pages)
  - i.e. remove all empty lines after parse (convert)
  - join lines during parse process
  - line breaks on output wiki page may occur only on explicit `<br>`, `<p>` or after some directives (like headers and list items)
  - should it check and fix (as possible) for wiki syntax errors after generating a wiki page? like multilined **bold/italic/etc.** (if somebody doesn't know it is one-liners)
- Good designed converter should use **delayed generation** of wiki constructs
  - complex pages like tables ~~~~which is usually used for page layout and not for real content~~~~ it can be detected almost at easy 😊— by absent content in most of cells for example, or by present another layout constructs inside a cell (like divs or other tables)
  - but unfortunately such things will extremely increase complexity of FiniteStateMachine (of cause good designed converter should be FSM based). However there is no guarantee that using big enough FSM can make parsing process **really very much better** in comparison to stupid (very simple) algorithm (actualy depend on FSM details). So, I'm in doubt should it be rewritten and when? — Looks like it will be an evolutionary process... 😊