# PyWikibot Update 241006

a new stable version '''9.4.0''' of Pywikibot was published. Among others the changes include the following:  
- Scripts for Pywikibot can now also be provided and installed via Python Package Index] (PyPI) . Instructions for this can be found at <https://doc.wikimedia.org/pywikibot/stable/entrypoint.html>.  
\* The supported Wikimedia projects have been expanded.  
\* Further changes are described at <https://doc.wikimedia.org/pywikibot/stable/changelog.html#current-release-changes>  
  
With version '''9.4.0''' and after a longer period of development all Pywikibot scripts can now be installed via the Python Package Index see <https://pypi.org/project/pywikibot-scripts/>. The command  
  
   pip install pywikibot-scripts  
  
installs the scripts and the associated Pywikibot framework including all required external packages as a site package. The script must be called via the pwb wrapper, see <https://doc.wikimedia.org/pywikibot/stable/utilities/install.html#module-pwb> e.g.  
  
  pwb touch -page:Wikipedia:Sandbox

WARNING: The script pwb.exe is installed in 'C:\Users\HP\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.12\_qbz5n2kfra8p0\LocalCache\local-packages\Python312\Scripts' which is not on PATH.

Consider adding this directory to PATH or, if you prefer to suppress this warning, use --no-warn-script-location.

(Search for advanced system settings, then under the advanced tab select environment variables.

# Error Handling

<https://www.w3schools.com/python/python_try_except.asp>

applied in summaries.py

# Colorama August 2024

https://pypi.org/project/colorama/

# Pywikibot Update August 2023

I had to uninstall and re-install because the pywikibot code has been updated. Also the MediaWiki software update to the web site changed the API access point!

Core folder remains the same

**pwb.py** did not come with the distribution: I had to retain the original one.

I had to generate a new family file:

This family file was auto-generated by generate\_family\_file.py script.

"""

Configuration parameters:

url = https://ehwa.mywikis.wiki/wiki/Main\_Page

name = ehwa

Please do not commit this to the Git repository!

"""

from pywikibot import family

class Family(family.Family): # noqa: D101

name = 'ehwa'

langs = {

'en': 'ehwa.mywikis.wiki',

}

def scriptpath(self, code):

return {

'en': '/w139',

}[code]

def protocol(self, code):

return {

'en': 'https',

}[code]

It gets generated in the **\families** folder, but it has to be moved to the folder **pywikibot\families**

I had to generate new user configuration file:

family = 'ehwa'

mylang = 'en'

usernames['ehwa']['en'] = 'jamest'

password\_file = ''

Then I was able to login successfully.

Now I am looking for a way to run a Python script to access a file text and upload a new file.

See **crosslink.py** before adding all the editing code:

''' crosslink.py

Script to automatically create internal links

'''

import pywikibot

import re

site = pywikibot.Site('en')

page = pywikibot.Page(site, 'Person:Frederick\_Allsop')

text = page.get()

print(text)

newtext = re.sub("Kalgoorlie","[https://en.wikipedia.org/wiki/Kalgoorlie Kalgoorlie]", text)

print("\n\n",newtext)

page.put(newtext,summary='This is a demonstration change',force=True, asynchronous=True)

Remaining work on the crosslinks project:

1. A version of crosslinks2 that merely generates a list of pages on the site, with the modified page names, the list goes to a file.
2. A version of crosslinks2 with the editing modules that works through a list of pages to be edited (page name, standardized name), gets the page text, edits, and puts it back.

# Numerical Methods

Numpy.org

DearPyGUI – fast GUI

https://github.com/hoffstadt/DearPyGui

VPython – 3D graphics

# Tree Notes

<https://www.geeksforgeeks.org/tree-traversal-techniques-in-python/>

<https://www.pythonforbeginners.com/data-structures/tree-data-structure-in-python>

<https://stackoverflow.com/questions/2358045/how-can-i-implement-a-tree-in-python>

# Panda Installation Notes

pip install pandas

WARNING: The script f2py.exe is installed in 'C:\Users\HP\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.12\_qbz5n2kfra8p0\LocalCache\local-packages\Python312\Scripts' which is not on PATH.

