

Senior Design Bi-Weekly Status Report

SE/CprE 492 Weekly Report #4

Team: sdmay21-35

Team Email: sdmay21-35@iastate.edu

Advisors

Ali Jannesari

Hung Phan

Team Members:

Ahmad Alramahi - Lead Developer

Austin Boling - Meeting Facilitator

Joseph Naberhaus - Project Lead

Ekene Okeke - Report Coordinator

Ethan Ruchotzke - Documentation Manager

James Taylor - Linguistics SME

Project: POS Tagger for Software Documentation

Report Period: March 1st - March 14th

Summary of Progress (Weekly Summary)

This week, the team set up, tested, and began manual tagging work and infrastructure. This week served as a small scale test of our tagging pipeline, as well as a good moment to bugfix our tokenization software.

In terms of tagging infrastructure, the team tested out the pipeline on a small scale set of documents (leetcode 0-5). These articles showed off the shortcomings of our autotagger, and allowed us to update the autotagger to make life easier.

Autotagger updates included usability changes (making bold text a different color) and extra additions to change the way in which numbers are parsed.

Overall, the testing of the manual tagging pipeline was a great success, and the changes we made to it this week will make the manual tagging process faster than expected.

Meeting Notes (3-8-2021, 3-15-2021)

- Our main meetings were cancelled due to Hung's graduate work.
- We updated Hung with progress as a replacement for the main meetings
 - Tokenizer updated and bugfixes galore
 - Start and divide of manual tagging
 - Download 50 javadoc articles for manual tagging, and tag them

Pending Issues

- Final Error fix for tokenizer
 - Fix of unicode issue on splitting negative numbers
 - More bugs will appear as manual tagging begins
 - Splitting apart negative numbers might still require some work
- Finalize manually tagged dataset
 - Split manual tagging among group
 - Each individual will get ~8 articles to tag
 - Report back next week with tagged data in order to start training with small dataset
 - Download 50 javadoc articles for tagging
 - Broaden / diversify the dataset

Austin

Leetcode 33-49

Ekene

Leetcode 18-33

Ethan

Leetcode 6-18

Javadoc 0-3

Joseph

Javadoc 4-18

James

Javadoc 19-34

Ahmad

Javadoc 35-50

(above) The split up for manual tagging work

Past Week Accomplishments - Current Period - Individual Accomplishments

Our hourly work done for the current reporting period is as follows.

Who	What	Individual Hours	Cumulative Hours (Starting 2/22/2020)
Ethan	Tokenizer - grouping of negative numbers, pre-split manual tagging, working through technical issues	5	15
Austin	Manual tagging	2.5	6
James	Examination of nouns and adjective in POS Tagged Software Documentation	4	16
Ahmad	Started writing code for Training Software	3	5
Joseph	Scraper & Training	4	10
Ekene	Running the Tokenizer with data from leetcode	2	3

Plans for upcoming Reporting Period

Our plans for this upcoming reporting period (**March 15th - March 29th**) are broken down as follows:

Who	What	Due When
Ethan	Fix the tokenizer's grouping of negative numbers (manual fixes until then)	ASAP (March 16)
Ethan	Download javadoc, split sources apart for group	ASAP (March 15)
All	Manual Tagging	March 22