

A SHORT INTRODUCTION TO

MINING

SOFTWARE

REPOSITORIES

Romain Robbes

**Mining the past evolution of software systems
allows us to assist their current evolution**

\$270,000,000

The version editor

```
String FindSource(String base, String dir) {
    DIR * dirp = opendir(dir);
    String result;    // The filename, if found
    for (int i = 0; i < NS; ++i) {        // Loop over suffix list
        String tmp = base + suffix[i];    // Target name to find
        for (dirent *de = readdir(dirp); de != NULL; de = readdir(dirp))
            if (tmp == de->d_name) {      // We found it, stop looking
                result = tmp;
                break;
                return tmp;
            }
        rewinddir(dirp);
    }
    closedir(dirp);
    return result; // Return the found name (may be null)
    return "";    // No match was found
}

Deleted by MR 595 by vz,97/11/15,approved [Stop source search at 1st match]
MR 467 by dla,97/09/21,integrated [Find source using list of suffixes]
"findsource.c", line 15 of 23
```

ve view (recent additions bold, deletions underlined)

We estimated a savings of \$270 million dollars over the ten year period (and that was in 1999, so there have been significant savings since then).

- David Atkins

What is

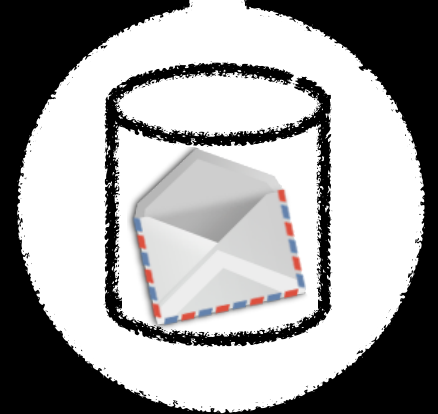
a software

repository?

