

# E E/Cpr E/S E 492 Bi-Weekly Report 5

## Intelligent Code Editor

Client & Advisor: Ali Jannesari

sdmay20-46

John Jago – Software Lead

Keaton Johnson – Systems Lead

Jon Novak – Machine Learning Lead

Matthew Orth – Meeting Facilitator

Garet Phelps – Report Manager

Isaac Spanier – Test Lead

## Weekly Summary

During this sprint, our group focused on labeling the 1,000 sample Java method invocation dataset. Each member was responsible for labeling the natural language part of 167 samples from the dataset. This work will allow us to train a model on a larger dataset that should result in more accurate translation results for our system.

## Past week accomplishments

John Jago

- Dataset Natural Language Labeling
  - Labeled 167 Java method invocations
- Translation server on AWS
  - Began setting up an EC2 instance to host the OpenNMT-py translation server which our IntelliJ plugin will communicate with

Keaton Johnson

- Translation Dataset Labeling
  - I labeled my assigned methods by translating Java to Natural Language

Jon Novak

- Natural language dataset labeling
  - Labeled the 167 java methods from the dataset

Matthew Orth

- Automatic Java Method Invocation Preprocessing:
  - Ran the automatic mined Java method invocation script to generate the dataset

- Worked through some configuration issues related to this such as errors and script hanging
- Dataset Natural Language Labeling:
  - Labeled 167 Java method invocations natural language statement for the dataset
    - Incorporated different word choice and sentence structure to allow better input generalization
    - Updated the Java method invocation portion of the dataset to remove package names

Garet Phelps

- Dataset Natural Language Labeling
  - Labeled the java method invocations with a natural language equivalent

Isaac Spanier

- Dataset Natural Language Labeling
  - Labeled Java Methods with natural language
- AWS Lambda Research
  - Studying other aws python lambda projects

## Individual contributions

<b>Name</b>	<b>Contributions</b>	<b>Bi-Weekly Hours</b>	<b>Cumulative Semester Hours</b>
John Jago	<ul style="list-style-type: none"> <li>● Dataset Natural Language Labeling</li> <li>● Translation server on AWS</li> </ul>	7	45
Keaton Johnson	<ul style="list-style-type: none"> <li>● Translation Dataset Labeling</li> </ul>	7	37
Jon Novak	<ul style="list-style-type: none"> <li>● Dataset Labeling</li> </ul>	8	36
Matthew Orth	<ul style="list-style-type: none"> <li>● Automatic Java Method Invocation Preprocessing</li> <li>● Dataset Natural Language Labeling</li> </ul>	14	67
Garet Phelps	<ul style="list-style-type: none"> <li>● Dataset Labelling</li> </ul>	8	39

Isaac Spanier	<ul style="list-style-type: none"> <li>● Java Labeling</li> <li>● AWS Research</li> </ul>	8	33
---------------	---	---	----

## Plans for the upcoming sprint

John Jago

- Help with AWS system integration
- Complete user interface testing and documentation

Keaton Johnson

- Optimize trained model results

Jon Novak

- Optimize trained model results

Matthew Orth

- Run the 1,000 sample Java method invocation dataset through OpenNMT-py and evaluate the results
  - Identify and research potential accuracy improvement methods

Garet Phelps:

- Complete Java user interface parameter and variable mapping and remapping

Isaac Spanier

- Complete AWS NLTK script integration and OpenNMT-py hosting

## Summary of weekly client/advisor meeting

### Meeting with Professor Jannesari and Hung Phan on 2020-03-09 at 4:30 pm

This meeting was a brief update on the progress we made on the automatic dataset mining and Java code preprocessing scripts. During and after this meeting we ran the mined Java method invocations through the automatic Java code preprocessing script and split the dataset up into 167 samples each to label.

### Meeting with Professor Jannesari and Hung Phan on 2020-03-23 at 4:30 pm

First, our Client, Adviser and Team decided that it would be best to use Zoom for all our remaining meetings this semester. During our first virtual meeting, we discussed our current dataset labeling progress. We also determined that we need to research potential ways to improve the accuracy of the system. This can help us if our results are not as accurate as we planned. We will continue labeling our dataset for next week's meeting.

### Meeting with Hung Phan on 2020-03-30 at 4:30 pm

During this meeting, we discussed our progress on labeling the Java method invocation dataset, which we had completed prior to the meeting. We then discussed our plans to have our entire system connected and working by the end of the next sprint, which is in two weeks. That work is contained in the Plans for the Upcoming Sprint section of this document.