

Name: Nurul Ariessa Binti Norramli

E-mail address: ariessa.norramli@gmail.com

Other information that may be useful to contact you: ariessa on #apertium

Why is it you are interested in machine translation?

Because machine translation platforms like Apertium, Google Translate and Bing Translator helps me a lot in writing. They are very useful when you need simple translations from one language to another. Other than that, the concept of machine translation itself is simply fascinating for me.

Why is it that you are interested in Apertium?

Because I love open source stuffs and Apertium sure is one! Plus, I want to contribute to open source projects and I believe Apertium is the platform for me to start and learn about contributing. To be honest, I've known about Apertium since 2015.

Which of the published tasks are you interested in? What do you plan to do?

None. I plan to add a new language pair, which is English-Malaysian (en-ms).

Title

Add English-Malaysian language pair (en-ms)

Why Google and Apertium should sponsor it? How and who it will benefit in society?

I saw that Apertium doesn't offer translation for English-Malaysian and vice versa. As a native Malay speaker, I feel that the English-Malaysian language pair should exist in Apertium. Because the Malaysian language or Bahasa Melayu is used in numerous countries like Singapore, Indonesia, Malaysia, Brunei and some part of Thailand. It is a vital language for the Southeast Asians. By doing this, Apertium can reach more audience. Consequently, it could attract more developers or translators to contribute to Apertium. This could then lead to an active and ongoing development of Apertium. Not just that, more language pairs will be added too since most Southeast Asians speak more than one language. In my case, I know English, Bahasa Melayu, basic Japanese and basic Arabic. All in all, Google and Apertium should sponsor this project as it can introduce the Southeast Asians to this machine translation platform.

Work plan

Before GSOC starts

- Installed Apertium on Ubuntu
- Installed ltoolbox
- Installed id-ms and practice with this language pair

- Read the Apertium wikis on how to add a new language pair
- Research about linguistics related to Bahasa Melayu and English
- Begin working on the morphological dictionary

Week 1 - Week 2:

- Work on the morphological dictionaries

Week 3:

- Work on the bilingual dictionary

Week 4:

- Get parallel corpus by translating Wikipedia articles from English to Malaysian
- Expand bilingual dictionaries with Giza++

Deliverable #1 Morphological dictionaries and bilingual dictionary

Week 5 - Week 6:

- Get parallel corpus by translating Wikipedia articles from English to Malaysian
- Write the transfer rules

Week 7:

- Work on bidix
- Add more words to the dictionaries

Week 8:

- 3-day break
- Work on bidix

Deliverable #2 Transfer rules

Week 9:

- Expand bilingual dictionaries
- Write more transfer rules

Week 10:

- Work on bidix

Week 11:

- Write an article about the grammatical difference between English and Malaysian

- Debug any problems

Week 12:

- Evaluate using testdoc, WEP and trimmed coverage
- Brush up the project

Project completed

List your skills and give evidence of your qualifications.

C, C++, Python, JavaScript, BASH and some HTML & CSS. I'm currently taking an introductory course about C language at a local university. As for the other languages, it was self-taught.

Tell us what is your current field of study, major, etc.

I'm currently taking a pre-university program called Foundation in Engineering at Universiti Teknologi Mara Dengkil, Malaysia.

Convince us that you can do the work.

I was a Google Code-In (GCI) 2015 finalist for MetaBrainz Foundation. So, I dare say that I can work on deadlines. I'm a very passionate and hardworking person when it comes to writing and coding. Moreover, I always give my all when working. On top of that, I'm also a freelance writer at Upwork. I'm used to work in 2-day deadline. Thus, I can assure you that I got what it takes.

In particular, we would like to know whether you have programmed before in open-source projects.

Yes but it was only a single pull request task for GCI 2015.

List any non-Summer-of-Code plans you have for the Summer, especially employment, if you are applying for internships, and class-taking.

Aside from participating in Google Summer of Code, I'm going to work as a freelance writer. Even so, my writing time is only 3 hours every day.

Be specific about schedules and time commitments. we would like to be sure you have at least 30 free hours a week to develop for our project.

I'm having my finals right now. But when GSOC starts, I'll be on a 5-months break. I'll be finishing school by early April 2017. So, I'm confident that I'll have at least 30 free hours a week.

Coding Challenge

I made a Malay translation file.

<https://github.com/ariessa/GSOC-2017/blob/master/ms.txt>