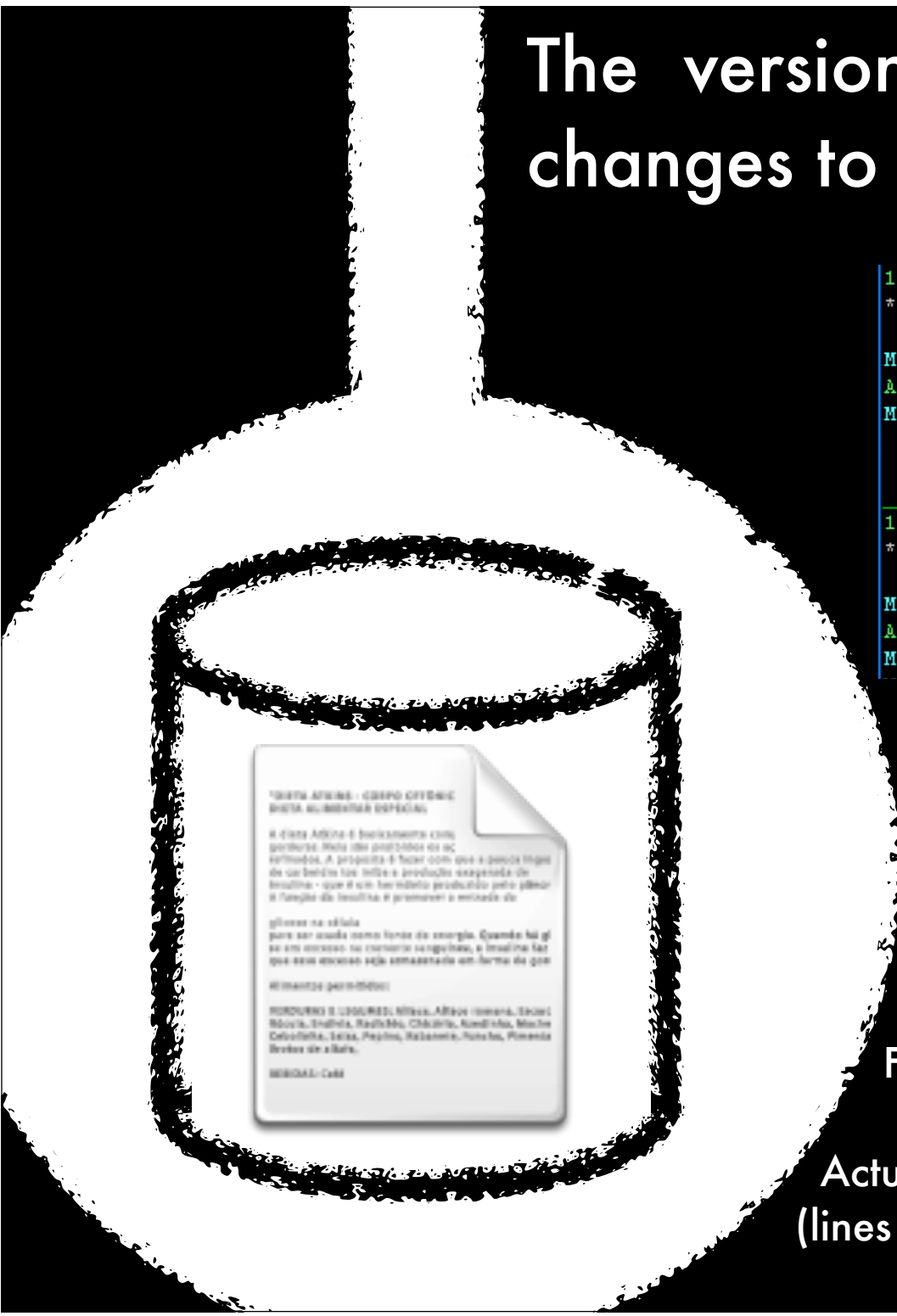
versioning system

issue tracking system

mailing lists



The versioning system stores all the changes to the source code



```
18. r2852 | tyler | 2007-04-30 17:18:39
* guard_method: Made sure that we set the guard variable back t

M gemables/qualitysmith_extensions/Rakefile
A gemables/qualitysmith_extensions/doc_include/ReleaseNotes-0.0
M gemables/qualitysmith_extensions/lib/qualitysmith_extensions/

17. r2850 | tyler | 2007-04-30 16:44:08
* Changed Module#bool_attr_accessor to allow a state of nil

M gemables/qualitysmith_extensions/Rakefile
A gemables/qualitysmith_extensions/doc_include/ReleaseNotes-0.0
M gemables/qualitysmith_extensions/lib/qualitysmith_extensions/
```

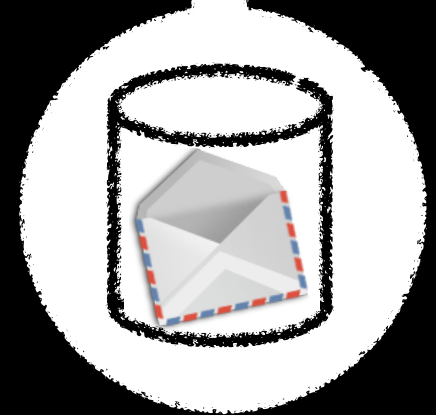
- Revision number
- Author
- Date
- Comment
- Files changed

Actual changes
(lines added/deleted)

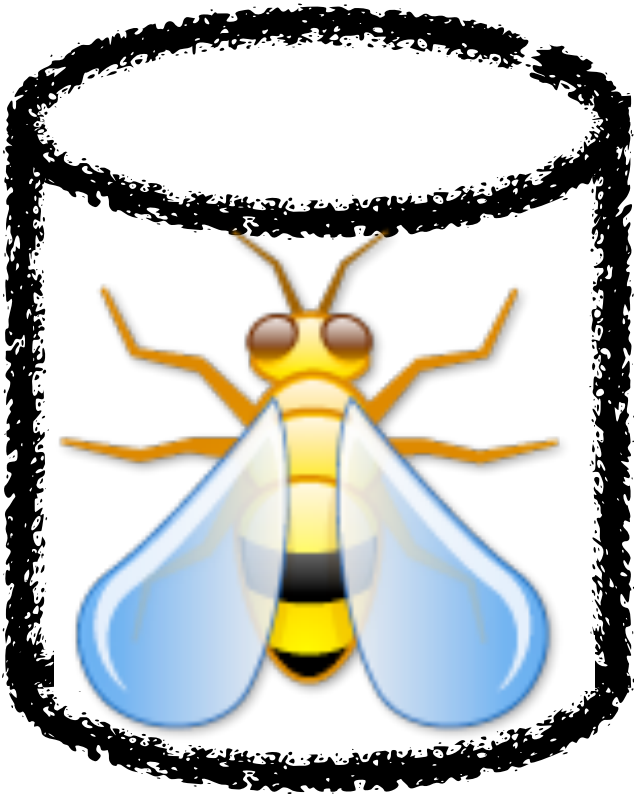
versioning system ✓

issue tracking system

mailing lists



The issue tracking systems stores all the information concerning bugs



ID:

Product and Component:

Status and Resolution:

Assigned To:

*Keywords:

Summary:

Version:

Priority:

Severity:

*Target:

Reporter:

Attachments:

*Dependencies:

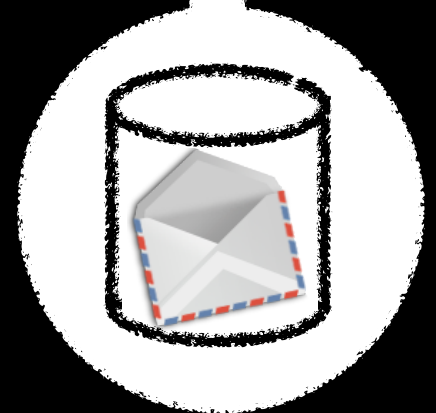
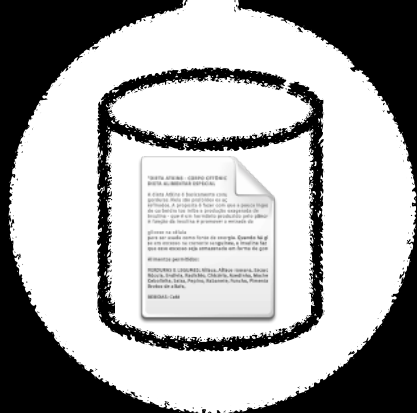
Platform and OS:

Additional Comments:

