Running basic scripts via Pywikibot

Run as part of the "Small wiki toolkits" project

WIKIMEDIA
FOUNDATION

Trainer notes
* Facilitate participant introductions in the chat. Add all usernames to etherpad.
* Introduce the speakers
* Introduce the Small wiki toolkits initiative, briefly highlight its purpose, past efforts coordinated as part of it, etc.
* Wikimedia Hackathon 2022 and Small wiki toolkits workshop
* Introduce the workshop topic
* Share housekeeping tips and start the recording.
** Keep your video turned off and audio muted when you are not talking to save yourself some bandwidth.
** Remember not to say things on video, audio, or chat that you wish would not have been recorded.
** Friendly reminder that the Wikimedia friendly space policy applies to this event: https://www.mediawiki.org/wiki/Code_of_Conduct No tolerance for harassment.
** Use the chat to ask questions.

Upcoming SWT workshops on bots development

➔ How to write a basic script via Pywikibot

➔ How to run a popular bot on your wiki

➔ How to host a bot on Toolforge

➔ How to maintain bots

➔ Intro to Wikimedia Tools (PAWS, Quarry, WDQS) for running queries and scripts

➔ How to interact with Wikidata via Pywikibot

WIKIMEDIA
F O U N D A T I O N

Trainer notes
* Share the previous workshop done on the introduction to Pywikibot framework and what was covered
* Highlight the several other workshops under the bots umbrella planned as part of SWT and the scope of those topics are not part of this very workshop: https://meta.wikimedia.org/wiki/Small_wiki_toolkits/Workshops
* Share the plan to distribute certificates.

Agenda

→ Working with Pywikibot scripts
→ Hands-on activities
→ Running Pywikibot on PAWS

WIKIMEDIA
FOUNDATION

Trainer notes
* Introduce the agenda

Working with Pywikibot scripts

WIKIMEDIA
F O U N D A T I O N

Trainer notes

* For those who missed the previous workshop, follow instructions in here:
https://www.mediawiki.org/wiki/Manual:Pywikibot/Installation (on your computer)

Managing categories

Pywikibot framework

Using 'category' script

➜   To mass add a category to a list of pages

```
(venv)core:$ python3 pwb.py category add

Category to add (do not give namespace): Test

Links from which page should be processed? User:SSethi (WMF)/Sandbox1

Retrieving 1 pages from wikipedia:test.

…

Script terminated successfully.
```

Trainer notes

* Also run the user-config command

* Mass adds a category from all pages in a category.

* Use "test.wikipedia.org" for playing around with the scripts.

* Asks for a page to fetch links from.

* Quickly highlight the options available with this command

https://www.mediawiki.org/wiki/Manual:Pywikibot/category.py#Options_for_add_action

Pywikibot framework

Using 'category' script

➜ To remove a category from all pages in a category

```
(venv)core:$ python3 pwb.py category remove
Please enter the name of the category that should be removed: Test
Retrieving 50 pages from wikipedia:test.
Page [[Ba]] saved
Page [[User:Aa48965]] saved
Sleeping for 9.0 seconds, 2022-03-18 16:43:20
…
```

Trainer notes

* Removes a category from all pages in a category.

* Use it with caution.

* Use "test.wikipedia.org" for playing around with the scripts.

* Quickly highlight the options available with this command

https://www.mediawiki.org/wiki/Manual:Pywikibot/category.py#Options_for_remove_action

Pywikibot framework

Using 'category' script

➜   To make a list of all articles that are in a category

```
(venv)core:$ python3 pwb.py category listify
Please enter the name of the category to listify: Test
Please enter the name of the list to create: Test list
Sleeping for 6.5 seconds, 2022-03-18 18:13:37
Page [[Test list]] saved
```

Trainer notes

* Makes a list of all articles that are in a category

* Creates a page > https://test.wikipedia.org/wiki/Test_list

* Quickly highlight the options available with this command

https://www.mediawiki.org/wiki/Manual:Pywikibot/category.py#Options_for_listify_acti
on

* Also demonstrate the following command:

"python3 pwb.py category listify -overwrite -showimages" with "Hello World" as
category and list name as "Test list""

Pywikibot framework

Using 'category' script

➜   To move a list of articles in one category to another

```
(venv)core:$ python3 pwb.py category move -from:"Test-B" -to:"Test-G"


Retrieving 1 pages from wikipedia:test.

Sleeping for 7.3 seconds, 2022-03-18 18:45:01

..

Sleeping for 9.4 seconds, 2022-03-18 18:45:39

Page [[Category:Test-G]] saved
```

Trainer notes

* Move a list of articles in one category to another

https://test.wikipedia.org/w/index.php?title=Category:Test-G

* Talk about how different generators / filters options such as 'logevents' available and options for several actions such as 'summary' this:

https://www.mediawiki.org/wiki/Manual:Pywikibot/category.py#Generators_and_filters_available

Try-out time!

