

sdmay23-20: Automated annotation of source code datasets for ML applications

Week 1 Report

September 21 - September 28

Team Members

Robby Rice — *Digital Content Coordinator*
 Tanner Dunn — *Agile Framework Organizer*
 Max Sutcliffe — *Client Communication*
 Gavin Canfield — *Quality Control*
 Amon McAllister — *Individual Component Design*

Summary of Progress this Report

Team split into two groups to study advanced ML topics that were unfamiliar to us – natural language processing tagging/ processing, and the study of lexers, parsers, and Python modules to parse natural language. Team also filled out team contract and defined leadership roles for each member. Started work on creating a list of reserved keywords and category they belong to begin tagging of source code datasets.

Pending Issues

There are still reserved keywords not assigned to a category. Categories aren't finalized. Decisions about sprints and stories still pending.

Plans for Upcoming Reporting Period

Organize into teams, finalize categories and tags, finish class assignments/ documents, solidify plan for sprints and semester plan.

Individual Contributions

Team Member	Contribution	Weekly Hours	Total Hours
Robby Rice	Created and organized team assignments, research on NLP	2	
Tanner Dunn	Lead work on generating keywords and categories, outlined sprint activities	3	0
Max Sutcliffe	Research on lexers, parsers, and already implemented Python modules	2	
Gavin Canfield	Major contribution to team contract, research on NLP parsing methods	3	0
Amon McAllister	Major contribution on team contract, research on lexers, parsers and already implemented Python modules	2	0

Gitlab Activity Summary

Nothing to report.
