Mining Software Repositories to Promote Reuse

Pankaj K Garg
Zee Source

May 25, 2004
Mining Repositories for Reuse

- Repository – collection of source code, email discussions, bug reports, etc., a la SourceForge
- Reuse – component, source code, or knowledge
  - Find relevant component
  - Use or adapt component
  - Maintain reusable component
LASER…

- Use WordNet for identifying concepts and relationships between concepts
  - e.g., synonyms, is-a, part-of
- Use WordNet derived relationships to suggest reusable components by analogy
  - e.g., client ~ patient => superclass person
LASER...

• Hierarchical reuse
  – recommend superclass

• Parallel reuse
  – recommend class structure

• Current experiments:
  – Do class and method names map to WordNet entries?
  – Do class and superclass names have a hypernym relationship?
Collaborative Filtering…

\[ v(i, c) = \# \text{ of use of component } c \text{ by class } i \]
\[ r(a, c) = \sum (v(i, c) \times \text{sim}(a, i)) \]
Recommend if:

\[ r(a, c) > Tau \]
Collaborative Filtering…

Experiment:
How accurately can we recommend correct components? Does this accuracy depend on the stage of development?
Template Mining...

- Analyze source code for lexical tokens
- Match templates against the tokens
- Construct results based on understanding of tokens and their relationships
  - exact match
  - generalization
  - reduction
  - name only
Template Mining…

• Example template
  – class Matrix
  – methods getMatrix, transpose, plus, minus

• Experiment:
  – Comparison of various methods of matching
Multiproject…

• Reuse *knowledge* from a project for other projects

• Such knowledge can be:
  – Which component implementation among several should a project use?
  – How to adapt a component for reuse?
  – Issues or problems with a component
Multiproject…

• Approaches for:
  – clustering similar components together
  – ranking components based on their use
  – automated adaptation of components for reuse
Summary

• Recent availability of software repositories makes mining for reuse attractive
• Mining may help in selecting reusable components
• Mining may help in adapting or maintaining components