➔ **How do you plan to use these scripts?**

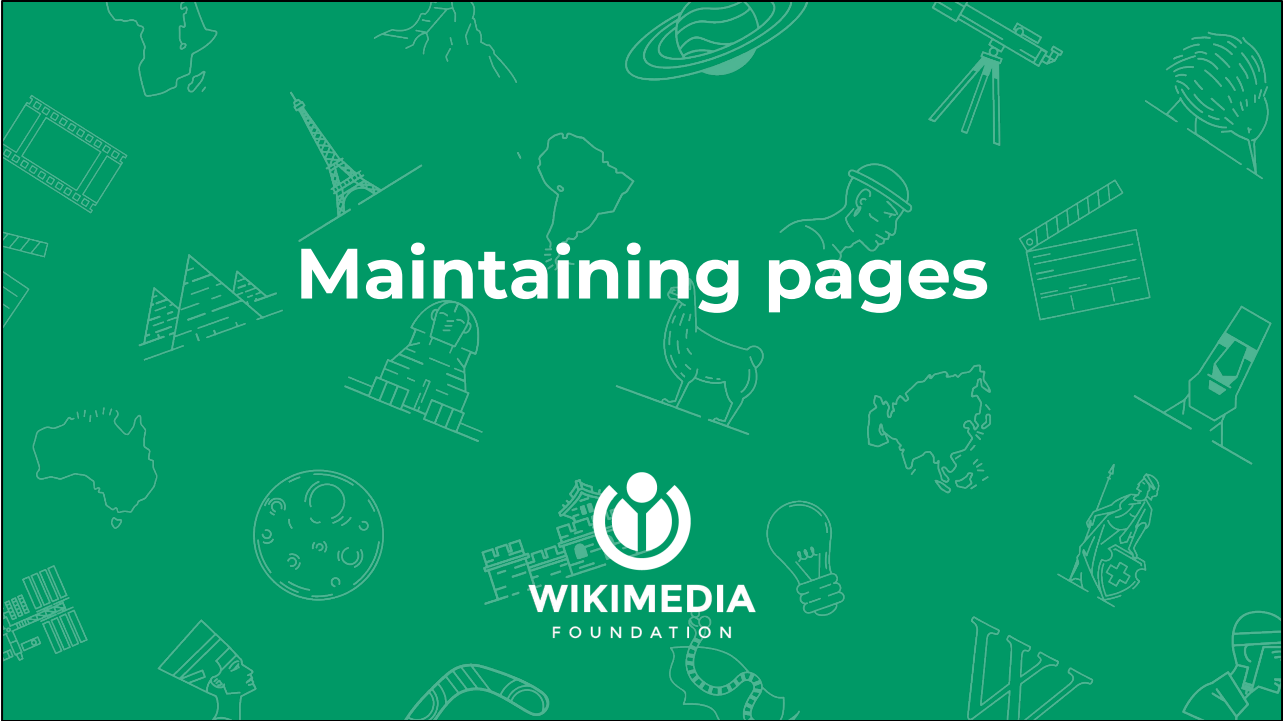➔ **Share the outcomes & errors in the chat as you try them out!**

WIKIMEDIA
F O U N D A T I O N

Maintaining pages

Pywikibot framework

Using 'clean_sandbox' script

➜ Cleans a sandbox with a predefined text

```
(venv)core:$ python3 pwb.py clean_sandbox -page:Test

Preparing to process sandbox page [[Test]]

Sleeping for 8.7 seconds, 2022-03-19 13:08:27

Page [[Test]] saved

@@ -1 +1,2 @@

…

Done.
```

Trainer notes

* Command is intended to work on https://test.wikipedia.org/wiki/Test

* There are so many optional parameters listed here:

https://www.mediawiki.org/wiki/Manual:Pywikibot/clean_sandbox.py

* You can also specify the -hours and -delay parameter specify after how long the script should repeat and the delay that should be there between first and the second run.

