Google Summer of Code 2019

FOSSology: Continuation of Atarashi OSS

horizontal line

29July **20**19 - 04 August 2019 [ Week 10 ]

# Delivered By: Ayush Bhardwaj ([HastagAB](https://github.com/hastagAB))

Mentored By:

* Anupam Ghosh ([ag4ums](https://github.com/ag4ums))
* Gaurav Mishra ([GMishx](https://github.com/GMishx))
* Aman Jain ([amanjain97](https://github.com/amanjain97))

# Goals for the Week

1. Finalizing the evaluation Script & Integration of Atarashi
2. Coding the New Algorithm: Semantic Text Similarity

# Tasks Accomplished

1. **Finalizing the evaluation Script & Integration of Atarashi**

PR was created lately but some changes were asked after the review. Some changes were done after that. The evaluation script is now almost ready to be merged.

Link to the PR: <https://github.com/fossology/atarashi/pull/57>

Integration with Atarashi needs some changes in the Agent\_Test folder. Few renaming was done but the Travis tests still fail. Some function and Makefile needs to be altered, need some help with that to finalize the PR.

Link to the PR: <https://github.com/fossology/fossology/pull/1408>

2. **Coding the New Algorithm: Semantic Text Similarity**

I started coding the algorithm. We need to train our own model in doc2vec so looked for some dataset. I didn’t find the perfect one but can use some for testing. The text preprocessing in SpaCy is awesome and won’t take many effort. We now need to train the model and work upon that.

# Conference Call

**30th July 2019(Tuesday) Timings : 6:00 PM onwards**

**Attended By:** Anupam Ghosh (Mentor), Aman Jain (Mentor), Ayush Bhardwaj

Discussed the algorithms and what needs to be done. The evaluation script and Atarashi Integration was discussed.