Code Retrieval on the Web

Rosalva E. Gallardo-Valencia
Susan Elliott Sim

University of California, Irvine

Institute for Software Research - Research Forum 2009
Code Retrieval on the Web

- The process of searching the Web for source code to help solve a software development problem.
- It is a phenomenon that has emerged as a result of evolving technologies and software development practices.
- Based on our empirical studies, we argue that it is similar to Code Search in an IDE, Information Retrieval, Software Reuse, and Program Comprehension.
Code Retrieval on the Web

- Different from Code Search in an IDE
  - Involves a great number of open source code and other resources on the Web.

- Different from Software Reuse
  - Involves also reference examples. Different process: Identify requirements ahead vs define them by searching.

- Different from Information Retrieval
  - Keywords represent features, requirements, functionality. Semantic gap: Natural Language and Programming Language.

- Different from Program Comprehension
  - Needs to evaluate many candidates faster. Relevance and Suitability Judgments.
Ongoing Work

- Empirical studies
  - Qualitative studies of search process and strategies.

- Development of better search engines for code retrieval
  - Extract code snippets from web pages such as tutorials, forums, blogs, etc.
  - Integrate code snippets in results from a code search engine (Sourcerer).