Pywikibot framework

Using 'weblinkchecker' script

➜ To find broken external links

```
(venv)core:$ python3 pwb.py weblinkchecker -page:Aeroflot

Retrieving 1 pages from wikipedia:test.

>>> Aeroflot <<<

*[[Aeroflot]] links to http://www.aeroflot.ru/eng/news.asp?ob_no=712&d_no=5609

- Not Found.

Waiting for remaining 5 threads to finish, please wait...

Saving history...
```

Trainer notes

* Reports dead link in a file in the code folder. Suppose to in a ".txt" file, also there is an option to report the dead links on the talk page.

* Several other options are available.

Pywikibot framework

Using 'delete' script

➜ Deletes a page

```
(venv)core:$ python3 pwb.py delete -page:"User:SSethi_(WMF)/Sandbox1"

Enter a reason for the deletion: No longer need

>>> User:SSethi (WMF)/Sandbox1 <<<

Can't delete [[test:User:SSethi (WMF)/Sandbox1]]; do you want to mark it for

deletion instead? ([y]es, [N]o, [a]ll): y

Page [[User:SSethi (WMF)/Sandbox1]] saved

Script terminated successfully.
```

Trainer notes

* https://test.wikipedia.org/wiki/User:SSethi_(WMF)/Sandbox1

* You can also restore pages

Try-out time!

➔ **How do you plan to use these scripts?**

➔ **Share the outcomes & errors in the chat as you try them out!**

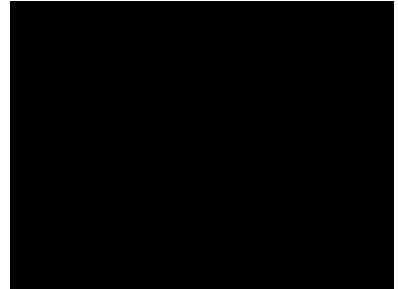# Running Pywikibot on PAWS

PAWS Overview

→ PAWS, a Jupyter notebook deployment hosted by Wikimedia

→ Allows creating and sharing documents that contain live code

→ Use PAWS to run Pywikibot scripts

→ Doesn't require installing any additional software or uploading files

WIKIMEDIA
FOUNDATION

Trainer notes

* Share with participants what PAWS is with help from the video and talk about its capabilities.

* Jupyter installation for Wikimedia

* Jupyter - concept that allows people to create programming notebooks that can run code in the browser.

* Quite popular in scientific community among educators - interactive computing.

* Data visualizations, analysis, running bots and so on.

* No need to require installation / software

* Why we are introducing PAWS now.

When to use PAWS

➔ Running scripts and bots for automated tasks
➔ For less intensive or non-periodic operations
➔ For exploration, not production
➔ Keeping notes on your work and sharing it with others
➔ Working with smaller wikis or datasets
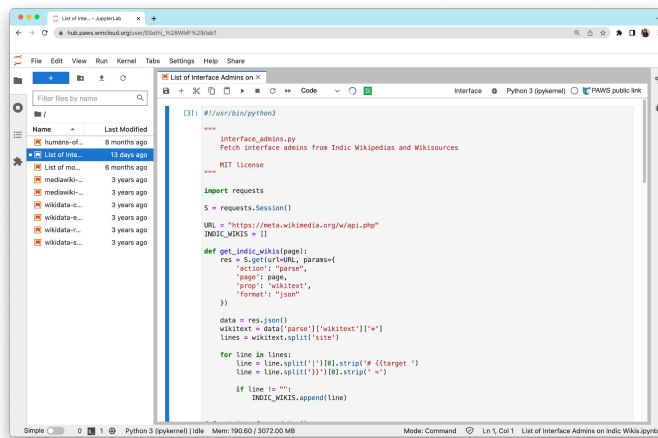➔ You don't need to collaborate with others on our code

WIKIMEDIA
F O U N D A T I O N

Trainer notes

* Talk about various scenarios when folks should / shouldn't use PAWS. Use information from the tutorial:
https://public.paws.wmcloud.org/User:SRodlund_(WMF)/Using-Pywikibot-with-Paws.ipynb

How to use Pywikibot in PAWS

WIKIMEDIA
FOUNDATION

Trainer notes

* Give a demo of the PAWS tool, show how one can access it, create a new Python 3 notebook, and run a short script that performs a simple task
https://public.paws.wmcloud.org/User:SRodlund_(WMF)/Using-Pywikibot-with-Paws.ipynb#Python3

* Go to https://hub.paws.wmcloud.org/hub/login. Show how login works.

