

# Running basic scripts via Pywikibot

Run as part of the “Small wiki toolkits” project



**WIKIMEDIA**  
FOUNDATION

# 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

# Agenda

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



# Working with Pywikibot scripts



**WIKIMEDIA**  
FOUNDATION

# Managing categories



**WIKIMEDIA**  
FOUNDATION

# 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.
```

# 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
```

```
...
```

# 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
```



# 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
```

# Try-out time!

- How do you plan to use these scripts?
- Share the outcomes & errors in the chat as you try them out!



**WIKIMEDIA**  
FOUNDATION

# Maintaining pages



**WIKIMEDIA**  
FOUNDATION

# 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.
```

# 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...
```

# 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.
```

# Try-out time!

- How do you plan to use these scripts?
- Share the outcomes & errors in the chat as you try them out!



**WIKIMEDIA**  
FOUNDATION

# Running Pywikibot on PAWS

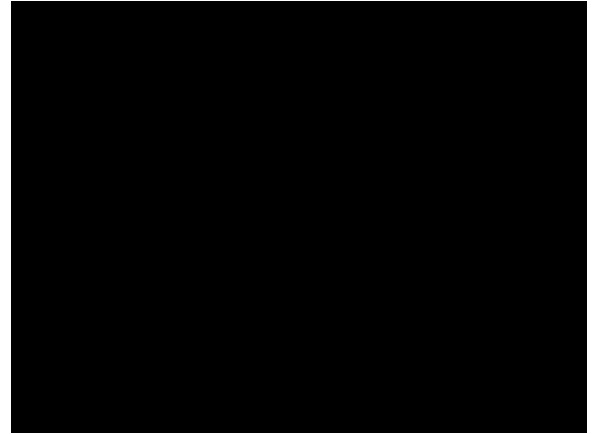


**WIKIMEDIA**  
FOUNDATION



# 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

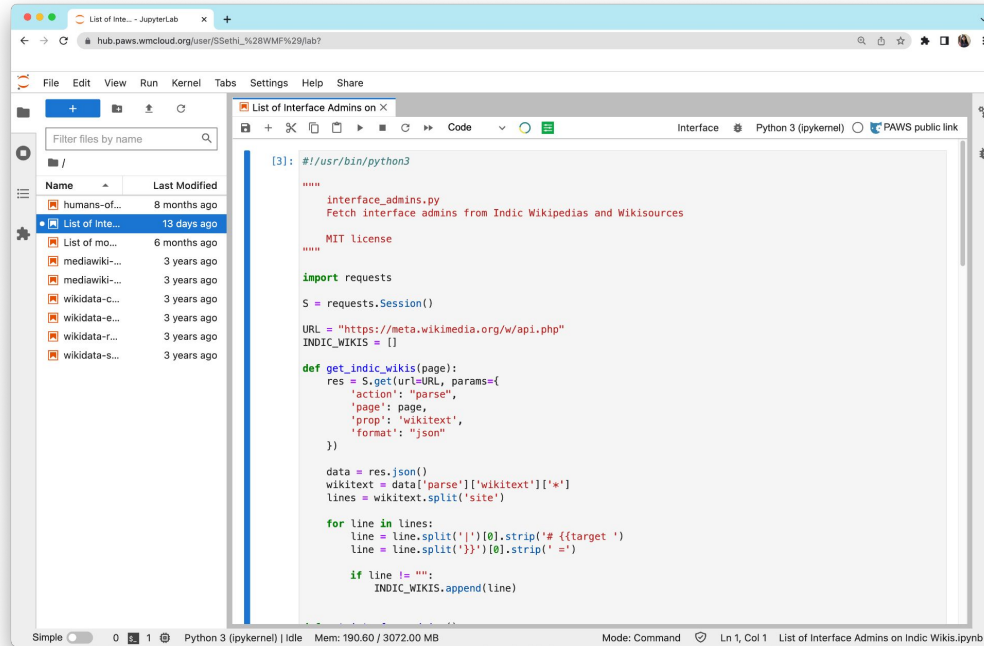


# When to use PAWS

- 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



# How to use Pywikibot in PAWS



The screenshot shows a JupyterLab environment. On the left, a file browser displays a list of files with columns for 'Name' and 'Last Modified'. The file 'List of Inte...' is selected. The main area is a code editor with the following Python code:

