Assist Converting Documents into MediaWiki Format

Contents

1 Assist Converting Documents into MediaWiki Format

This guide is intended to show the processes that can be followed to convert a document (DOCX by preference) into an Assist page.

There are multiple methods and processes, including some not documented here, such as using LibreOffice or a Word extension, neither having produced reasonable results or time savings.

The processes below will automate vast tracts of the process and make the conversion of documents much easier and less time consuming.

1.1 Tools

Recommended initial editing tool - NOTEPAD++

- https://notepad-plus-plus.org/downloads/

Create Macro - Assist Cleanup

- See documentation - Assist Cleanup Macro

Create Macro - Vertopal Cleanup

- See documentation - Assist Clean Up Vertopal Macro

Recommended - PowerToys

- https://learn.microsoft.com/en-us/windows/powertoys/install

AI

- Aurora/Direct ChatGPT
- Gemini

Online Conversion tool:

- https://www.vertopal.com/en/convert/docx-to-wiki-mediawiki

1.2 Process using On-line Converter

Convert document using Vertopal online conversion tool

- https://www.vertopal.com/en/convert/docx-to-wiki-mediawiki

Use DOC/DOCx - PDF doesn't work very well.

NOTE: This does not convert embedded objects like excel, attachments, Visio or Word diagrams. If your document contains these, then save these as PNG images to add to the document later.

Download and Extract to subfolder

Rename and Edit wiki file

Run Macro - Assist Cleanup

Run Macro - Clean Up Vertopal

Essentially this does the following:

```
<span.*</span> with nothing

\[\[File\:vertopal_392608dac46847cb99daf0bb8d5090ed/media/image with \[\[File\:TTM-

\.png\|.*?\]\] with \.png\|800px\]\] (if mostly huge images) or \.png\]\] (if mostly small images)
```

Ready for **What's Next, Now**™

```
<\/?blockquote> with nothing

<ol.*>\n<li> with \n

</li></ol> with nothing

= <br />\n with "= " (without the quotes)

\{\| with \{\| class="wikitable" border="1"

!width="\d*%"\| with !
```

💡 **Note:** Replace in the above - this assumes replacing for a document for TTM - replace TTM- with the name of your Assist and doc being converted e.g. WMS-, PORTAL-TTM-ARCH-, EPOD-DEVICE-, CTLTMS-, etc

The Vertopal Cleanup macro will put in "SYS-DESC-", so replace that.

Go through and ensure images are named with a reasonable name

e.g. if in home section in TTM, change file from TTM-1.png to TTM-HOME-1.PNG

Check all images now have a name - search for File\:

Any inline icons, make size 16px - search for file\: and any that are within paragraphs will need manipulation

Put into Assist

Rename images - Recommend using PowerRename from PowerToys.

- Select all images to be renamed, then right-click and Select Rename with PowerRename.
- If from Vertopal, the images are in a media subdirectory, names as image1, image2, etc.
- Tick "User regular expressions"
- From name: image(.*)
- To name: SYS-DESC-$1

Example:

- From name: image(.*)
- To name: TTM-WMS-$1

Results

From	To
Image1	TTM-WMS-1
Image2	TTM-WMS-2
Image10	TTM-WMS-10
Image11	TTM-WMS-11
Image20a-text	TTM-WMS-20a-text

💡 **Note:** If there is any media in funky formats (like WMF/WMV. Visio or other objects directly embedded), then these will need to be manually converted - use screenshots, paint, etc to achieve that. This is not covered in this guide.

When renamed, use the Upload Multiple Files Special page to upload the files in bulk:

- Go to Special pages
- Click Upload multiple files.
- Enter a description for the images (otherwise this will default to something useless).
- Drag your images to be uploaded to the appropriate place on the page.
- The images will upload automatically, showing the progress on the screen.
- If any images should fail, try re-uploading them - sometimes bulk uploading >40 images at a time will cause some issues.

If you are doing this manually instead, upload the images by clicking on the image in the document and then manually upload the correct "imageX" file using the numbers as guide.

Finally, check the sizes of all pictures - remove or add |800px if necessary.

Ready for **What's Next, Now**™

1.3 Process using AI/Manual

This process does not use an on-line converter process, but instead leverages AI to do a lot of the conversion work for you (after you have trained it appropriately). This is slightly more long-winded than using an online converter, but is much quicker than typing, copying and pasting manually.

