

# E E/Cpr E/S E 492 Bi-Weekly Report 3

## 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 our biweekly report 3 timeframe, we focused on implementing the user interface method parameter and variable mapping to Java type, Java code parameter type automation, Java method invocation dataset mining, configuring the AWS server to host our preprocessing, and updating our design documentation. We also finalized our dataset labeling format to require the user to specify the explicit method call when there are nested or chained methods. This work and some work during our next sprint will prepare us for our dataset labeling and evaluation task.

## Past week accomplishments

John Jago

- Mined Java statement normalization (dataset)
  - Created a Java command line application to replace variables with their types and perform whitespace formatting given a list of method invocations and their respective Java source files
  - This will be used to automate the task of preprocessing the mined Java method invocations in the dataset
  - See `/dataset-scripts/invoation-formatter` on GitLab
    - <https://git.linux.iastate.edu/hungphd/sdmay19-intelligent-code-editor-gitlab/tree/mined-method-invoation-formatter/dataset-scripts/invoation-formatter>
      - If merged, this branch may have been deleted. See the master branch instead.

## Keaton Johnson

- Worked on labelling first 100 code snippets
- Continued work on Data Mining Algorithm
  - Currently retrieving all methods within a repo
  - Implemented a basic Octokit wrapper for easy access to API

## Jon Novak

- Worked on labeling the 100 code snippets
- Will work with Keaton on the data mining algorithm

## Matthew Orth

- Dataset Labeling Method and Automation:
  - Created, tested, and finalized dataset labeling method
    - Documented step-by-step procedure for dataset labeling and processing
    - Automated some steps to make dataset labeling faster and more reliable
  - Ran the labeling method through on samples from the 100 sample evaluation dataset
- Integrated Custom PoS Tagger, Verb-Noun Preprocessing Script, and User Interface:
  - Replaced the default NLTK tagger with our custom PoS tagger
    - Trained the default tagger using the brown tagged corpus
  - Configured the front-end system locally and familiarized myself with the code
  - Used Java process to call NLTK python script from IntelliJ plugin (determined it would be best to create a server microprocess to call Python script due to NLTK dependencies)
    - Modified Verb-Noun preprocessing script to accept command line arguments
- Detailed Design Documentation:
  - Created user flow diagram with the function, input, and output of each component
  - Created list of challenges and potential solution for each component
  - Documented the tools and research used for each component

## Garet Phelps

- Set up IntelliJ plugin local development
- Adjusted the plugin to reflect part of our new preprocessing plan:
  - Swap out variables and primitives for their types.

## Isaac Spanier

- Worked on setting up both the AWS Lambda and EC2 Server to do all of our processing.
  - Made sure to configure both servers for our needs.
  - Looked into setting up a group account for multiple to access the servers
  - Looked into setting up the scripts on the servers
- Modifying the overall schematics for the project

## Individual contributions

Name	Contributions	Biweekly Hours this sprint	Cumulative Semester Hours
John Jago	<ul style="list-style-type: none"><li>Mined Java statement normalization (dataset)</li></ul>	14	20
Keaton Johnson	<ul style="list-style-type: none"><li>Worked on labelling first 100 code snippets</li><li>Continued work on Data Mining Algorithm</li></ul>	15	19
Jon Novak	<ul style="list-style-type: none"><li>worked on 100 code snippets</li></ul>	14	18
Matthew Orth	<ul style="list-style-type: none"><li>Dataset Labeling Method and Automation</li><li>Integrated Custom PoS Tagger, Verb-Noun Preprocessing Script, and User Interface</li><li>Detailed Design Documentation</li></ul>	14	39
Garet Phelps	<ul style="list-style-type: none"><li>Set up IntelliJ plugin local development</li><li>Adjusted the plugin to reflect part of our new preprocessing plan</li></ul>	12	19
Isaac Spanier	<ul style="list-style-type: none"><li>Worked Setting up the AWS Servers and modifying digrams</li></ul>	14	20

## Plans for the upcoming sprint

John Jago

- Continue working on dataset labeling automation script

Keaton Johnson

- Continue mining the 5,000 sample dataset (mine both the Java file and method invocations)

Jon Novak

- Continue mining the 5,000 sample dataset (mine both the Java file and method invocations)

Matthew Orth

- Revise the 100 sample evaluation dataset's natural language statements and test with system

Garet Phelps:

- Continue working on user interface parameter mapping

Isaac Spanier

- Continue setting up AWS server to host NLTK scripts and user group access

## Summary of weekly client/advisor meeting

### **Meeting with Hung Phan on 2019-02-17, 4:30 pm**

We discussed the amount of preprocessing that should be done in order to make an accurate model. We considered two options: one involving generalizing variables to myVar1, myVar2, etc., and another that involves the types of the parameters. In the latter case, instead of myVar1, the preprocessing would replace the value with its type (int, for example). We came to the conclusion that this should produce the best predictions because what methods a developer calls on an object depends heavily on the types involved.

We also agreed to limit the scope for this sprint, requiring the user to type more specific natural language so that the complexity of our preprocessing does not get out of hand.

### **Meeting with Professor Jannesari on 2019-02-24, 4:30 pm**

During this meeting, we updated our adviser, Professor Jannesari, on the progress we made during this sprint. We all agreed that our current project status and future planned work is on the right track, but we do realize that there is only two months left in the semester. Finally, we determined that we will focus on finishing the user interface parameter mapping, automatic dataset labeling script, setting up the AWS server for the NLTK script, Java method invocation mining, and evaluation dataset labeling and evaluation in our system. Once these steps are completed, we will move on to dataset labeling and then running that through our final system.