*Copy paste the following code:

import pywikibot

site = pywikibot.Site('test', 'wikipedia')

page = pywikibot.Page(site, 'test')

page.text = 'Hello world!'

page.save('test edit')

page.text

* Show output and explain how run, pause, renaming, and public link works.

Logging into PAWS

https://hub.paws.wmcloud.org/hub

[Use your mediawiki account]

WIKIMEDIA
FOUNDATION

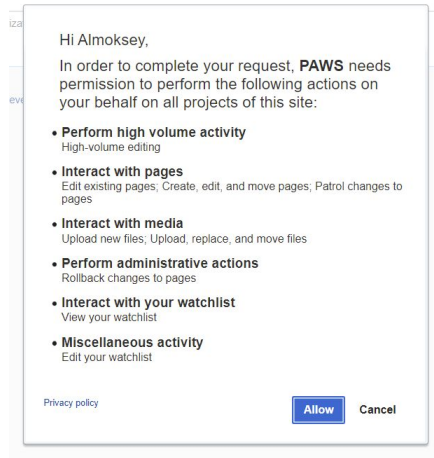**Trainer notes**

* The PAWS interface was updated 4 days ago. So if you've seen the previous one, you might notice some changes
* A notebook is running in a container and can be stopped and started at anytime. So it advisable not to run jobs that will last longer than an hour
* Use a chromium-based web browser - it supports copy-pasting
* Current Pywikibot version on PAWS is 6.6.5 so take not when following tutorials etc
* Unlike local installation, Pywikibot on PAWS is installed at /srv/paws. user-config.py etc are found here
* With Pywikibot 7. most deprecations will be dropped. Pywikibot 7 does not support Python 3.5.0 - 3.5.2 any longer
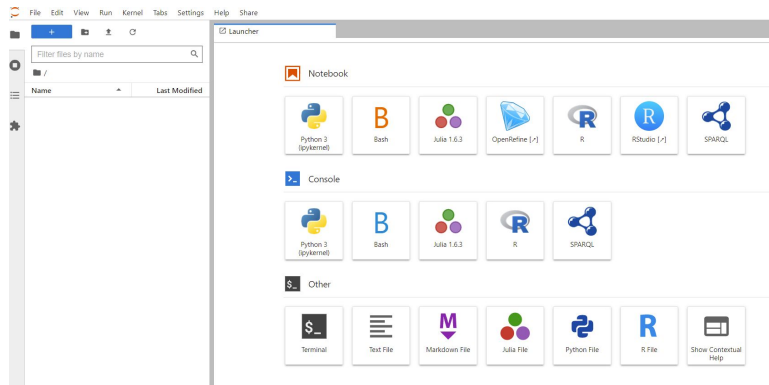
Working with PAWS

Accept PAWS permissions

Hi Almoksey,

In order to complete your request, **PAWS** needs permission to perform the following actions on your behalf on all projects of this site:

- **Perform high volume activity**
  High-volume editing
- **Interact with pages**
  Edit existing pages; Create, edit, and move pages; Patrol changes to pages
- **Interact with media**
  Upload new files; Upload, replace, and move files
- **Perform administrative actions**
  Rollback changes to pages
- **Interact with your watchlist**
  View your watchlist
- **Miscellaneous activity**
  Edit your watchlist

Privacy policy   Allow   Cancel

Trainer notes

* 6 necessary permission you need to interact with wikis from PAWS

PAWS Interface

Notebooks, Console, etc

Trainer notes

* PAWS already configures test.wikipedia.org as the default wiki. To change it, use generate_user_files or use -site:xxx:yy with your commands
* Select terminal

