



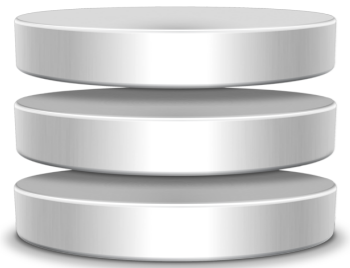
# *Mining Software Repositories*

*Seminar 2010 - The Mining Project*

*Yana Mileva & Kim Herzig*

# Predicting Defects for Eclipse

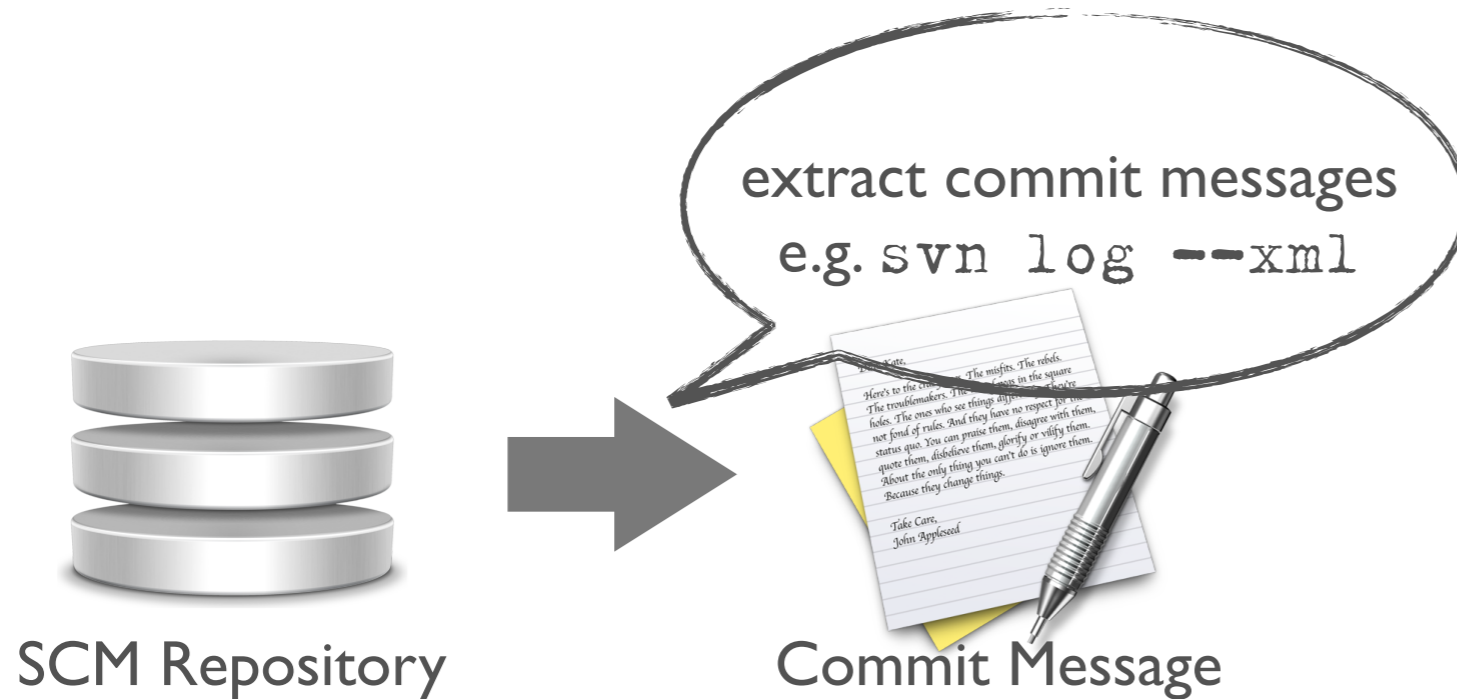
[Zimmermann et al.]



SCM Repository

# Predicting Defects for Eclipse

[Zimmermann et al.]



# Predicting Defects for Eclipse

[Zimmermann et al.]



# Predicting Defects for Eclipse

[Zimmermann et al.]