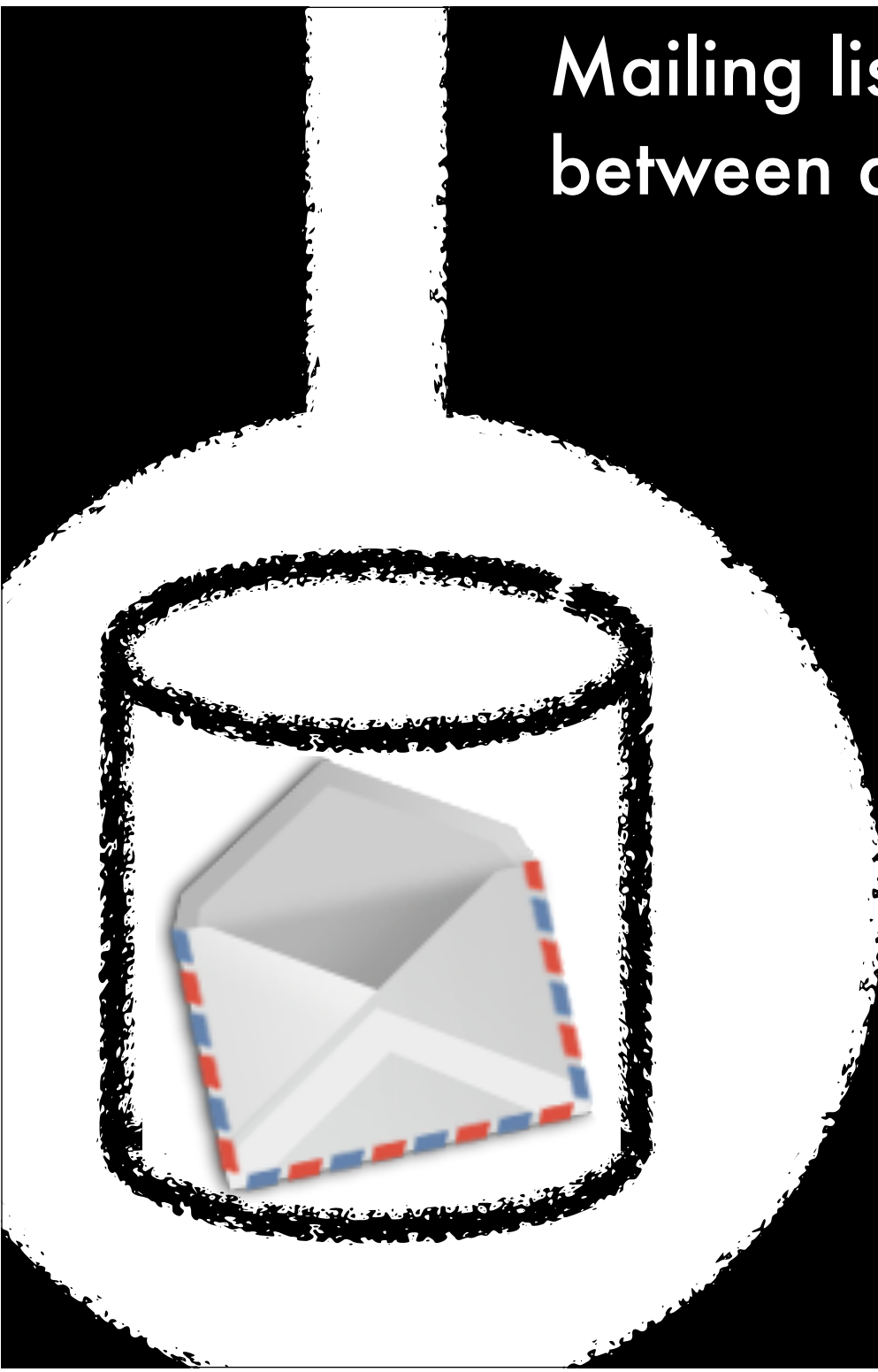
versioning system ✓

issue tracking system ✓

mailing lists



Mailing lists store all the discussions between developers and the users



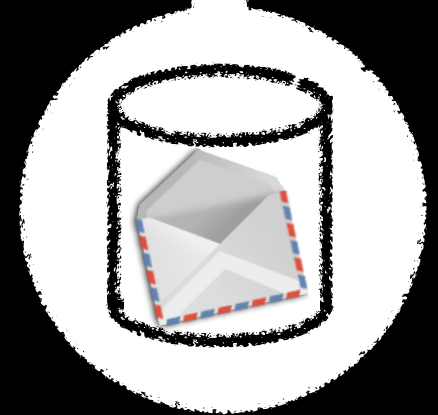
```
Return-Path: <bogdan@fx.ro>
Received: from srv01.advenzia.com (root@localhost)
  by emailaddressmanager.com (8.11.6/8.11.6) with ESMTTP id i2OApwQ14083
  for <support@emailaddressmanager.com>; Wed, 24 Mar 2004 10:51:58 GMT
X-ClientAddr: 193.231.208.29
Received: from corporate.fx.ro (corporate.fx.ro [193.231.208.29])
  by srv01.advenzia.com (8.11.6/8.11.6) with ESMTTP id i2OApvs14078
  for <support@emailaddressmanager.com>; Wed, 24 Mar 2004 10:51:57 GMT
Received: from mail.fx.ro (mail3.fx.ro [193.231.208.3])
  by corporate.fx.ro (8.12.11/8.12.7) with ESMTTP id i2OAtxBt025924
  for <support@emailaddressmanager.com>; Wed, 24 Mar 2004 12:55:59 +0200
Received: from localhost.localdomain (corporate2.fx.ro [193.231.208.28])
  by mail.fx.ro (8.12.11/8.12.3) with ESMTTP id i2OAtQe006624
  for <support@emailaddressmanager.com>; Wed, 24 Mar 2004 12:55:50 +0200
Date: Wed, 24 Mar 2004 12:55:50 +0200
Message-Id: <200403241055.i2OAtQe006624@mail.fx.ro>
Content-Disposition: inline
Content-Transfer-Encoding: binary
MIME-Version: 1.0
To: support@emailaddressmanager.com
Subject: How to read email headers
From: bogdan@fx.ro
Reply-To: bogdan@fx.ro
Content-Type: text/plain; charset=us-ascii
X-Originating-Ip: [80.97.5.101]
X-Mailer: FX Webmail webmail.fx.ro
X-RAVMilter-Version: 8.4.3(snapshot 20030212) (mail)
Status:
```

