CSC 510 Software Engineering
Project Proposal - Empirical Track

Members	Unity ID
Abhishek Sarangan	asarang
Mahek Chheda	mschheda

Project Title - Examining and analyzing APIs in different communities in an open source community

Introduction:

When we think of an open source community, we find projects for achieving a myriad of different solutions spanning across various real world challenges. So, our goal through this project is to find information about the API usage i.e. classes and programming language usage in big and small communities (like stack overflow, koders.com, Apache, Github etc) and collect their frequency of usage. Moreover, we will track the data as to what percentage of the questions posted are resolved.

Motivation:

As a trend in communities,some libraries/APIs are used more frequently than others. Our goal is to explore the factors that explain why certain libraries and particular subsets of libraries become more popular than others. So we will analyse and try to gauge the usage of such libraries/APIs present, so the frequent ones can be identified and see if there exists a trend in them. We shall be trying to build up on the work done by Dr. Parnin et al in *Crowd Documentation: Exploring the Coverage and the Dynamics of API Discussions on Stack Overflow*.

Project Summary (Research Questions):

We aim to mine the data sets to collect frequency of usage for the APIs spanning over time and the methods that are defined inside them. In addition, we will be tracking the issues that have come up for the libraries in the communities and find out what percentage of them are open or have been resolved.

Implementation:

We plan to use Apache Hadoop and its ecosystem to complete our goals for this project. We will parse through the javadoc and maintain a dictionary of each class and their respective methods which will be used to search in the data sets. We will use d3 and Highcharts for representing the results in form of graphs and charts using Javascript.

Deliverables:

We will summarize our findings in a formal report through visualizations and a source code for running it on different data sets.

Conclusions:

With these metrics and statistics, we aim to make the API developers and other developer community aware of the APIs and the trends, which might help them in understanding the needs and improve them for future usage.

References:

Chris Parnin, Christoph Treude, Lars Grammel, and Margaret-Anne Storey. *Crowd Documentation: Exploring the Coverage and the Dynamics of API Discussions on Stack Overflow*. Technical Report GIT-CS-12-05, Georgia Tech, 2012.