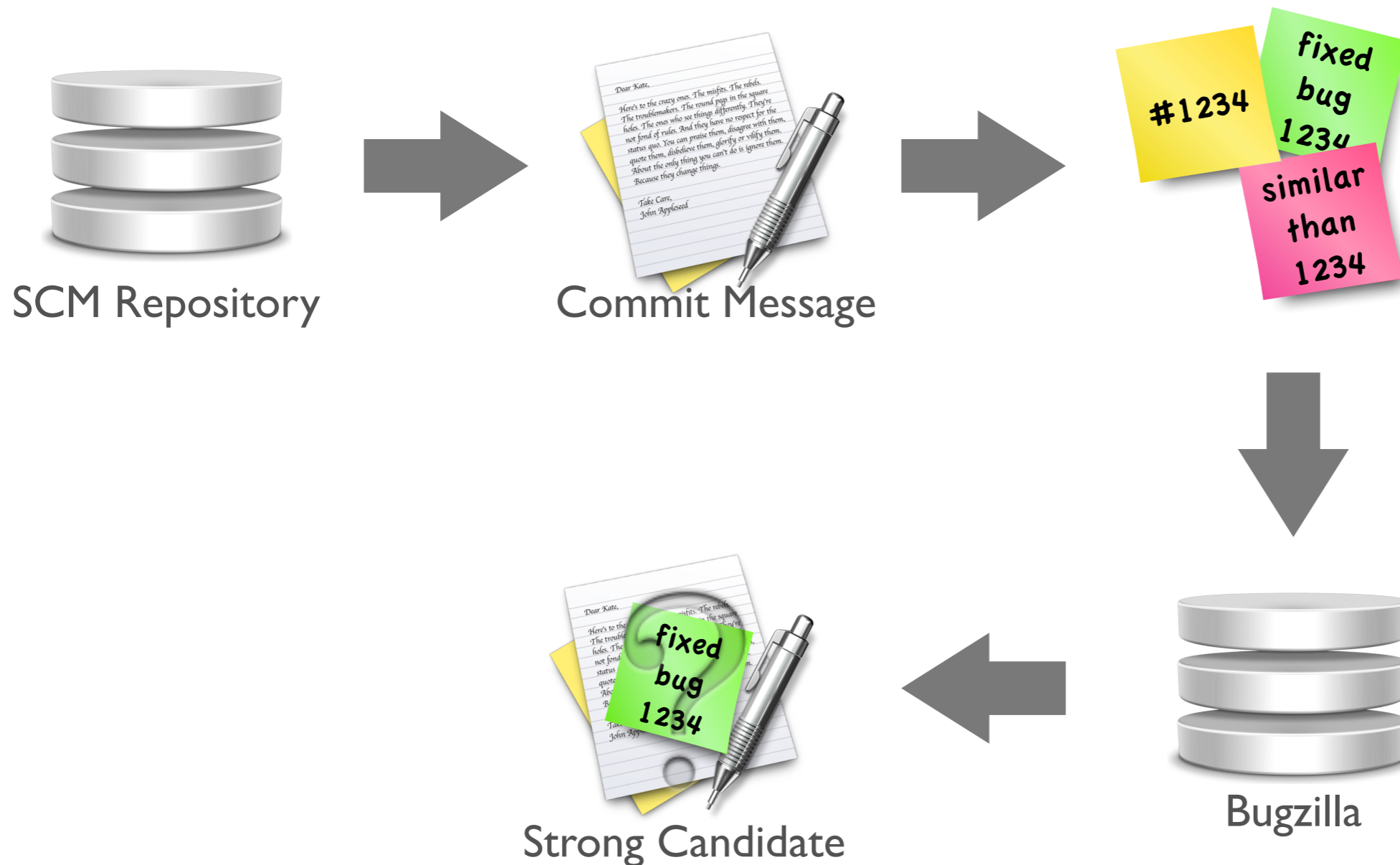
# Predicting Defects for Eclipse

[Zimmermann et al.]



# Predicting Defects for Eclipse

[Zimmermann et al.]





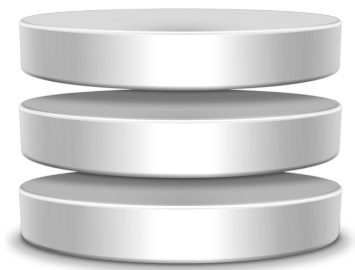
# Predicting Defects for Eclipse

[Zimmermann et al.]