Pywikibot examples on PAWS

1. **pagefromfile**: create wiki page(s) from uploaded text file

2. **archivebot**: archives talk page discussions

3. **checkimages**: check recently uploaded files for description etc

4. **Python session**: invoking methods and classes

WIKIMEDIA
FOUNDATION

Running Scripts on PAWS

Using 'pagefromfile' script

➔ Create a set of wiki page(s) from uploaded text file

```
@PAWS: pwb.py pagefromfile -file:/home/paws/content.txt

Reading '/home/paws/content.txt'...

>> Pywikibot Workshop In March 2022 <<<

Page List of Pywikibot Workshops already exists, ***overwriting!

Edit summary: Bot: Automated import of articles *** existing text overwritten
***

Do you want to accept these changes? (      )es, [N]o, [a]ll, [q]uit): a
```

WIKIMEDIA
FOUNDATION

Trainer notes

* By default, the title of the content becomes the page name

* Existing pages are skipped

* Use -force to overwrite an existing page

Running Scripts on PAWS

Using 'pagefromfile' script

→ Content of uploaded text file

```
{{-start-}}

'''Pywikibot Workshop In March 2'''

This was held on the 25th of March, 2022

==List of Participants==

*Participant 1

….

Previous: [[Pywikibot Workshop In February | February, 2022]]

{{-stop-}}
```

WIKIMEDIA
FOUNDATION

Trainer notes

 * Multiple pages can be put in the uploaded file

Running Scripts on PAWS

Using 'archivebot' script

→   Archive a discussion page

```
@PAWS: pwb.py archivebot -page:"Talk:Pywikibot March Workshop"
"User:MiszaBot/config"
Processing 3 threads
Archiving 1 thread(s).
Sleeping for 9.0 seconds, 2022-03-24 23:58:16
Page [[Talk:Talk:Pywikibot March Workshop] saved
#Archives a talk page of a specific...
```

WIKIMEDIA
FOUNDATION

Trainer notes

* Archive a discussion/talk page

* For a section to be archived, it must be a level 2 section, marked with ==. This is
actually the thread or the topic

Running Scripts on PAWS

Using 'archivebot' script

→   Configuring a talk page with MiszaBot/config marker template

```
{{User:MiszaBot/config

|archive = Talk:Pywikibot March Workshop/Archive %(counter)d

|algo = old(10s)

|counter = 1

|minthreadsleft = 2

|minthreadstoarchive = 1

}}

#'archive' is the only required parameter here
```

WIKIMEDIA
FOUNDATION

Trainer notes

* Template must be at the very top. Must come before the first ==second level header==

* The marker template of "User: MiszaBot/config is necessary. It is what the bot looks for.

NOTE: Even though the doc says old(xx) takes seconds without the 's', this doesn't seem to work with 6.6.5, you must add the 's' suffix.

* The current code shows that it looks for the 's' suffix. It fails otherwise.

NOTE: To prevent a specific thread from being archived, insert an entry with a signature timestamp in a distant future. Use 'DoNotArchiveUntil' user …

NOTE: There must be no leading spaces before the pipes, and no blank lines.

Running Scripts on PAWS

Using 'checkimages' script

➜ Check recently uploaded files for description, license, etc.

```
@PAWS: pwb.py checkimages -simulate -limit:10

Bot: Marking newly uploaded untagged file

SIMULATION: edit action blocked.

Page [[File:Foobar.svg]] saved

[[Andrya Test Heart.png]] seems ok, license found: {{Cc-by-sa-4.0}}...

Player-documentation-ooui.png seems to have a fake license:
Template:Uncategorized, reporting...    Mar 2022 14:34:54 (UTC)

SIMULATION: edit action blocked.
```

Trainer notes

* Checks images for content

Using Python session on PAWS

Creating a wiki page

```python
from pywikibot import Site, Page


site = Site('test', 'wikipedia') #we connect to test.wikipedia.org

page = Page(site, 'March Pywikibot Workshop') #title and page name


page.text = 'This is the second in our series of workshops' #page content

page.save()
```

Trainer notes

* Create a page from a Python session.

* Show that page does not exists

Try-out time!

➜ **How do you plan to use these scripts?**

➜ **Share the outcomes & errors in the chat as you try them out!**

THANK YOU

Seyram Komla Sapaty, Srishti Sethi
✉ ssapaty@wikimedia.org, ssethi@wikimedia.org
🐦 @komlasapaty, @srish_aka_tux

WIKIMEDIA
F O U N D A T I O N

Trainer notes
* Encourage people to join the telegram chat https://t.me/+Z_b1MR8O0wAzZmVh
* Share about upcoming workshops
* Address discussion questions in the etherpad