MSA PROJECT
Seminar on Mining Software Repositories
2010
Imagine

- 30,000 classes
- ~ 5.5 million lines of code
- ~3000 defect per release
- 700 developers

You as Quality Manager

Your product
Your Boss

Test the system!
You have 6 months!
And don’t miss a bug!
Your Testers
The Problem

- Not enough **time** to test everything
  - What to test? What to test first?

- Not enough **money** to pay enough testers
  - To which extend?
Where are the most defect prone entities in my system?
Can we learn from history?
Your Task

- Map bug reports to code changes (transactions)
  - Which code change introduced which bug?

- Map bug reports to files
  - How many defects have been fixed per source code file?
Predicting Defects for Eclipse
by Zimmermann et al. (Session 3)
Your Task

- Bug description
- SVN transactions
- Bug report(s)

Transactions

Bug #19558442

Transaction ID

Extract

Find

Search
Your Task (2)

- Search
- Extract

Bug description

Transaction ID

#19558442

 SVN transactions

changed source code files
Your Input

- Project: JEdit
  - Programmer's text editor
  - Sourceforge project: http://sourceforge.net/projects/jedit/

- We will provide you with
  - Bug reports (XHTML files)
  - Tarball containing SVN (Subversion) mirror

- Download from Webpage
  - http://www.st.cs.uni-saarland.de/edu/msa10/jedit_reports.tar.gz
  - http://www.st.cs.uni-saarland.de/edu/msa10/svn_repo_14_01_2010.tar.gz
Bug Reports
Bug Reports

XHTML

Hint: One transaction can fix more than one bug!
SVN tarball

pearl:~ kim$ svn info file:///Users/kim/Documents/Teaching/msa2010/project/svn_repo_14_01_2010/
Path: svn_repo_14_01_2010
URL: file:///Users/kim/Documents/Teaching/msa2010/project/svn_repo_14_01_2010
Repository Root: file:///Users/kim/Documents/Teaching/msa2010/project/svn_repo_14_01_2010
Repository UUID: 25ceb265-6a78-4dce-b758-64b437aadf78
Revision: 16942
Node Kind: directory
Last Changed Author: ezust
Last Changed Rev: 16942
Last Changed Date: 2010-01-17 06:05:58 +0100 (Sun, 17 Jan 2010)

To parse log you can use:

- `svn log --xml`

or use any SVN library
Programming Language

We don’t care!
But: It has to run on CIP pool computers!
Hand In

- A CSV file containing pairs <transaction_id, bug_id>
  - Which bug was fixed in which transaction?
- A CSV file containing pairs <svn_path, #bugs>
  - How many bugs could be mapped to file.
- Executable program or script (CIP-pool)
- Documentation
  - How to reproduce the provided files?
  - How did you manage the mappings? (heuristics)
  - Why is your program/script good, where could it be improved?
Questions