

Outreachy Internships to create Wikidata bots

[[User:Mike Peel]]

31 October 2021

What is Outreachy?

- Outreachy.org: "Outreachy provides **internships in open source** to people subject to systemic bias and impacted by under-representation in the technical industry where they are living."
- Runs twice yearly, fully remote
- Internships last around 3 months, and are paid US\$6,000
- WMF participates each round, typically around 4 projects/round



What projects?

- Round 22: "Synchronising Wikidata and Wikipedias using pywikibot" - <https://phabricator.wikimedia.org/T276329>
- Round 23: "Automatically matching new Wikipedia articles with Wikidata items using Python" - <https://phabricator.wikimedia.org/T290718>

Ideas for projects [edit]

- [Automatically matching new Wikipedia articles with Wikidata items using Python](#)
- [Create Documentation for MW-Docker](#)
- [Develop a tool for large-scale open image similarity](#)
- [Develop learning toolkits and videos to demonstrate the use of essential tools for Wikimedia](#)
- [Improve Wikidata support on Programs & Events Dashboard](#)
- [Refactor Mediawiki tests to use WebdriverIO Async](#)

https://www.mediawiki.org/wiki/Outreachy/Round_23



How is the intern selected?

- No curricular vitae or application letters!
- Applicants are pre-approved by Outreachy, and then select projects
- One-month 'contribution period', where applicants go through a set of starter tasks to demonstrate their capabilities
- 'Final application' then describes what they did with these tasks, provides extra context (e.g., past engagement with open source), and includes an estimated timeline of the work they want to do.
- Then have to pick one of them!! (or if you're lucky, two...)
- For the mentor: 5-10 hours work/week for 4 weeks

Example starter tasks

1. **T278860** Look through a class of articles (e.g., books, authors, games, etc.) and identify how the information in the article matches with the Wikidata properties for that topic
2. **T278863** Set up pywikibot on your computer, and understand how it interacts with Wikidata
3. **T278997** Write a function that reads in a Wikipedia article, parses it for specific information, and prints it to screen
4. **T279288** Write a function that writes new structured data into Wikidata (ideally linked to the function above)
5. **T279289** Write a function that searches Wikidata for a match to a term
6. **T279290** Loop through a Wikipedia tracking category and import IDs into Wikidata

From Round 22 (<https://phabricator.wikimedia.org/T276329>).

For Round 23, #1, 2, and 5 were reused.



What happens during the internship?

- Students work full-time on the project. Mentors need ~5hours/week.
- Coding/documenting + blogging (expected to blog every 2 weeks)
- You mentor them remotely, e.g., by Slack/Zulip, or by telephone or video calls
- Make sure they're fully integrated into the project and Wikimedia community, and aware of community norms etc.
- If they have problems, expected they'll try to solve them, and come to you if they can't.
- Description of mentor role at <https://www.mediawiki.org/wiki/Outreachy/Mentors>

Outcomes from Round 22 project

- Ammar and Nirali both set up Wikidata bots + had tasks approved
- For User:AmmarBot:

Approved tasks [\[edit \]](#)

- (Outreachy) Importing [official website \(P856\)](#) from English Wikipedia; Status: Running
 - (Outreachy) Importing [LepIndex ID \(P3064\)](#) from English Wikipedia; Status: Running
 - (Outreachy) Importing [Find A Grave memorial ID \(P535\)](#) from English Wikipedia; Status: Running
 - (Outreachy) Importing [maximum capacity \(P1083\)](#) from English Wikipedia; Status: Running
 - (Other) Fix typo in Belarusian description on select items; Status: Running
 - (Other) Replace [stated as \(P1932\)](#) qualifier with [named as \(P1810\)](#) on select items with [GND ID \(P227\)](#) statement; Status: Running.
-
- NiraliBot similar. ~28k edits between them.
 - Importing data from Wikipedia infoboxes, identifiers and lists

What happens afterwards?

- Write it up! For example, on diff blog:
<https://diff.wikimedia.org/2021/10/01/synchronising-wikidata-and-wikipedia-an-outreachy-project/>
- Hopefully students continue engaging with the project/community after internship ends (but not required)
- Interns can go on to co-mentor for future projects, if interested.

How to get involved

- Community members can mentor projects!
- There is a call for proposals every round, watch wikimedia-!l
- You have to draft a project description, define starter tasks, and ideally find a co-mentor. Then propose it for inclusion in the next round.
- I'm looking for a co-mentor with Lua experience for future proposals! (co-mentors with Python experience also would be great!)

Thanks for your attention!

Get in touch with us:

[[User talk:Mike Peel]]

Credits

Information about the licence of your slides, the pictures used in your slides, links to additional resources, etc.

