Software Engineering Seminar (WiSe 2019/20)

Deep-learning-based Automated Code Summarization

Description

High-quality code summarization can help developers comprehend programs as well as maintain software. To automate code summarization, two types of techniques are mainly applied – information retrieval and deep learning, where the latter is the main focus of this seminar topic. The deep-learning-based automated code summarization approaches usually apply Recurrent Neural Networks (RNNs) or Convolution Neural Networks (CNNs) with an attention mechanism to generate code comments.

The student is to examine and to discuss the current state of the art of deep-learning-based automated code summarization techniques.

References

[1] Xing Hu, Ge Li, Xin Xia, David Lo, and Zhi Jin. Deep code comment generation with hybrid lexical and syntactical information. *Empirical Software Engineering*, pages 1–39, 2019.

[2] Alexander LeClair, Siyuan Jiang, and Collin McMillan. A neural model for generating natural language summaries of program subroutines. In *Proceedings of the 41st International Conference on Software Engineering*, pages 795–806. IEEE Press, 2019.

[3] Yao Wan, Zhou Zhao, Min Yang, Guandong Xu, Haochao Ying, Jian Wu, and Philip S Yu. Improving automatic source code summarization via deep reinforcement learning. In *Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering*, pages 397–407. ACM, 2018.

Contacts

Minxing Tang (`tanminxi@informatik.hu-berlin.de`)
Software Engineering Group
Institut für Informatik
Humboldt-Universität zu Berlin