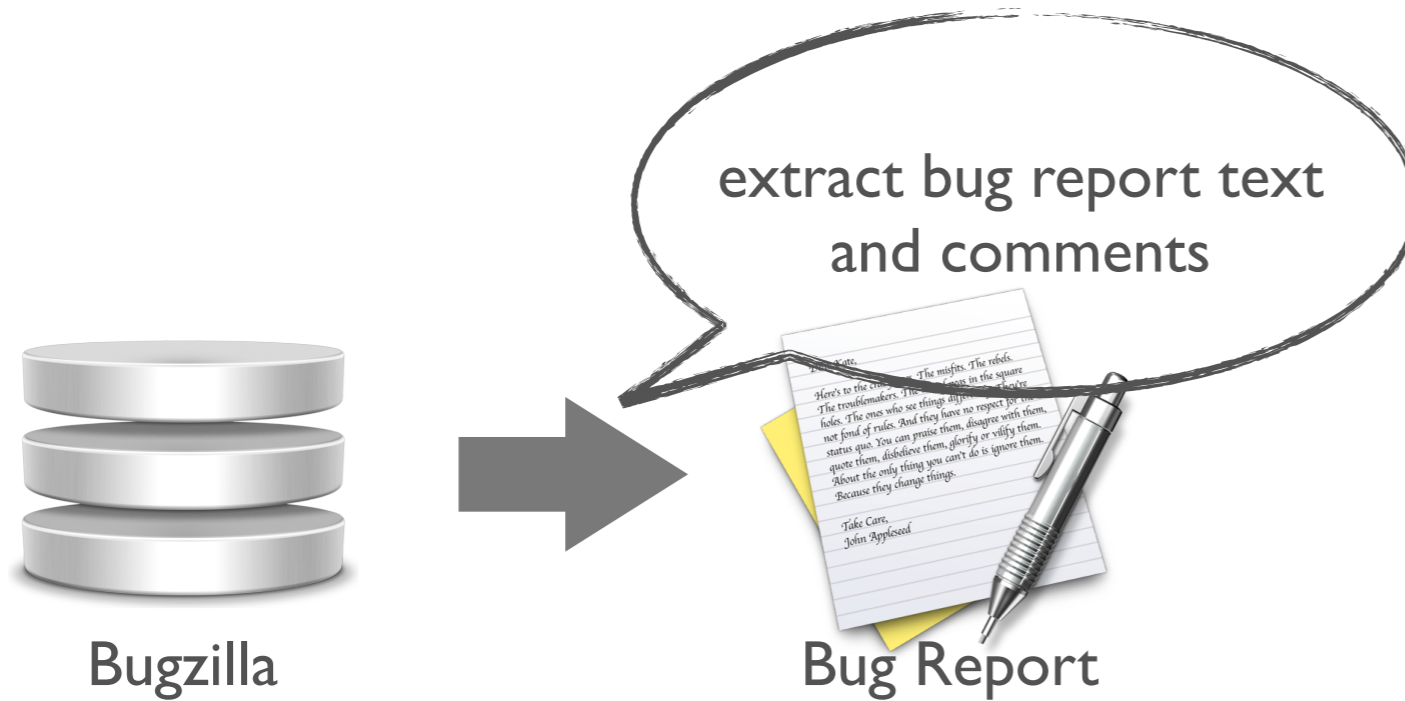


# Your Job



Bugzilla

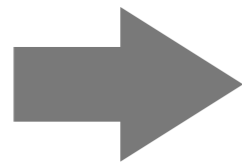
# Your Job



# Your Job



Bugzilla



Bug Report



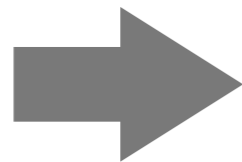
extract reference candidates  
e.g. using regExp



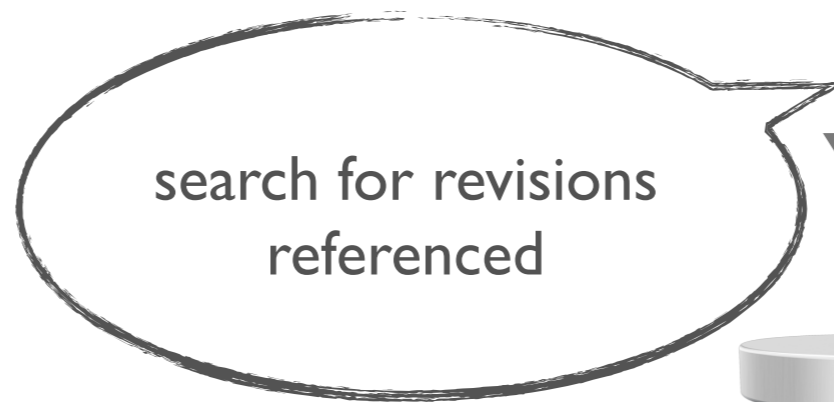
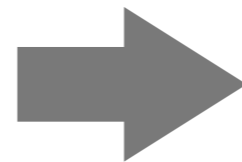
# Your Job