Consider adding this directory to PATH or, if you prefer to suppress this warning, use --no-warn-script-location.

pip install openpyxl

# PyWikibot Installation Notes

James Trevelyan, March 2024

I am using Windows 11.

Follow instructions here: <https://www.mediawiki.org/wiki/Manual:Pywikibot/Installation>

You need to know where to unzip the stable pywikibot installation which I only found by using the recommended command “python -m pip install pywikibot” and then searching for the “pywikibot” folder on my “C” drive. My installs have all been to a folder specific to my user files since the installer, apparently, could not access the normal system install folder. I did this command to uninstall the “pip” version: “python -m pip uninstall pywikibot”.

Then I unzipped the “stable distribution” referred to above into a folder, and copied everything to the “core folder”, which is the folder above the “pywikibot” folder.

There is a useful guide here: <https://doc.wikimedia.org/pywikibot/stable/introduction.html>, but you need to configure for a non-standard wiki:

<https://www.mediawiki.org/wiki/Manual:Pywikibot/Third-party_Wiki_Quick_Start>

“Core” folder

C:\Users\<username>\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.12\_qbz5n2kfra8p0\LocalCache\local-packages\Python312\site-packages

In my case, the username is HP

Core folder:

C:\Users\HP\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.12\_qbz5n2kfra8p0\LocalCache\local-packages\Python312\site-packages

My configuration files, in the core folder

## user-config.py:

family = 'ehwa'

mylang = 'en'

usernames['ehwa']['en'] = 'jamest'

password\_file = 'user-password.py'

## user-password.py

(jamest,\*\*\*\*\*\*\*\*\*\*\*\*)

# password obscured for security reasons, put the actual password into the file.

Families folder (./pywikibot/families/)

C:\Users\<username>\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.12\_qbz5n2kfra8p0\LocalCache\local-packages\Python312\site-packages\pywikibot\families

## ehwa\_family.py

"""Family module for ehwa wiki."""

#

from pywikibot import family

# ehwa family module

class Family(family.Family):

name = 'ehwa' # Set the family name; this should be the same as in the filename.

langs = {

'en': 'ehwa.mywikis.wiki', # Put the hostname here.

}

# Translation used on all wikis for the different namespaces.

# Most namespaces are inherited from family.Family.

# Check the family.py file (in main directory) to see the standard

# namespace translations for each known language.

# You only need to enter translations that differ from the default.

#namespaces[4] = {

# '\_default': 'Engineering Heritage Western Australia', # Specify the project namespace here.

#}

#namespaces[5] = {

# '\_default': 'Engineering Heritage Western Australia talk', # Specify the talk page of the project namespace here.

#}

def version(self, code):

return "1.35.13" # The MediaWiki version used.

def scriptpath(self, code):

return '/w'

# The relative path of index.php, api.php :

# look at your wiki address.

# This line may need to be changed to /wiki or /w,

# depending on the folder where your mediawiki program is located.

# Note: Do not \_include\_ index.php, etc.

I found that the scriptpath return had to be ‘/w’ by trial and error. I found that when I first tried, “namespaces” was an undefined symbol so I commented out the lines assigning values to it.

## Getting Started

Once done, I found I had to work in the “core” folder.

I managed to login: >python pwb.py login

I managed to do this: >python pwb.py upload -help

The help file appears below.

Then I was able to do these commands manually and in a batch file.

python pwb.py upload C:\D\2024\240315\_EHWA\ehwa\_downloads\lettuce.jpg -ignorewarn -noverify -keep -descfile:C:\D\2024\240315\_EHWA\ehwa\_downloads\lettuce\_desc.txt

python pwb.py upload C:\D\2024\240315\_EHWA\ehwa\_downloads\lettuce1.jpg -ignorewarn -noverify -keep -descfile:C:\D\2024\240315\_EHWA\ehwa\_downloads\lettuce\_desc.txt

Upload help text

Script to upload images to Wikipedia.