+contents, attachments

versioning system ✓

issue tracking system ✓

mailing lists ✓



Why

do we

mine?

Does code decay?

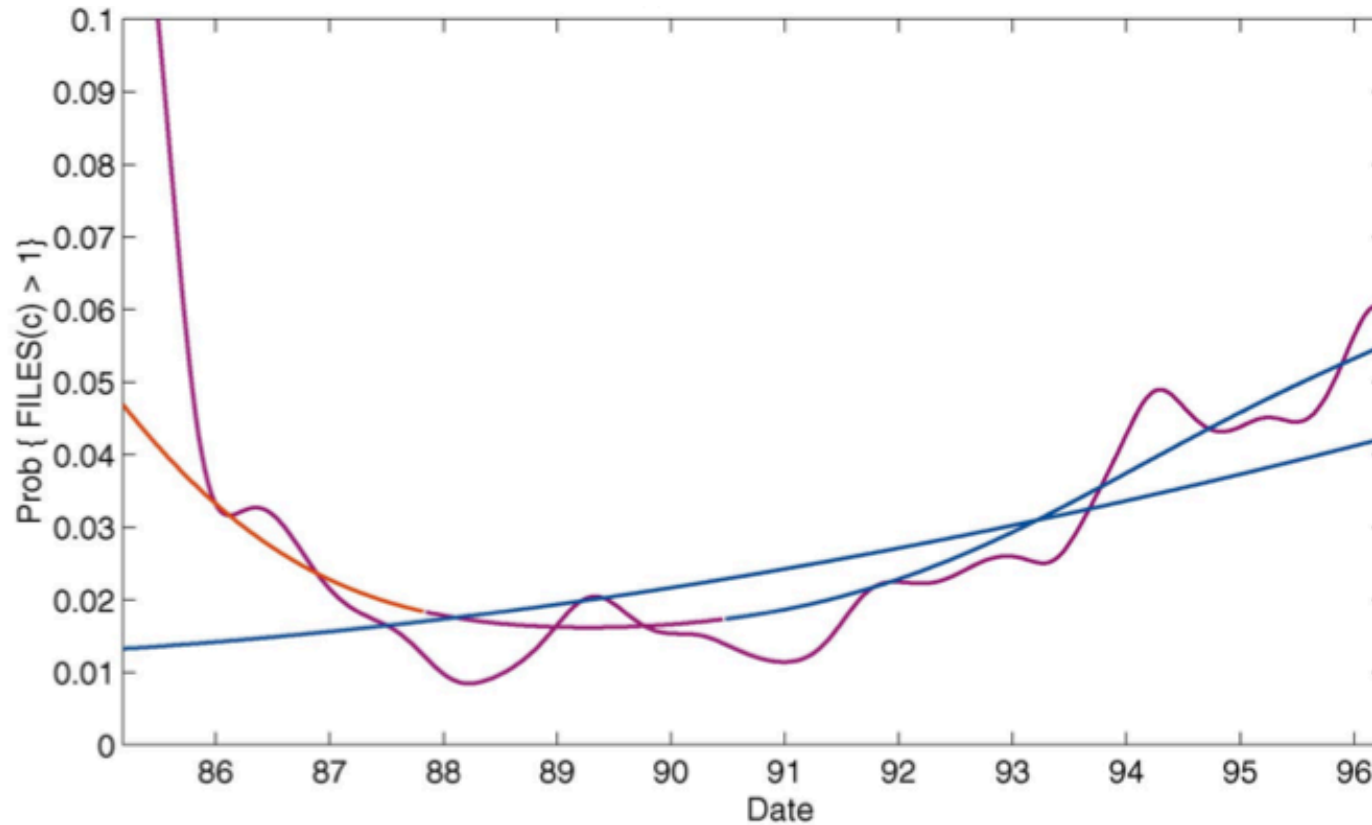
[Eick et al., 2001]

“The quality of software systems will appear to be declining unless they are rigorously maintained and adapted to changes.”

– Lehman’s 7th law of software evolution

Does code decay? Yes!