Bugzilla



Bug Report

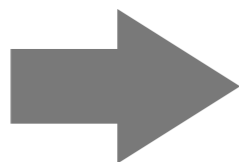


SCM Repository

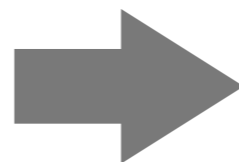
# Your Job



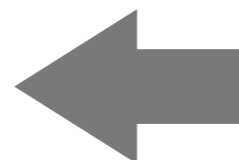
Bugzilla



Bug Report



SCM Repository

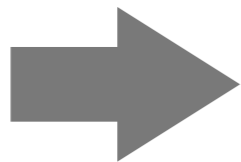


Strong Candidate

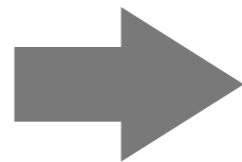
# Your Job



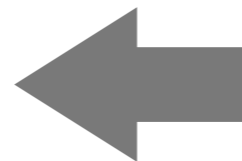
Bugzilla



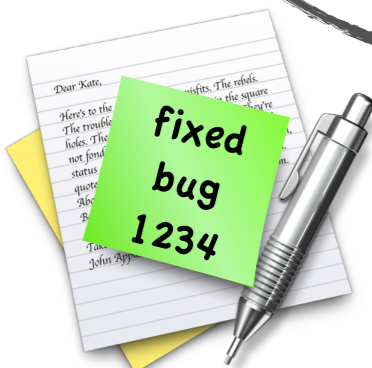
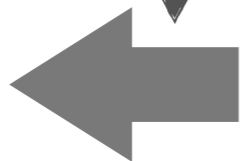
Bug Report



SCM Repository



Strong Candidate



Confirmed Reference

# Your Input

- Project: JEdit
  - ▶ Programmer's text editor
  - ▶ Sourceforge project: <http://sourceforge.net/projects/jedit/>
- We will provide you with
  - ▶ Bug reports (XHTML files)
    - ▶ expect error messages
  - ▶ Tarball containing SVN (Subversion) mirror
  - ▶ Download from Webpage



# Bug Reports

## jEdit

[Share](#) [Donate](#)

[Summary](#) | [Files](#) | **[Support](#)** | [Develop](#) | [Hosted Apps](#) | **[Tracker](#)** | [Mailing Lists](#) | [Code](#)

[Add new](#) [Browse](#)

Tracker: Bugs

[Monitor](#)

5 Can't disable "Restore previously open files on startup" - ID: 2920352 Last Update: Comment added ( [shlomy](#) )

**Details:** Although I have unchecked "Restore previously open files on startup" in Global Options, and the setting appears to be saved (it's still unchecked when I restart jEdit), it still always reloads previously open files on startup anyway.

Using jEdit 4.3, Java 1.6.0\_17, Windows XP SP3.

**Submitted:** Eric Rosenberger ( [twofiftysix](#) ) - 2009-12-23 22:30:09 UTC

**Assigned:** Shlomy Reinstein

**Priority:** 5

**Category:** None

**Status:** Closed


**Group:** None


**Resolution:** Fixed

**Visibility:** Public

### Comments ( 8 )

Date: 2009-12-25 17:24:52 UTC  
Sender: [shlomy](#)  
Fixed in SVN revision 16731.

Date: 2009-12-24 05:12:20 UTC  
Sender: [ezust](#)   
if no files are restored on startup then there should be only an untitled buffer, which should be in the same directory as that specified by global options - file system browser - general - default path.

Date: 2009-12-24 05:09:04 UTC  
Sender: [ezust](#)   
yes, reproduced with -noplugins. It only restores the active open file, none of the others.

# Bug Reports

## jEdit

[Share](#) [Donate](#)

[Summary](#) | [Files](#) | **[Support](#)** | [Develop](#) | [Hosted Apps](#) | **[Tracker](#)** | [Mailing Lists](#) | [Code](#)

[Add new](#) [Browse](#)


Tracker: Bugs

[Monitor](#)

5 Can't disable "Restore previously open files on startup" - ID: 2920352 Last Update: Comment added ( [shlomy](#) )

Details: Although I have  
bal G

Resolution



### Comments ( 8 )

- |   |   |
|---|---|
| Date: 2009-12-25 17:24:52 UTC<br>Sender: <a href="#">shlomy</a>     | Fixed in SVN revision 16731.  |
| Date: 2009-12-24 05:12:20 UTC<br>Sender: <a href="#">ezust</a> ↗ \$ | if no files are restored on startup then there should be only an untitled buffer, which should be in the same directory as that specified by global options - file system browser - general - default path. |
| Date: 2009-12-24 05:09:04 UTC<br>Sender: <a href="#">ezust</a> ↗ \$ | yes, reproduced with -noplugins. It only restores the active open file, none of the others.   |

# Bug Reports

## XHTML

```
<div id="comment_table_container">
  <table class="comments"><thead></thead><tfoot></tfoot></table>
  <div class="yui-gf">
    <div class="yui-u first">
      <input type="hidden" name="arti" value="1501"></input>
      <p>
        Date: 2009-01-30 22:11<br clear="none">
        Sender: <a shape="rect" href="#">k_satoda</a><a shape="rect" href="/help/icon_legend.php?co
      </p>
    </div>
    <div class="yui-u">
      <p>Fixed in r14534.<br clear="none"></p>
      I didn't know the way to remove lines from a visible line to an<br clear="none">
      invisible line. Thanks.<br clear="none">
      <br clear="none">
      <br clear="none">
    </p>
  </div>
</div>
```

**Fixed in r14535**



Hint: One nug can be fixed using multiple transactions!

