**Project Proposal**

Chinmay Joshi caj2163

Aditya Das ad3574

Siddhant Gada scg2151

In this proposal, we discuss few topics that we have thought of implementing for our project. Our full midterm papers were very different from each other and extending the same idea further didn’t seem possible in any of them. Adiya wrote his paper on Metamorphic and Differential Testing, which is very domain oriented and particular to the software. We contemplated working on a software that could generate these test cases but it was not feasible because every software is extremely different in itself. Chinmay had worked on automatic commit message generation. One way to extend this is discussed in Topic 1. Siddhant worked on a paper where he discussed on how the research is being done on a new architecture of deep neural networks which have the potential to beat the performance and accuracy of the CNNs (Convolution Neural Networks) which are used for various computer vision tasks like Object Detection, Image Classification, Image Segmentation, Facial Recognition etc. Although this topic is very interesting, finding ‘software engineering’ related projects is quite difficult. We discuss few of the topics below:

Topic 1: Testing and comparing different automatic commit message generation methods

In this topic, we look to merge Aditya’s and Chinmay’s full paper topics. We will perform both metamorphic and differential testing on 3 different commit message generators. This topic has lots of dependencies in terms of availability, implementation and working of the generators. We will venture this further in the coming week and try to figure out if it is indeed possible. This seems like a good idea if we can have 3 running models with us.

Topic 2: Code Summarization

We have thought about making a project on a topic which can have a lot of practical use in the industry. We plan to make an application which would take a code snippet as an input, and the output generated would be a summary of what the code is intended to do. This project will be accomplished by using a deep neural network. We will also be testing and analyzing the application to check how the model performs when it is supplied with some vague and wrong inputs.

Topic 3 : Clickbait detection

The web is full of various posts which sound tempting and the users tends to get lured into clicking them. However, not all such topics / links are true or useful, which leads to a waste of time. What we intend to do is, develop a model which will help classify topics as clickbait or news.

We will also conduct multiple surveys which will help us understand what the users feel about clickbaits and if they are able to effectively identify clickbaits. The surveys would include a few topics which we will ask the user to classify as clickbait or news, this will help us analyze the user tends and will also serve as a training dataset.

Dataset : Buzzfeed topic, reddit topics, CNN topics, Times topics.

Topic 4: StackOverflow Tag generation

StackOverlow is a platform where programmers ask questions and answer them. It is a privately held website. There is a dataset on Kaggle for this task but surprisingly don’t have the actual question in them.

We feel that we aren’t ready to commit to a topic as of now and thus will continue to look for more projects.