[Eick et al., 2001]



#files changed for each MR



Is it safe to change this function?

[Sliwerski et al., 2005]

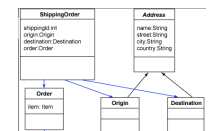
The screenshot shows the Eclipse IDE interface with the file `StandardSourcePathProvider.java` open. The code is as follows:

```
    } else {  
        // recover persisted source path  
        entries = recoverRuntimePath(configuration, IJavaLaunchConfigurationConstants.  
    }  
    return entries;  
}  
  
/* (non-Javadoc) */  
public IRuntimeClasspathEntry[] resolveClasspath(IRuntimeClasspathEntry[] entries, IL  
List all = new ArrayList(entries.length);  
for (int i = 0; i < entries.length; i++) {  
    switch (entries[i].getType()) {  
        case IRuntimeClasspathEntry.PROJECT:  
            // a project resolves to itself for source lookup (rather than the cl  
            all.add(entries[i]);  
            break;  
        case IRuntimeClasspathEntry.OTHER:  
            IRuntimeClasspathEntry2 entry = (IRuntimeClasspathEntry2)entries[i];  
            String typeId = entry.getTypeId();  
            IRuntimeClasspathEntry[] res = null;  
            if (typeId.equals(DefaultProjectClasspathEntry.TYPE_ID)) {  
                // add the resolved children of the project  
                IRuntimeClasspathEntry[] children = entry.getRuntimeClasspathEnt  
                res = JavaRuntime.resolveSourceLookupPath(children, configuration
```

Two annotations are present on the left side of the code editor:

- A **low risk** annotation with a black arrow pointing to the `recoverRuntimePath` call.
- A **high risk** annotation with a black arrow pointing to the `resolveClasspath` method signature.

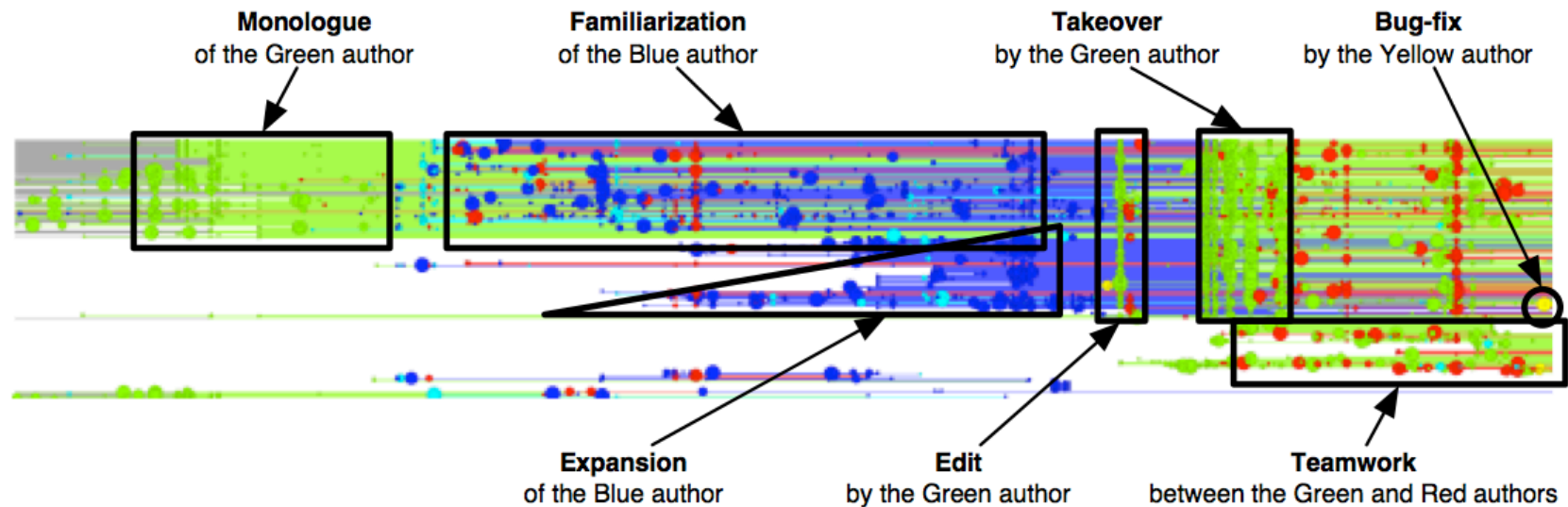
“[...] all changes to `resolveClasspath()` resulted in later fixes.”



Who knows about this? Who should fix it?

[Gîrba et al., 2005]

[Anvik et al., 2006]



“We have reached precision levels [on bug assignment] of 57% and 64% on the Eclipse and Firefox development projects respectively.”



Programmers who changed X also changed ...

[Zimmermann et al., 2004]

[Gall et al., 1998]

The screenshot shows the Eclipse IDE with the following components:

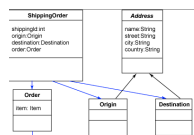
- Package Explorer:** Shows a project structure with files like `CompareEditorContributor.java`, `CompareMessages.java`, `CompareNavigator.java`, `ComparePreferencePage.java`, and `import declarations`. A new preference `NEW_PREFERENCE` is visible in the list.
- Code Editor:** Displays the source code for `ComparePreferencePage.java`. A new line of code is inserted into the `fKeys` array: `new OverlayPreferenceStore.OverlayKey(OverlayPreferenceStore.BOOLEAN, NEW_PREFERENCE),`. The `initDefaults` method is also visible, showing default values for various preferences.
- Related Changes:** A table showing the impact of the change on other parts of the codebase.

