SEMANTIC CODE RECOMMENDATION USING DEEP LEARNING TECHNIQUES FOR ONLINE Q&A PLATFORMS

Shanil Arjuna Saram

A dissertation submitted in partial fulfilment of the requirement for

Bachelor of Engineering (Honours) degree in Software Engineering

Department of Computing

Informatics Institute of Technology, Sri Lanka

in collaboration with

University of Westminster, UK

2020

Abstract

Developers face many problems in their day to day life. These problems vary from environment setup problems, development problems to deployment problems. They use various resources in their process of finding solutions for these problems they face. Q&A platforms play a vital role helping developers resolve their problems since these platforms provide experience from industry experts. Out of these Q&A platforms, StackOverflow can be mentioned as one the standout Q&A platforms developers turn to when they need to resolve a problem. StackOverflow consists of code examples which play an important role in making it famous among developers.

Even though these code examples are very helpful for developers, they have to spend a considerable amount of time to find relevant examples browsing through multiple posts due to the lexical gap between natural language and code. Currently developers use few keywords in order to search which means the search is limited to those few keywords. Providing a solution to find code examples fairly quickly would impact on the productivity of the developers.

The main objective of this project is to research and develop a solution to directly search for code examples using natural language from Q&A platforms like StackOverflow. This project focuses on following a deep learning approach along with natural language processing techniques to provide a solution to the above mentioned problem.

After analyzing existing systems from general code search, this project has implemented a mechanism with optimized sequence to sequence models using custom question and code embeddings to bridge the lexical gap between code and natural language. At the initial stage the models are trained to recommend code examples from the python programming language. The project is also tested using actual user queries which were extracted. The project was evaluated by technical and domain experts and they have given positive feedback and suggestions for future improvements

Keywords – Software Development, Code Recommendation, Deep Learning, Natural language   processing