Software Engineering Seminar

Automated Semantic Code Search

Description

Searching for program source code is a common task among programmers, with the ultimate goal of reuse. To automate this development task, developers need to provide some form of query for finding a desired code fragment. Such a query may be automatically evaluated on a collection of programs, e.g. obtained from an (open) source code repository. The query result, i.e. the potential matches, may be in turn inspected by a developer to pick the best one. Most approaches proposed so far are syntactic in the sense that a query is based on the programming languages' syntax, e.g. by giving a list of important key words. However, the query results of syntactic approaches are often highly inadequate, which has motivated code search approaches taking program semantics into account. For example, [1, 2] proposes an approach to semantic code search where developers specify lightweight specifications in the form of input/output pairs as queries. An SMT solver identifies potential matches in programs encoded as constraints representing their behavior.

The student should examine and discuss the current state-of-the-art in semantic code search.

References

[1] Kathryn T. Stolee and Sebastian G. Elbaum. Toward semantic search via SMT solver. In *20th ACM SIGSOFT Symposium on the Foundations of Software Engineering (FSE-20), SIGSOFT/FSE'12, Cary, NC, USA - November 11 - 16, 2012*, page 25, 2012.

[2] Kathryn T. Stolee, Sebastian G. Elbaum, and Daniel Dobos. Solving the search for source code. *ACM Trans. Softw. Eng. Methodol.*, 23(3):26, 2014.

Contacts

Simon Heiden (`heiden@informatik.hu-berlin.de`)
Software Engineering Group
Institut für Informatik
Humboldt-Universität zu Berlin