Symbol	File	Support	Confidence
<code>initDefaults(IPreferenceStore store)</code>	<code>ComparePreferencePage.java</code>	8	1.0
<code>org.eclipse.compare.plugin.properties</code>	<code>plugin.properties</code>	7	0.875
<code>org.eclipse.compare/buildnotes_compare.html</code>	<code>buildnotes_compare.html</code>	6	0.75
<code>TextMergeViewer(Composite parent, int style, CompareConfiguration configuration)</code>	<code>TextMergeViewer.java</code>	6	0.75
<code>propertyChange(PropertyChangeEvent event)</code>	<code>TextMergeViewer.java</code>	6	0.75
<code>createGeneralPage(Composite parent)</code>	<code>ComparePreferencePage.java</code>	5	0.625
<code>createTextComparePage(Composite parent)</code>	<code>ComparePreferencePage.java</code>	5	0.625
<code>handleDispose(DisposeEvent event)</code>	<code>TextMergeViewer.java</code>	4	0.5

A) The user inserts a new preference into the field fKeys[]

B) ROSE suggests locations for further changes, e.g. the function `initDefaults()`

Change coupling uncovers potentially hidden relationships



Honorable mentions

Refactoring Detection	[Weißgerber et al., 2006] [Dig et al., 2006]
Refactoring-aware evolution	[Henkel et al., 2003][Dig et al., 2007]
Clone analysis	[Kim et al., 2005]
Class hierarchy analysis	[Gîrba et al., 2005]
Framework evolution	[Dagenais et al., 2008][Schäfer et al., 2008]
Package repositories analysis	[Gonzalez-Barahona et al., 2009]
Automatic code formatting	[Reiss, 2007]
Concern mining	[Breu et al., 2006]
Licensing Issues	[Germàn et al., 2009]
...	[REVEAL, 2004-20??]



What

about

you?

Actions

[RSS feed](#)
[Register Member](#)
[Register Group](#)
[Register Project](#)

Authentication

[Login](#)

Home

Welcome to SqueakSource, the smart Monticello code-repository for [Squeak](#) and [Pharo](#). To get started register your personal account and create any number of projects to host on our server. You'll immediately get all the necessary permissions to create and manage your account, projects and versions. Detailed instructions can be found on the [Help](#) page.

This service is brought to you for free: it is backed up daily. Still, please make sure that you have your own proper backups and note that we cannot give any accessibility guarantees. However, if you wish you can set up your own SqueakSource server. It is self-hosted and can be downloaded from [SqueakMap](#).

Please report any problems or suggestions to the [Squeak Mailing-List](#) or the [Pharo Mailing-List](#). Enjoy!

We gratefully acknowledge the support of the [Software Composition Group](#) and the [University of Bern](#) for hosting this service.

Statistics

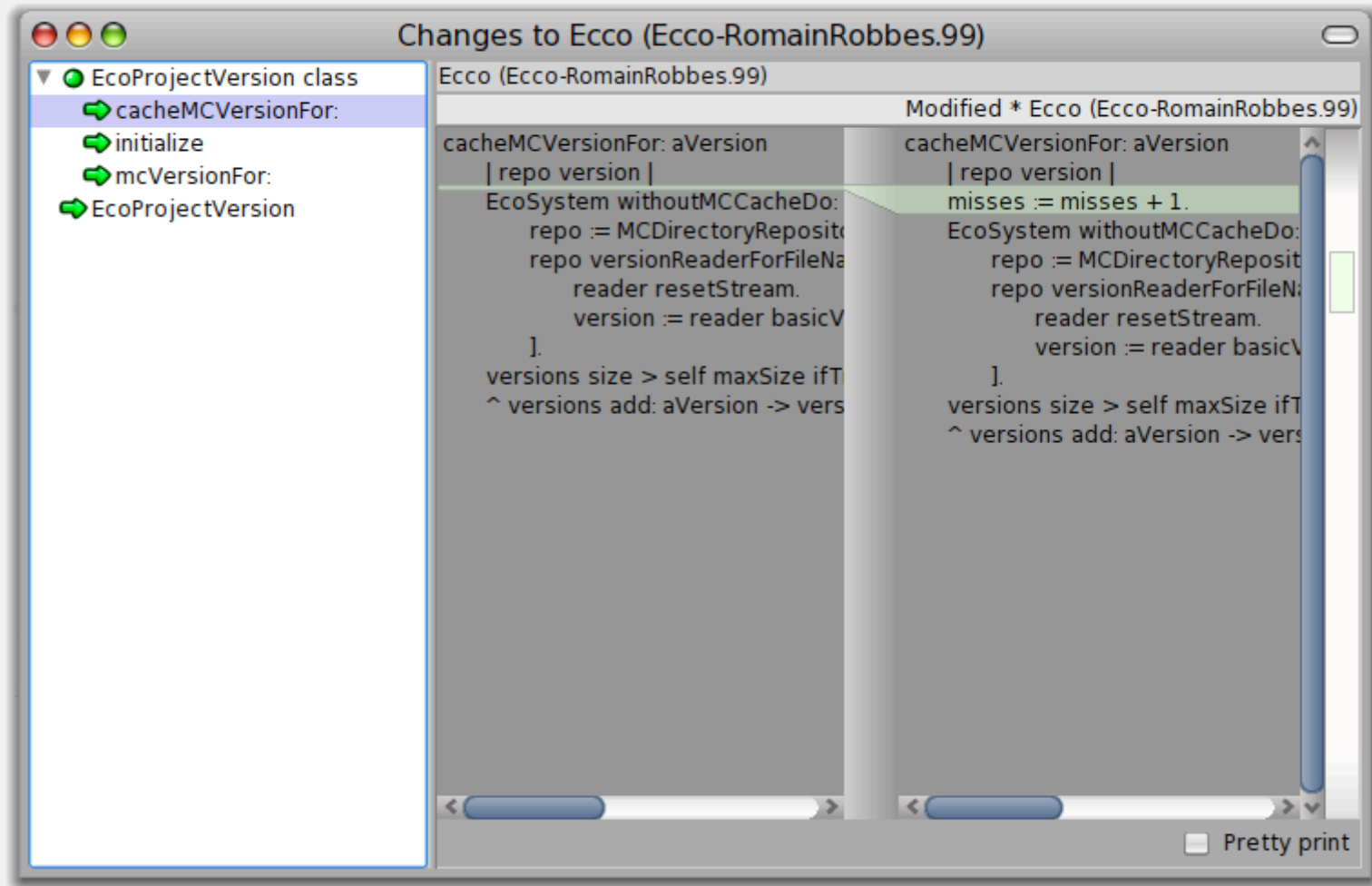
Members (2725)