```
[3]: #!/usr/bin/python3

"""
    interface_admins.py
    Fetch interface admins from Indic Wikipedias and Wikisources
    MIT license
"""

import requests

S = requests.Session()

URL = "https://meta.wikimedia.org/w/api.php"
INDIC_WIKIS = []

def get_indic_wikis(page):
    res = S.get(url=URL, params={
        'action': 'parse',
        'page': page,
        'prop': 'wikitext',
        'format': 'json'
    })

    data = res.json()
    wikitext = data['parse']['wikitext']['*']
    lines = wikitext.split('site')

    for line in lines:
        line = line.split('|')[0].strip('# {{target }}')
        line = line.split('}}')[0].strip(' =')

        if line != '':
            INDIC_WIKIS.append(line)
```

# Logging into PAWS

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

[Use your mediawiki account]

# Working with PAWS

Accept PAWS permissions

Hi Almoksey,

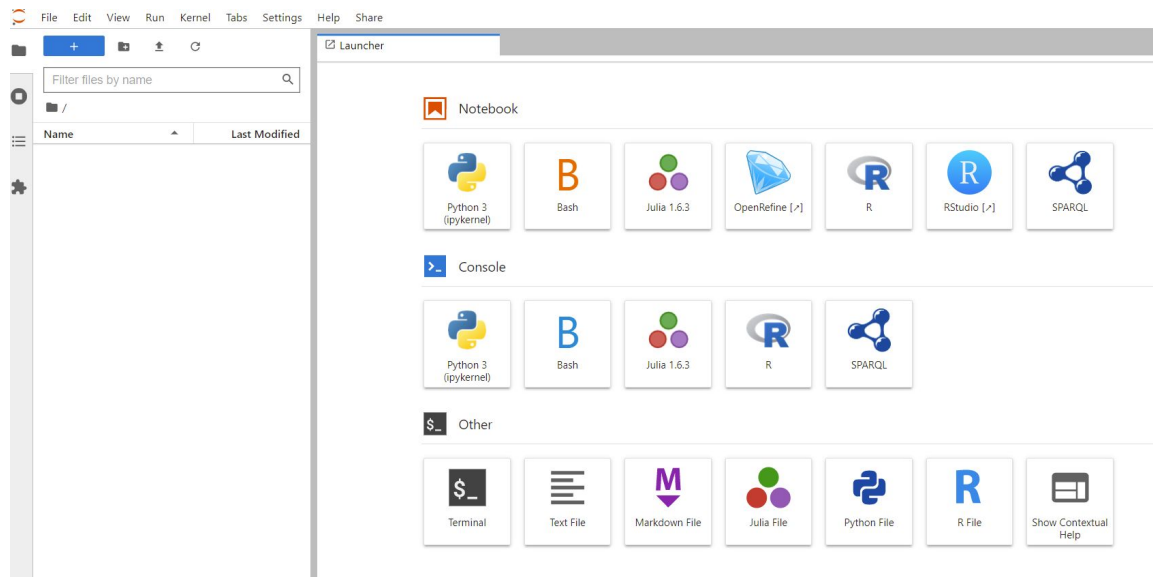
In order to complete your request, **PAWS** needs permission to perform the following actions on your behalf on all projects of this site:

- **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](#)

# PAWS Interface

Notebooks, Console, etc



# 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

# 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
```





# 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

# 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

# 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



# 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-ouui.png seems to have a fake license:
```

```
Template:Uncategorized, reporting...  21 Mar 2022 14:34:54 (UTC)
```

```
SIMULATION: edit action blocked.
```

# Using Python session on PAWS

## Creating a wiki page

```
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()
```

# Try-out time!

- How do you plan to use these scripts?
- Share the outcomes & errors in the chat as you try them out!



**WIKIMEDIA**  
FOUNDATION

# THANK YOU

**Seyram Komla Sapaty, Srishti Sethi**

✉ [ssapaty@wikimedia.org](mailto:ssapaty@wikimedia.org), [ssethi@wikimedia.org](mailto:ssethi@wikimedia.org)

🐦 [@komlasapaty](https://twitter.com/komlasapaty), [@srish\\_aka\\_tux](https://twitter.com/srish_aka_tux)



**WIKIMEDIA**  
FOUNDATION



Copyright © 2022, [Seyram Komla Sapaty](#), [Srishti Sethi](#) and the [Wikimedia Foundation](#).  
This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](#).