The following parameters are supported:

-keep Keep the filename as is

-filename: Target filename without the namespace prefix

-prefix: Add specified prefix to every filename.

-noverify Do not ask for verification of the upload description if one

is given

-abortonwarn: Abort upload on the specified warning type. If no warning type

is specified, aborts on any warning.

-ignorewarn: Ignores specified upload warnings. If no warning type is

specified, ignores all warnings. Use with caution

-chunked: Upload the file in chunks (more overhead, but restartable). If

no value is specified the chunk size is 1 MiB. The value must

be a number which can be preceded by a suffix. The units are:

No suffix: Bytes

'k': Kilobytes (1000 B)

'M': Megabytes (1000000 B)

'Ki': Kibibytes (1024 B)

'Mi': Mebibytes (1024x1024 B)

The suffixes are case insensitive.

-async Make potentially large file operations asynchronous on the

server side when possible.

-always Don't ask the user anything. This will imply -keep and

-noverify and require that either -abortonwarn or -ignorewarn

is defined for all. It will also require a valid file name and

description. It'll only overwrite files if -ignorewarn includes

the 'exists' warning.

-recursive When the filename is a directory it also uploads the files from

the subdirectories.

-summary: Pick a custom edit summary for the bot.

-descfile: Specify a filename where the description is stored

It is possible to combine -abortonwarn and -ignorewarn so that if the specific

warning is given it won't apply the general one but more specific one. So if it

should ignore specific warnings and abort on the rest it's possible by defining

no warning for -abortonwarn and the specific warnings for -ignorewarn. The

order does not matter. If both are unspecific or a warning is specified by

both, it'll prefer aborting.

If any other arguments are given, the first is either URL, filename or

directory to upload, and the rest is a proposed description to go with the

upload. If none of these are given, the user is asked for the directory, file

or URL to upload. The bot will then upload the image to the wiki.

The script will ask for the location of an image(s), if not given as a

parameter, and for a description.

GLOBAL OPTIONS

==============

For global options use -help:global or run pwb.py -help

I also managed to remove the category accidentally added to all the imported WA pages EHWA\_Main.

python pwb.py category remove EHWA\_Main

18-May

Bro Code <https://www.youtube.com/watch?v=lyoyTlltFVU> – Tk in Python

Also PyCharm development environment

<https://stackoverflow.com/> - Q&A site very valuable!

<https://iqss.github.io/dss-webscrape/index.html> - filling in web forms with Python

Early code which worked….

os.chdir('C:/Users/HP/OneDrive - Close Comfort Pty Ltd/Documents')

print(os.getcwd())

with open('eh.xml', 'rb') as file:

content = file.read()

import xml.etree.ElementTree as ET

root = ET.fromstring(content.decode('utf-8'))

python

import re

import xml.etree.ElementTree as ET

# Regular expression patterns

page\_name\_pattern = r'<title>(.\*?)</title>'

image\_file\_pattern = r'(?<=File:)(.+?)(?=\|)'

# Function to extract page names

def extract\_page\_names(text):

return re.findall(page\_name\_pattern, text)

# Function to extract image files

def extract\_image\_files(text):

return re.findall(image\_file\_pattern, text)

# Parse the XML file

tree = ET.parse('eh.xml')

root = tree.getroot()

# Iterate through the XML elements

page\_names = []

image\_files = []

for page in root.findall('.//page'):

title = page.find('title').text

print(title)

revision = page.find('.//revision')

if revision is not None:

text = revision.find('text').text or ''

page\_names.extend(extract\_page\_names(title))

image\_files.extend(extract\_image\_files(text))

# Print the extracted data

print("Page Names:")

for name in set(page\_names):

print(name)

print("\nImage Files:")

for file in set(image\_files):

print(file)