Steps:

Convert the text to MediaWiki format

- Open the document
- Get rid of titles and final appendices - these will be added later using templates.
- Break the text down into chunks - it's easier to work with smaller segments
- Ask your favoured AI to convert the text to MediaWiki format.
    - ♦ 💡 **Note:** These steps are proven working for the AI's listed.
    - ♦ E.g. Gemini process including prompts:
        - ◊ *Please convert this text to Mediawiki format.*
        - ◊ Paste in text
        - ◊ *Please remove the table of contents and the first page before that.*
        - ◊ *Please lose the DIV elements*
        - ◊ *The first heading is level 1, so please reflect that and reduce all other headings by 1 level*
        - ◊ *Please remove the numbering from the headings.*
        - ◊ *Please remove any smart quotes or dashes and replace with plain quotes and hyphens.*
        - ◊ *The text I pasted in had images. Can you identify where those images were and put a placeholder in there of [[File:SYS-DESC.png|800px]]*
        - ◊ *Can you make the placeholder count please? e.g. SYS-DESC-1.png, SYS-DESC-2.png, etc*
        - ◊ *Any A.# headings should be level 2. Same for B.#, C.#.*
    - ♦ E.g. Aurora/ChatGPT including prompts:
        - ◊ *Please convert my text into mediawiki format*
        - ◊ Paste in text
        - ◊ *Please make sure there are two line breaks between each section*
        - ◊ *The first heading is level 1, so please reflect that and reduce all other headings by 1 level*
        - ◊ *Please remove any smart quotes or dashes and replace with plain quotes and hyphens.*
        - ◊ *The text I pasted in had images. Can you identify where those images were and put a placeholder in there of [[File:SYS-DESC.png|800px]]*
        - ◊ *Can you make the placeholder count please? e.g. SYS-DESC-1.png, SYS-DESC-2.png, etc*
    - ♦ More prompts may be required on your document to get the format right, such as numbered lists, tables etc. In my tests, these all converted well, but you may have other preferences. For example:
        - ◊ *Please add "apt-searchable" as a class to the wikitables*
        - ◊ *Please add width="100%" to the tables*
        - ◊ *Please add border="1" to the tables.*
    - ♦ Depending on your document, you may want to make the images start at a particular number. You can ask the AI, for example:
        - ◊ *Please start the image numbering from the last section*
        - ◊ *Please reset image placeholder numbering to 1*
        - ◊ *Please start image placeholder numbering at 3*
    - ♦ The AI may be limited to the amount of text that can be uploaded, or the amount of text that can be output. In the latter case, you may then be able to prompt the AI to give you the next section, piece by piece. E.g., for ChatGPT:
        - ◊ *Please now output the next section of the text, starting where you left off*
        - ◊ *Please continue*
    - ♦ The output may be formatted by the AI renderer (for example, numbered lists look bold and large, line breaks are missing. You can ask for the plain wikitext code as opposed to the rendered output.
        - ◊ *Please display the raw wikitext*
        - ◊ *Please put 3 backticks at the start of the output*
    - ♦ However, when you have trained your AI to give you the correct output, this should stick for further document conversions.
        - ◊ 🔴 **Warning**: You WILL lose all of your training is you close down the chat with your AI, so keep it open to preserve your requests for formatting.
- Create a new page in Assist
- Copy the converted text into your page, and continue until complete.

Extract Images/Media:

- Make a copy of your DOCX file and rename it to ZIP instead.

Ready for **What's Next, Now**™

- Open the zip
- Go to word/media
- Copy all of the image files from here into folder.

💡 **Note:** If there is any media in funky formats (like WMF/WMV. Visio or other objects directly embedded), then these will need to be manually converted - use screenshots, paint, etc to achieve that. This is not covered in this guide.

Rename Images:

Recommend using PowerRename from PowerToys.

- Select all images to be renamed, then right-click and Select Rename with PowerRename.
- If from Vertopal, the images are in a media subdirectory, names as image1, image2, etc.
- Tick "User regular expressions"
- From name: image(.*)
- To name: SYS-DESC-$1

Upload Images

Use Assist batch upload to upload the images:

- Go to Special Pages
- Click Upload multiple files.
- Enter a description for the images (otherwise this will default to something useless).
- Drag your images to be uploaded to the appropriate place on the page.
- The images will upload automatically, showing the progress on the screen.
- If any images should fail, try re-uploading them - sometimes bulk uploading >40 images at a time will cause some issues.

Manual insertion of Images:

If you have not trained the AI to insert image placeholders as suggested above, you may need to manually insert images in the correct place in the document.

- You can use the VisualEditor copy and paste or upload.
- You can instead batch upload the images first.
- You can paste in [[File:SYS-DESC-1.png||800px]] for the first image, and then update as you go along.

Finally, check the sizes of all pictures - remove or add |800px if necessary.

Ready for **What's Next, Now**™