Software Evaluation Testbeds

Alexander Lam
USC Center for Systems and Software Engineering
alexankl@usc.edu

USC-CSSE Annual Research Review
February 13, 2007
Research Problem

• Formulate and evaluate the principles and practices needed for developing software technology evaluation testbeds that will be used to accelerate technology maturation and transition
  – Concepts of operation
  – Requirements
    • Determine the requirements needed for a software evaluation testbed; what will a researcher need to do a technology evaluation
  – Architecture
    • Determine effectiveness and importance of each testbed component
  – Usage
    • Allow researchers to compare their technologies against others
    • Allow practitioners to find technologies right for them
Current Technology Evaluation Testbed Usage

• Examples: RoboCup, Trading Agent Competition (TAC)

• Helps in the validation of a technology
  – Criteria may be in terms not helpful to a software engineer
    • For example, points scored, lowest ticket cost

• Results may not be representative of users’ missions

• Comparison of technologies could be difficult and easy to draw incorrect conclusions

• Problem scope is limited
Testbed Architecture
SCRover – Testbed Instance

• Representative of NASA planetary rover
  – But usable by non-citizen researchers
  – Campus public safety robot
  – Using JPL’s Mission Data System software
Experiments

• Conducted experiments on SCRover with several technologies
  – Mae - an architecture evaluation tool developed here at USC-CSSE by Roshanak Roshandel
  – AcmeStudio- an architecture evaluation tool developed at CMU by David Garlan
  – Peer Review Process

  – Other technologies evaluated as well
    • STRESS, ROPE, and Maude
Peer Review/Mae/Acme Results by Defect Type

Number of Defects

Defect Type

Directional Usage Incomplete Signature Pre-Post Conditions

Peer Review Acme Mae

The bar chart shows the number of defects by defect type for Peer Review, Acme, and Mae. The categories are Directional, Usage, Incomplete, Signature, and Pre-Post Conditions.
Overall Results

• Multiple technology families were evaluated demonstrating that the testbed is capable of evaluating a wide range of technologies.
• The seeded defect approach was effective in identifying the degree to which Mae and AcmeStudio could identify defects of various classes and provided a good way to combine results; prevented researchers from designing their technologies to pass a specific mission/scenario.
• Made researchers rethink how their technologies could be applied to software systems and how the technology should work.
• Results were representative
  – For example, AcmeStudio, Mae, and Maude researchers showed the JPL MDS group how their technology would work on their system.
• Initial SCRover testbed available for pilot operation and has been used by NASA and researchers to evaluate technologies.
• Testbeds are a good way to accelerate technology maturation and transition
  – Better than 18 years as indicated by Redwine and Riddle.
Future Work

• Strengthen ability of current testbeds to be able to better evaluate technologies and accelerate technology maturation and transition

• Looking for large organizations needing technology evaluation testbeds

• Done for NASA originally
  – NASA had to reprioritize their projects so this project wasn’t continued but the problem still exists for NASA