



Common Themes & Future Directions

A new opportunity: Lots of software repositories ...

- Limited access to industrial repositories, but ...
- OSS gives us a corpus of software archives
- Easier to validate predictors of bugs/changes
- Easier to validate utility of software metrics
- Conclusive studies of software evolution
- Record of what developers do: process as it occurs
- Etc etc etc

Themes of Sessions at MSR-05

1. **Evolution & change patterns** – Can repositories help us understand how software changes?
2. **Defect analysis** – Can we use repositories to detect/predict bugs/changes?
3. **Education** – What can we learn from repositories of student programs?
4. ...Lightning talks...
 - a. **Text mining**
 - b. **Infrastructure**, extraction and querying
5. **Integration & collaboration**
6. *(Nothing on software reuse this year)*

1. Evolution & change patterns

- Software evolution: Lots of data. How does change occur?
- Evolution of text vs structure (e.g. AST)
 - Law: “Change increases with depth” [U Md]
 - Enforcing/recognizing rules, eg, open-close [U Md’]
- Evolution of program vs product family (BSD) [Vienna U Tech]
- Clones & their evolution [U Wash]

2. Defect analysis

- Where do bugs come from & where do they go? How well can we predict them?
- Bugs & churn predict bugs & churn. How well? [Planck Institute]
 - “Fix” = Change to correct a previous buggy change
 - When do changes occur? Answer: On Fridays!
- Auto detection of incomplete factoring [Saarland U]
- Not this year:
 - How buggy is this function/file/subsystem/system?
 - “Complexity” predicts bugs, or does it?
 - How know when to “ship” software product?

3. Education

- Mining student programs
- Shadowing students' progress via CVS
 - What does this tell us? About students? About programs? About bugs?
- Warnings predict bugs [U Md'']
- Nothing predicts students' grade in course [U Toronto]

4a. Text mining

- Finding context keywords [Tokyo IT]
- “Talking” comments [IBM Watson]
 - Someone left a note for you in the code
- Analyst-guided mining [U Kentucky]

4b. Infrastructure, extracting and querying

- Taxonomy of mining
 - Kinds/granularity of mining [Kent State]
 - Who/why/what of mining [UVic]
- Querying repositories [UVic]

5. Integration & collaboration

- Data integration: Combining data from multiple sources
 - What sources of data beyond code?
- Mining & 6 sigma to improve process [Fla Atlantic U]
- Identifying developers across repositories [U Rey Juan Carlos]
- Social networks across projects [Nara IST]
- Collecting status of OSS projects [Syracuse U]
- Discovering social/module networks [Chiao Tung U]
- Ethics: Is big brother watching?

Overall...

- MSR: Lots of interesting ongoing work
- Dynamic community
- Lots of open problems
- Lots of opportunity for interesting research
- Future
 - Std extractors? Std guinea pigs? Std benchmarks? Std statistical methods?
 - We're watching the birth of a new way of doing software



Stephan Diehl, Harald Gall, and Ahmed Hassan

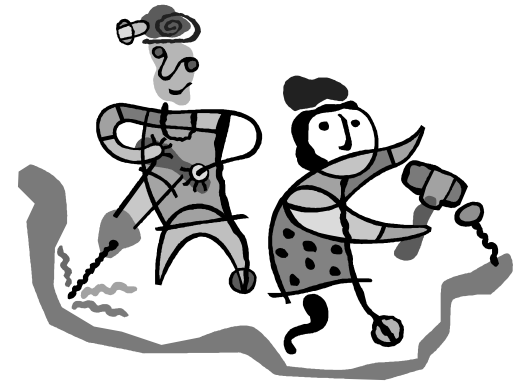
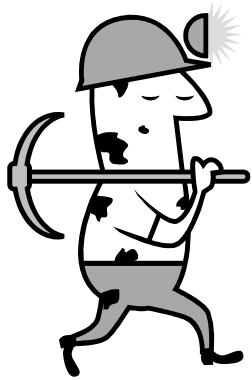
MSR 2006

- MSR 2006 @ ICSE 2006 @ Shanghai
- Umbrella vs. Lightning Talks
- 2 Days ?
 - posters, demos
 - more discussions
 - long and short papers
 - challenge
- Journal



MSR 2006

Mining Software Repositories



MSR 2006 Challenge



MineSweet - MSR Mining Contest 2006

- Challenge: run your mining tools on a common benchmark OSS system
- **PostgreSQL** and preparatory data
- List of analyses tasks
- Offsite and onsite analysis
- Prize for the winner

mine as mine can ;-)



MSR 2006

Mining Software Repositories

Lightning Award

- Mining Version Histories for Verifying Learning Process of Legitimate Peripheral Participants
 - Shih-Kun Huang, and Kang-Min Liu (National Chiao Tung University)

Dinner Details

- Meet in lobby at 6:40
- Broadway Oyster Bar: 736 S. Broadway
- Covered Patio with music