# SVN tarball

```
pearl:~ kim$ svn info file://<absolute_path>/svn_repo_14_01_2010/
```

```
Path: svn_repo_14_01_2010
```

```
URL: file://<absolute_path>/svn_repo_14_01_2010
```

```
Repository Root: file://<absolute_path>/svn_repo_14_01_2010
```

```
Repository UUID: 25ceb265-6a78-4dce-b758-64b437aadf78
```

```
Revision: 16942
```

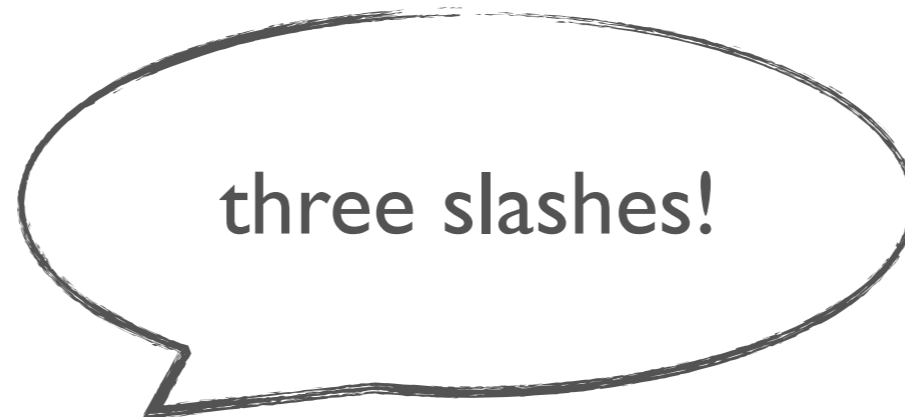
```
Node Kind: directory
```

```
Last Changed Author: ezust
```

```
Last Changed Rev: 16942
```

```
Last Changed Date: 2010-01-17 06:05:58 +0100 (Sun, 17 Jan 2010)
```

# SVN tarball



```
pearl:~ kim$ svn info file:///<absolute_path>/svn_repo_14_01_2010/
```

```
Path: svn_repo_14_01_2010
```

```
URL: file:///<absolute_path>/svn_repo_14_01_2010
```

```
Repository Root: file:///<absolute_path>/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)
```

# Challenges

- Parsing XML
- Full text instead of short messages
  - Many more false positives
- Multiple authors, many comments
  - Order of comments is relevant
- Error Handling

# Possibilities

- Using RegExp
- Using similarity between text
- Order the comments by time
- Comments are very likely to be later than patch



# Programming Language

# Programming Language

**We don't care!**

# Programming Language

**We don't care!**

But: It has to run on CIP pool computers!

# Handing In



- A CSV file containing pairs `<transaction_id, bug_id>`
  - Which bug has been fixed in which transaction?
- A CSV file containing pairs `<svn_path, #bugs>`
  - How many bugs could be found in this file?
- Executable program or script (CIP-pool)

# Handing In (2)



- Documentation

- ▶ What have you done?
- ▶ How can we reproduce the result?
- ▶ What's good, what's bad regarding your approach?

*"It's better to be vaguely  
precise than precisely wrong!"*

Clem Sunter (ICSE 2010)

