

Leveraging Big Software Data to Improve Software Quality

Speaker: Dr. Baishakhi Ray, *University of California, Davis (Ph.D., UT-Austin)*

Date and Time: February 24 (Tuesday), 2015 – 11:00am-12:00pm

Location: 3437 SEC

Abstract:

Improving software quality has been one of the primary concerns of Programming Language (PL) and Software Engineering (SE) research over decades. Researchers developed different techniques, e.g., new languages, automatic bug finding tools, code review processes, etc. to reduce defects. However, there is no evidence to indicate how well these techniques work in real software development scenario. In this talk, I will demonstrate how the abundant archived data from GitHub can be used to evaluate different SE/PL methods. I will also discuss how such empirical insights can further be leveraged to build new development tools to improve software quality and programmer productivity.

In this talk, I will discuss some of my work that analyze different aspects of existing software engineering practices. First, I will describe how a choice of programming language affects software quality. Studying the 728 most popular GitHub projects, written in 17 different languages, we found that language did have a significant effect on certain kinds of defects. Second, I will show what kinds of bugs are caused by copy and paste and how do we detect them automatically? Finally, I will discuss the current adversarial testing practices in different SSL/TLS implementations and demonstrate a practical approach to automatically detect security vulnerabilities in SSL/TLS certificate validation code.

Biography:

Dr. Baishakhi Ray is a postdoctoral researcher at the University of California, Davis. She finished her Ph.D. in 2013 from the University of Texas at Austin. Her research focuses on Software Engineering, in particular, empirical studies, program analysis, and software evolution. Baishakhi analyzes large-scale software repositories to understand on-going software engineering practices. Then, leveraging this data-driven knowledge, she builds novel program analysis techniques and development tools to improve software quality and programmer productivity. Baishakhi has received Best Practical Paper Award from the 2014 IEEE symposium on security and privacy (Oakland) and was also nominated for distinguished paper award from the 2013 IEEE Automatic Software Engineering Conference (ASE). Her research has also been nominated for publication in the “Research Highlights” (RH) section of the CACM and has been widely covered in trade media.