UNIVERSITY OF DELAWARE

Sana Malik (2011), Emily Hill, Lori Pollock, K. Vijay-Shanker
Computer & Information Sciences

Sponsored by the CRA-W DREU program

Applying a Software Word Usage Model to Other Programming Languages

Motivation

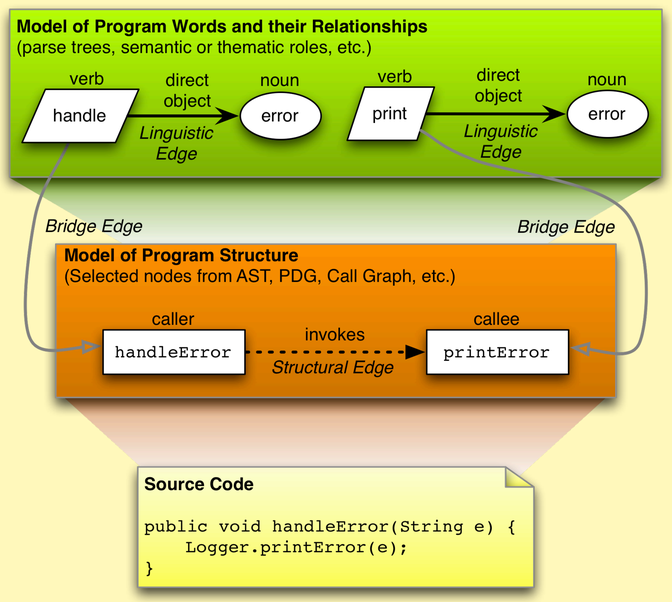
- Programmers spend 60-90% time maintaining software
  - Need automated support from software engineering tools
- **Problem:** Existing tools use either program structure OR linguistic information
- **Overall Research Strategy:** Create a model that captures both linguistic and structural information about a program (SWUM)
  - SWUM is currently implemented only for Java

Research Question

How can the SWUM implementation be generalized for other programming languages?

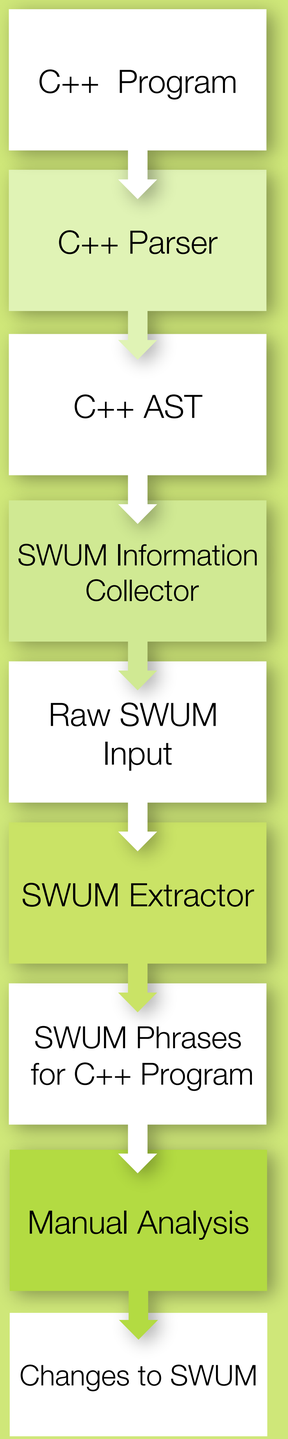
Current Focus: C++ method signatures

SWUM Overview



Given the source code, SWUM captures program word relationships and links them with program structure

Taking SWUM From Java to C++



1. Take a C++ program and **parse** it to make an abstract syntax tree (AST)

2. Give the AST to the **SWUM information collector**, which produces the raw input for SWUM

3. Run the **SWUM extractor** on this information; this gives SWUM phrases for the C++ program

4. **Analyze** each phrase and note the changes that must be made and which SWUM components can remain unchanged

Preliminary Observations

Keywords

- C++ has additional keywords, e.g. `virtual`.
- Public and private are not explicit for each statement in C++.

Classes

- Many functions are not contained within a class, unlike Java.
- Some class methods are defined outside the class.

Return types

- Similar implications as Java, no major differences

Constructors

- C++ also has destructors, which have a similar signature as constructors.

Parameters

- C++ AST visitor does not visit parameters under functions, which are necessary to extract relevant information.

Outcomes

- Developed C++ information collector for input to SWUM extractor
- Suggested SWUM modifications

Next Steps and Future Work

- Expand SWUM to include more features of C++
- Create an Eclipse plug-in from the SWUM for C++ implementation
- Investigate other programming languages
  - Loosely typed languages, e.g. Python