# Questions?



# What do we Expect?



*"Show that you understood the problem and that you are capable of solving it!"*

# Why do we Make you do that?

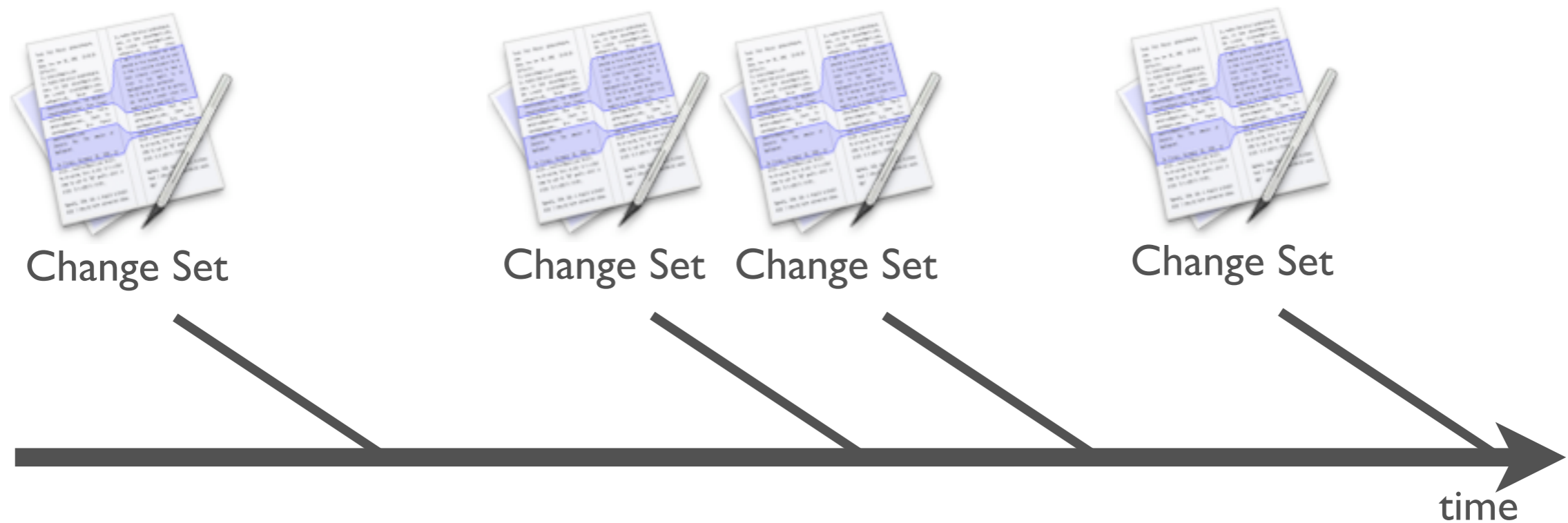
*"Mining is a technical challenge.  
You need to think first but  
finally you have to do it."*

# Bachelor/Master - Theses

- Improve Bug-Mapping techniques
- Decomposing Change Sets
- Trend Mining
- Project Similarity for Prediction



