Google Summer of Code 2019

FOSSology: Continuation of Atarashi OSS

horizontal line

15July **20**19 - 21 July 2019 [ Week 08 ]

# 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. Coding the evaluation Script
2. Testing the Evaluation script with various test data
3. Studying and coding the New Algorithm : Semantic Text Similarity

# Tasks Accomplished

1. **Coding the evaluation Script**

The evaluation script was the priority as it will be helpful to evaluate any of the newer as well as the older algorithms. The script was coded to know for accuracy as well as the speed of any algorithm.

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

2. **Testing the Evaluation script with various test data**

Two test dataset was provided : **spdx-test-data** & **NomosTestfiles**

Spdx test dataset was perfectly organized but the problem was it contains 100% of the license statement which can give 100% accuracy. NomosTest data gives less accuracy but the problem was that it took more than 4 hours to run the test.   
I created one more dataset manually from NomosTest data as well as Spdx Test dataset to make a perfect dataset for unified evaluation.

3. **Studying and coding the New Algorithm : Semantic Text Similarity**

Studied and formulated the approach to code the algorithm in an Organized manner.

# Conference Call

**18th July 2019(Tuesday) Timings : 6:30 PM onwards**

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

Discussed the working of evaluation script. Showed the code of Evaluation script to the mentors. Both the mentors gave the suggestion for the Evaluation Script.

Anupam Ghosh ([ag4ums](https://github.com/ag4ums)) suggested to finalize the work of Integration of Atarashi to FOSSology after changing the names of the files.