Master Thesis Topic

Code Generation from Natural Language Documentation

Motivation and Background
Code generation has been one of the most popular applications of natural language processing (NLP) over the last few years. Previous studies [1, 2, 3, 4] viewed code generation as a probabilistic problem of maximizing the conditional probability of generating code $y$ based on a given document $c$, where neural network is usually leveraged to support an encoder-decoder approach. Even though many of the state-of-art approaches have achieved good performance with respect to BLEU (bilingual evaluation understudy) [5], the accuracy of the generated code is still low and the functional correctness remains a main issue.

Goals
The student should develop a prototype to generate general-purpose code (i.e. in programming languages) from a given documentation. This could be achieved by optimizing previous techniques such as Bi-RNN and AST modeling. The goal is to generate functional correct code with higher accuracy, when given a functional description of the target code.

Description of the Task
A detailed description of the task and the underlying techniques will be given personally on interest.

Research Type
Theoretical Aspects: *****
Industrial Relevance: ****
Implementation: *****

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

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

References

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: “CG-NLD” to se-career@informatik.hu-berlin.de