
Bachelor/Master Thesis Topic

Automated Documentation of Source Code

Motivation and Background

Code documentation is often incomplete or incorrect. To alleviate this problem, many approaches have been proposed to automatically generate code documentation, where different techniques such as static analysis combined with natural language processing (NLP) [1, 2] and dynamic analysis [3] are involved. However, the application of the existing tools is usually limited due to the synthesis templates, i.e. they can only generate documentation for certain types of code fragments, methods or classes.

Goals

The student should develop a prototype tool to automatically summarize a given code fragment into natural language documentation, which can involve the description of parameter/field usage, functional behavior, and context information.

Description of the Task

This task will involve program analysis as well as natural language processing and synthesis techniques to analyze and extract relevant information from the source code and then generate corresponding documentation. A detailed description of the task will be given personally on interest.

Research Type

Theoretical Aspects: *****
Industrial Relevance: *****
Implementation *****

Prerequisite

The student should be enrolled in the bachelor/master of computer science program, and has completed the required course modules to start a bachelor/master thesis.

Skills required

Programming skills in (preferably) Java, understanding of, or willingness to learn, the architectural and statistical foundations needed for the project.

References

- [1] Giriprasad Sridhara, Lori Pollock, and K. Vijay-Shanker. 2011. Automatically detecting and describing high level actions within methods. In Proceedings of the 33rd International Conference on Software Engineering (ICSE '11). ACM, New York, NY, USA, 101-110.
- [2] Paul W. McBurney and Collin McMillan. 2014. Automatic documentation generation via source code summarization of method context. In Proceedings of the 22nd International Conference on Program Comprehension (ICPC 2014). ACM, New York, NY, USA, 279-290.
- [3] Sulír, M.; Porubán, J. Source Code Documentation Generation Using Program Execution. Information 2017, 8(4), 148; <https://doi.org/10.3390/info8040148>.

Contacts

Minxing Tang, Humboldt-Universität zu Berlin, Institut für Informatik, Lehrstuhl Software Engineering, Unter den Linden 6, 10099 Berlin, Germany

Application

Please contact during office hours or write an email with the title: “AD-SC” to se-career@informatik.hu-berlin.de