# Solving a General Problem



# Solving a General Problem

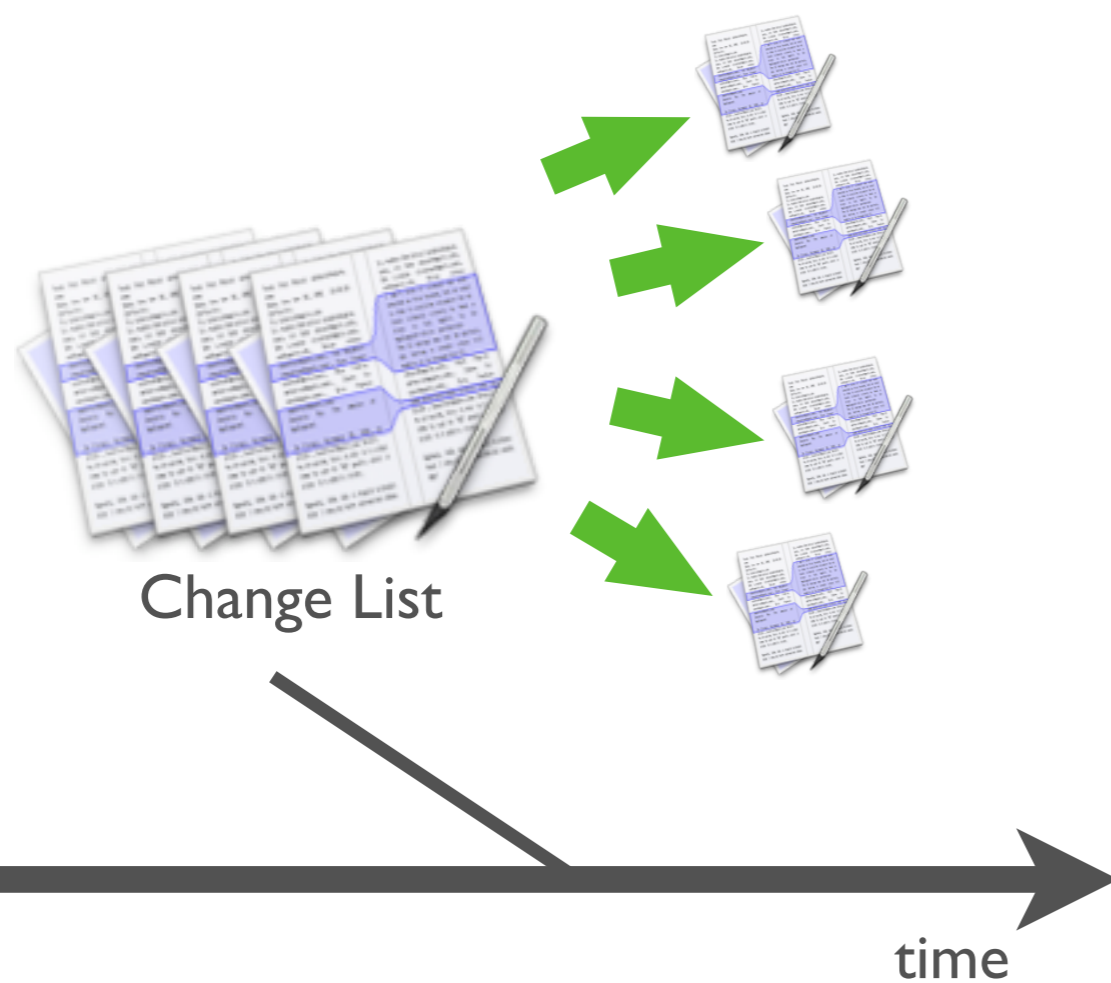


Change List



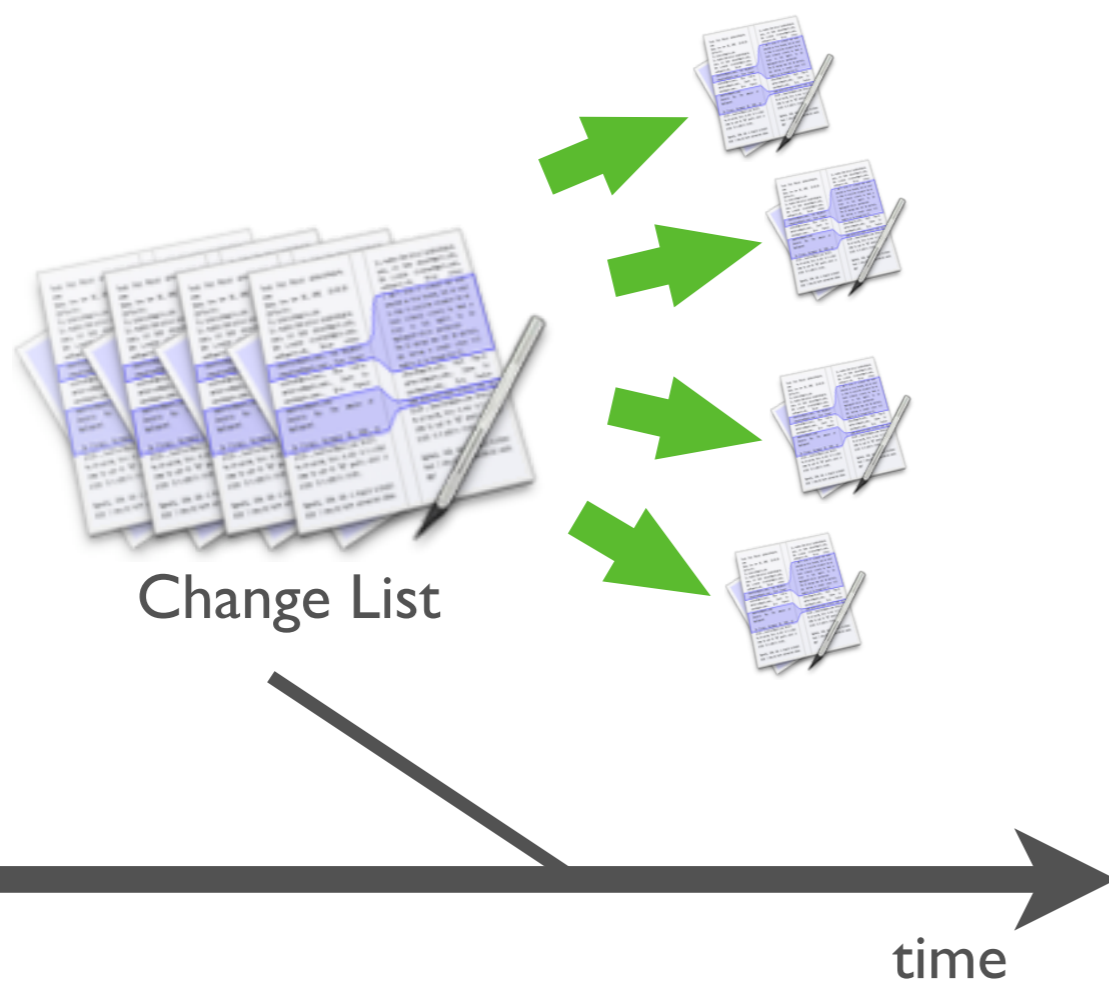
time

# Solving a General Problem



Can we separate  
these change sets  
again?

