SE/CprE 492 Weekly Report #5

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 15th to March 29th

Summary of Progress (Weekly Summary)

This week the tagging process was undertaken in full force, with over 10,000 individual tags completed manually, with many more (almost 4x as many) autotagged by our tagging pipeline.

Each member of the team was given 16 files to manually tag, but due to a lack of deeper reading into the files, the javadoc files were far longer than the leetcode files, meaning the javadoc members had much more work to do. To alleviate this, the second week was a partial "load balancing" period, where we helped eachother finish up the tags. As of the current moment, there are less than 10 files remaining to be tagged, meaning training can begin.

Meeting Notes (3-22-2021, 3-29-2021)
- Very uneventful meetings due to the pure tagging work being done
- As a deliverable for 4-5-2021
  - More formal presentation of results this semester (Ali Jannesari will be attending)
  - Initial training results - how far did 100 samples get us?

Pending Issues

- Training Pipeline Bug Fixing
  - Character inserting extra whitespace
    - LEAD: Weird whitespace tag (possibly a tab or)
    - RESOLVE: ASAP
- Training Pipeline: Validation
  - Create a program to tag using the trained model
  - Compared model-tagged results to manually tagged results
- Training: Completion
  - Figure out the preferred split up of training vs testing vs validation data
  - Train model
  - Compile Results
- Formal Presentation of Results
  - Create a nice presentation to review semester progress
  - Include results of training regime in presentation

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	Manual Tagging	9	24
Austin	Manual tagging	3	9
James	Manual tagging	10	26
Ahmad	Started writing code for Training Software	3	5
Joseph	Scraper & Training	4	14
Ekene	Running the Tokenizer with data from leetcode	2	3

Plans for upcoming Reporting Period

Our plans for this upcoming reporting period **(March 29th - April 12th)** are broken down as follows:

Who	What	Due When
Ahmad	Training Pipeline: Bugfixing	3/31/2021
James, Ethan	Training Pipeline: Validation	3/31/2021
Everyone	Training Pipeline: Training	4/5/2021
Everyone	Formal Presentation of Results	4/5/2021
Everyone	Completion of Manual Tagging	ASAP (3/31/2021)