

Computer Science Colloquium

Tuesday, February 26, 2013

Nguyen Engineering 4201

11:00am-12:00pm



Improving Text Retrieval Applications in Software Engineering: A Case on Concept Location

The source code of large scale, long lived software systems is difficult to change by developers. Finding a place in the code where to start implementing a change, also known as *concept location*, can be particularly challenging. Recent approaches based on Text Retrieval leverage the textual information found in the identifiers and comments found in source code in order to guide developers during this task. In addition, text retrieval has been used to address many other software engineering tasks, but wide adoption in industry and education is still ahead of us. Among the factors that deter broader adoption, researchers observed the difficulties of developers to formulate good queries in unfamiliar software and the quality of identifiers present in software.

In order to support developers for writing better queries we propose two *query reformulation* techniques. One is based on feedback provided by the developers, whereas the other one is completely automated and employs machine learning techniques to learn from past user queries. Empirical validation shows that the queries reformulated using our approaches lead to better results in concept location, compared to the original queries and to previous techniques.

In order to improve the quality of identifiers in source code we define a catalog of the most common identifier problems, called *lexicon bad smells*, and propose a series of refactoring operations in order to correct them. We show that the refactored identifiers lead to an improvement in the results of Text Retrieval-based concept location.

In addition, we also investigated how to present to developers the information retrieved from source code during concept location. We studied the use of automated text summarization techniques to generate *code summaries*, which are suited for the task, as they are short, yet informative.

Sonia
Haiduc

Biography

Sonia Haiduc is a PhD candidate at Wayne State University, in Detroit, MI, USA, where she performs research in software engineering. Her research interests include software maintenance and evolution, program comprehension, and source code search. Her work has been published in several highly elective software engineering venues. She has also been awarded the Google Anita Borg Memorial Scholarship for her research and leadership.