

CS 48000 / CS 58005 - Special Topics in CS: Software Design Patterns

Project Proposal

**AUTOMATICALLY MINING AND SUGGESTING STACK OVERFLOW
ANSWERS FOR COMPILATION AND SYNTAX ERRORS IN A SPECIFIC
PROGRAMMING LANGUAGE***

*the programming language is yet to be decided by the team

Team Members

- *Dilara Tekinoğlu – 27868 – dilaramemis@sabanciuniv.edu*
- *Nisa Defne Aksu – 28104 – nisadefne@sabanciuniv.edu*
- *Ayşegül Rana Erdemli – 26744 – aysegulrana@sabanciuniv.edu*
- *Mustafa Göktürk Yandım – 24134 – mgokturk@sabanciuniv.edu*

PROJECT DESCRIPTION

1. MOTIVATION

Developing software free from any syntax and/or compilation errors is a must in any development process. As it is well-known in the software community, this is not always an easy task. Instead, most developers face errors while coding and search for solutions on the Internet. This process, of course, requires manual human-effort and delays the production time.

Our motivation is to reduce this development effort by automating the process of finding and suggesting existing solutions for compilation and syntax errors; thus saving time. With this motivation, we will use Stack Overflow -which is one of most commonly used platforms for finding answers for errors shared by developers all around the world- in our project.

2. GOALS & RESEARCH QUESTIONS

The project aims to help -especially beginner- developers to get quick explanations and suggestions for syntax and compilation errors in their code. Stack Overflow, one of the most popular Q&A websites for programmers, has a lot of useful content that can be used. The goal of our project is to use data from Stack Overflow such that the most-voted answers will automatically be extracted both as code and/or verbal explanations and suggested to the developer. Extracting the related summarized information from Stack Overflow answers for an error could be possible by using ML and NLP techniques and hopefully, a software that is based on those summaries and which can be useful for developers will be developed at the end.

RQ1: To what extent can popular Stack Overflow answers help developers to quickly solve syntax and compilation errors?

RQ2: To what extent is it likely that our program finds a correctly- matching solution from Stack Overflow for a given error?

RQ3: Is our tool useful for finding reliable answers faster than manual search on the Internet?

RQ4: Is it possible to find and suggest most-reliable Stack Overflow answers based on their voting and/or repetitiveness?

3. ARTIFACTS SELECTED FOR THE PROJECT

The software artifacts selected in the study are mainly buggy/erroneous codes, most likely to be retrieved from GitHub repositories or certain buggy-code datasets, their corresponding error messages generated by the IDE and data from Stack Overflow. We will mainly focus on compilation and syntax errors and match the proper Stack Overflow answers with them.

4. ABOUT THE TECHNOLOGIES

We are planning to use ML and NLP technologies as well as scripts written for the extraction of the data and a basic UI to provide the outcome of the ML models to developers. The languages that will be used for the implementation are not decided yet. However, it is most likely that we will use Python for data analysis and ML parts.

5. EXPECTED OUTCOMES

For our first research question -RQ1- we plan to conduct surveys for software developers. Thus; first outcome of the project will be survey results gathered from the industrial experience.

As the main deliverable of our project, we aim to develop a tool that will suggest proper solutions from Stack Overflow to encountered errors in scripts in a specific programming language. Our program will automatically scan Stack Overflow for a suitable solution to that specific error. In case of multiple solution suggestions (which is usually the case), our program is expected to extract the most useful, relevant or approved information.

Finally, we also aim to do a literature review and compare existing techniques with the performance of our tool. Together with the implementation details and results, we plan to prepare a paper in which we will discuss our findings from the conducted literature review. Thus; term paper will be the final outcome of our project.