

Improve the article recommendation pipeline

Profile Information

Name: Muhammad Usman

IRC nickname on Freenode: muhdusman

Email: me@usmanmuhd.com / muhdusman98@gmail.com

Webpage: <https://usmanmuhd.com>

GitHub: <https://github.com/usmanmuhd>

Location: Bangalore, India

Typical working hours: 3 AM to 3 PM GMT. Will be available anytime on prior notice.

Timezone: +5:30 (IST)

Synopsis

The project aims to improve article recommendation pipeline by solving the various issues in the article-recommender projects. The various issues currently ready to be worked upon are:

- Task 1: **T216721: Remove duplicate Wikidata items from article recommendations**
- Task 2: **T215222: Recommendation API translation endpoint stopped working**
- Task 3: **T216750: Article recommendation API: replace WDQS with MW API**
- Task 4: **T211980: 'morelike' recommendation API: Bulk import data to MySQL in chunks**

The project aims to solve more issues going forward.

The steps that will be followed while solving each issue will be:

- Setup development environment.
- Reproduce the error.
- Fix the error in development.
- Write unit tests for the changes.
- Document the changes.
- Deploy the fix.
- Check for any more issues in production.

Mentor: [@bmansurov](#)

Have you contacted your mentor already? Yes

Timeline

Period	Task
May 6 to May 26	Community Bonding Period. Discuss the existing issues and understand more potential issues with the article-recommender pipeline. Setting up the development environment for the various projects throughout the pipeline. Understanding the deployment process being used to push code to production. Adding the corresponding smaller todo's in Phabricator.
May 27 to June 2	<u>T216721: Remove duplicate Wikidata items from article recommendations</u> Reproduce the bug and fix it in development. Write test cases as needed.
June 3 to June 9	Document the patch. Merge into main repository. Deploy to production and make sure that the fix works correctly.
June 10 to June 16	<u>T215222: Recommendation API translation endpoint stopped working</u> Find out the exact location which is causing the error. Fix the error in development.
June 17 to June 23	Write test cases for the changes made. Document the patch.
June 24 to June 30	Evaluation June 24-28, Merge the patch into main repository. Deploy to production and make sure that the fix works correctly.
July 1 to July 7	<u>T216750: Article recommendation API: replace WDQS with MW API</u> Work on replacing the call to Wikidata Query Service with a call to Mediawiki API.
July 8 to July 14	Make changes to tests as needed by the new patch. Document the changes. Merge into main repository.
July 15 to July 21	Deploy to production and make sure that the fix works correctly.
July 22 to July 28	Evaluation July 22-26, Buffer time.
July 29 to August 4	<u>T211980: 'morelike' recommendation API: Bulk import data to MySQL in chunks</u> Work on splitting the data into chunks and adding it into MySQL.

August 5 to August 11	Write tests to ensure the data import works as desired. Document the changes. Merge into main repository.
August 12 to August 19	Deploy to production and make sure that the fix works correctly.
August 19 to August 26	Final Evaluation
Future Work	Continue to work on bugs. Mentor students in GCI and GSoC.

Note: This is a tentative timeline and will be adjusted in consultation with the mentor.

Deliverables

- Fix the above 4 issues and other issues as needed.
- Documentation and unit tests for the changes.
- Deploy fixes and make sure that they work correctly in production.

Participation

- Will be working around 40 hours each week.
- Work on my forked repository and merge it to the main repository as and when a particular issue is solved.
- Available on IRC during my working hours.
- Available on Email outside working hours.
- Update the status of the issues being worked upon as a comment in Phabricator.
- Publish a blog once every two weeks.

About Me

I am a pre-final year undergraduate student, pursuing Bachelor of Technology in Computer Science and Engineering from PES University, Bangalore, India. I have been contributing to open source since August 2017. My first big contribution to open source was adding Kannada Language support to [cltk project](#). Since then I have made numerous contributions to various open source projects. I have developed a keen interest in working on open source projects. It gives me immense pleasure to work on real world issues, solve them and see my code working perfectly in production.

I first heard about GSoC last year but could not apply as my exams were overlapping with the program. This time I have no other commitments during the duration of the program. I hope this program will provide me a great chance to delve into bigger open source projects and provide me exposure to the open source community.

I am very much inspired by the mission of the Wikimedia Foundation. It is something I very much feel myself and would like to make it happen. Coming from India, it has provided me a very close view of the disadvantages that the underprivileged face due to the costs and logistics involved with gaining quality knowledge. This project would give me a chance to work on fulfilling that mission and help provide everyone an equal chance at gaining quality education and help the world grow.

Past Experience

I have experience with working on Python, JavaScript, Java, HTML and CSS along with other technology stacks as well. I have worked with MySQL, PostgreSQL and MongoDB in database technologies. I have used git for all personal and open source projects.

Microtasks completed:

- **T219505: Recommendation API: output source language**

Some of the other open source projects that I have significantly contributed to are:

- Gitcoin:
 - <https://github.com/gitcoinco/web/pull/2334>
 - <https://github.com/gitcoinco/web/pull/3221>
 - <https://github.com/gitcoinco/web/pull/2548>
 - <https://github.com/gitcoinco/web/pull/3505>
- cltk:
 - <https://github.com/cltk/cltk/pull/597>
- Hydrus:
 - <https://github.com/HTTP-APIs/hydrus/pull/290>
 - <https://github.com/HTTP-APIs/hydrus/pull/287>

All my other open source contributions and personal projects are available on <https://github.com/usmanmuhd>.

Link to the proposal on Phabricator: <https://phabricator.wikimedia.org/T218971>