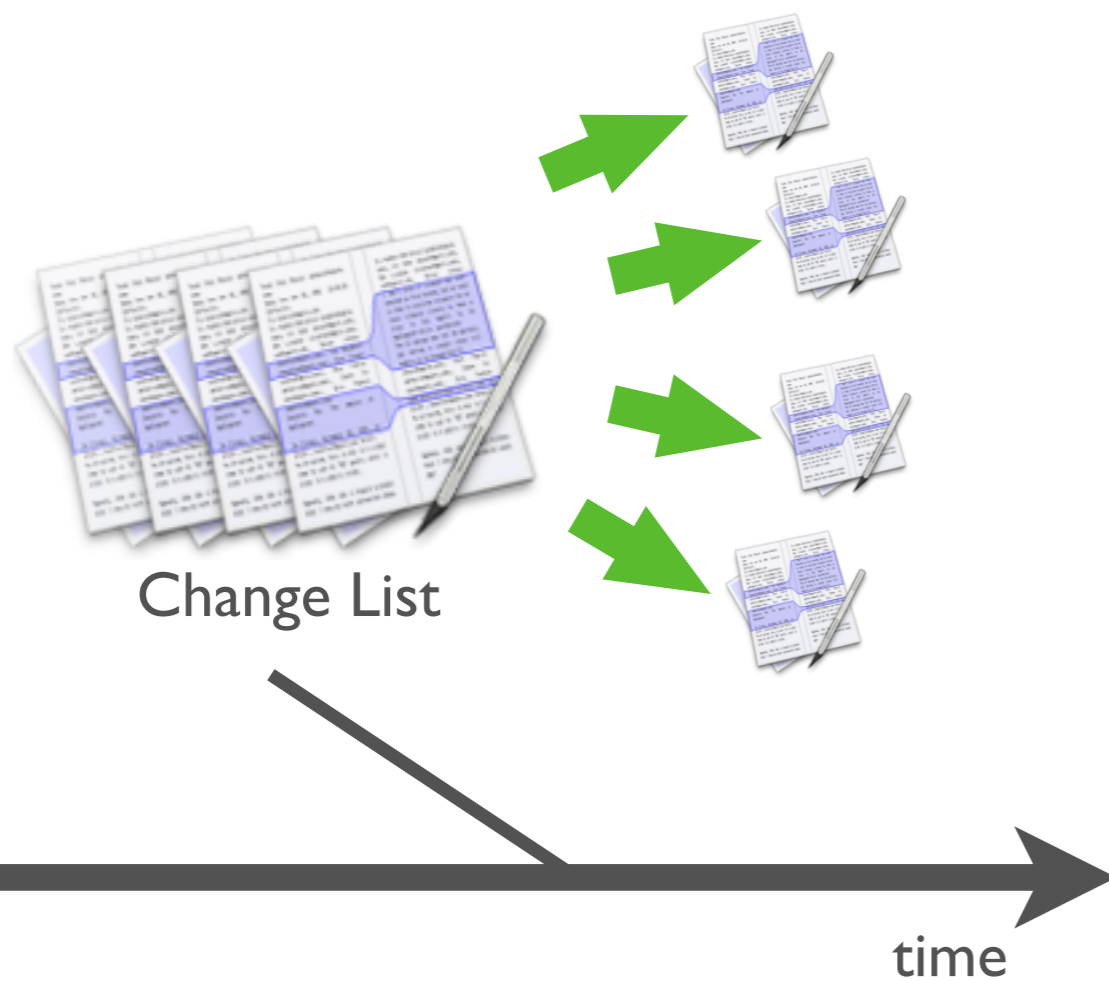
# Solving a General Problem



Can we separate  
these change sets  
again?

- Call graphs
- Dependency data
- Test data
- DeltaDebugging
- ...

# Solving a Google™ Problem



Can we separate  
these change sets  
again?

- Call graphs
- Dependency data
- Test data
- DeltaDebugging
- ...



*You did it! And now?*

*Which change fixed the bug and  
which one added the feature?*

# Change Set Classification

- **keywords** do not work
  - ▶ best predictors are: “should” and “wtf”
- changes visible in **API?**
  - ▶ reverse engineering
- **Test** suites
- **Dependencies** to previous changes

*You did that too?*





Monday, 17 May, 2010





# 33rd International Conference on Software Engineering

Waikiki, Honolulu, Hawaii  
May 21-28, 2011



## MSR 2011

Mining Software Repositories





# 33rd International Conference on Software Engineering

Waikiki, Honolulu, Hawaii  
May 21-28, 2011



## MSR 2011

Mining Software Repositories