Recently Joined: [Brian Rogers](#), [Ming Xun Huang](#), [Andreas Pauley](#), [Nicolas Paez](#), ...
Most Active: [Fernando Olivero](#), [Romain Robbes](#), [Tudor Girba](#), [Toon Verwaest](#), ...

Projects (2261)

Recently Created: [PBE Lights Out Game](#), [Quinto done light](#), [Tesseract Trainer](#), [ST-Image](#), ...
Recently Used: [PharoInbox](#), [PharoInbox](#), [PharoInbox](#), [PharoInbox](#), ...
Most Active: [Gaucho](#), [PinocchioVM](#), [Ecco](#), [Pharo](#), ...
Most Downloads: [Pharo](#), [Seaside](#), [Seaside 3.0](#), [Magma Tester](#), ...

Monticello = SVN for Pharo ...

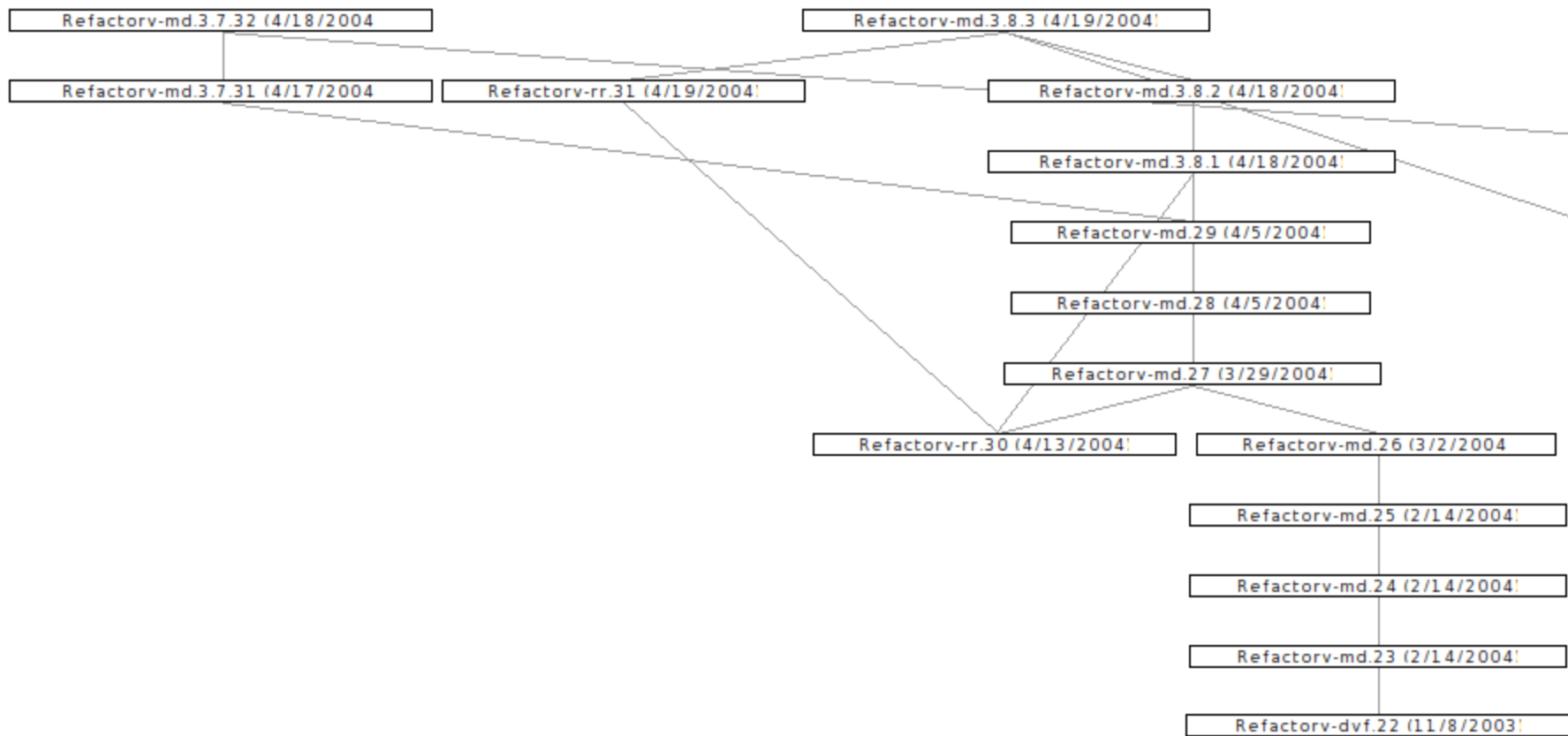


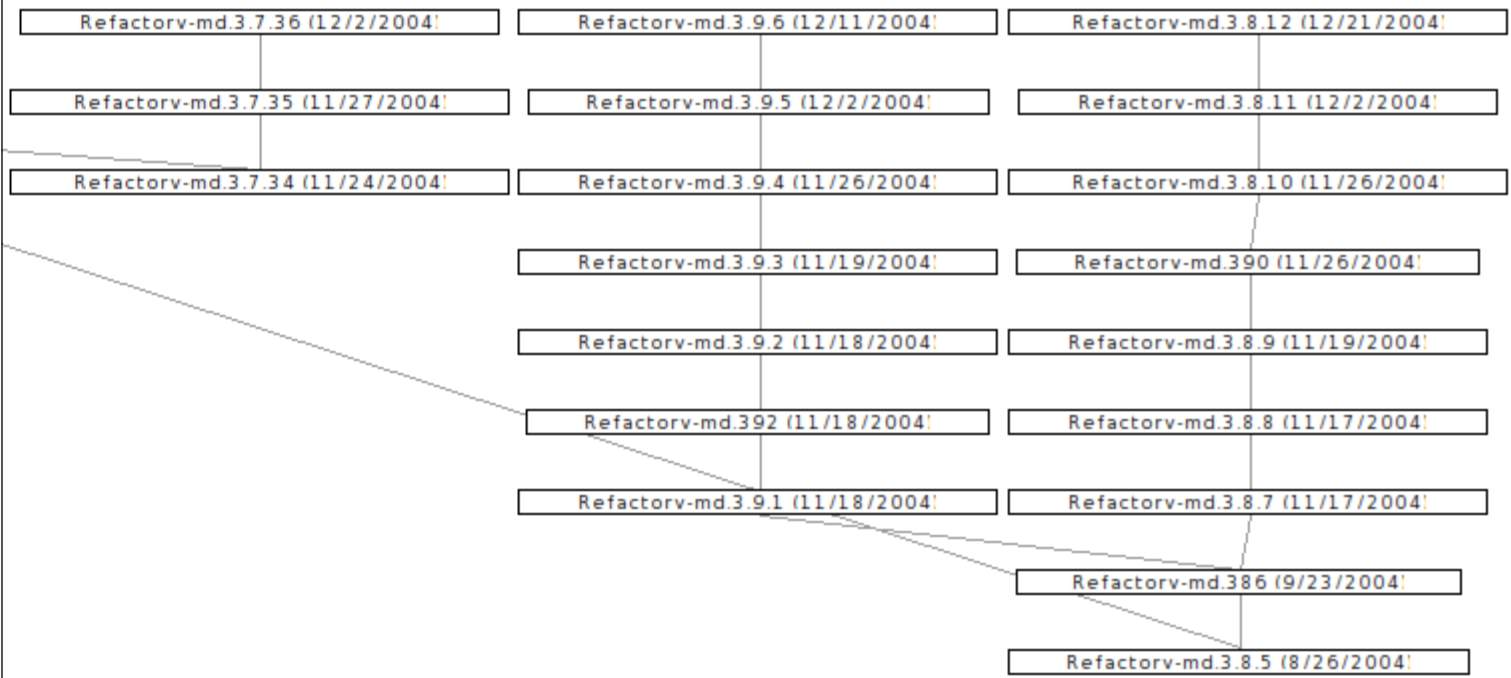
... but better!

Refactory, 2003 - 2004

34 versions

200+ classes





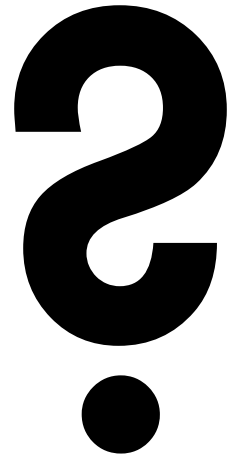
Squeaksource, 2003 – 2010

2,500+ developers

2,000+ projects

50,000+ versions

50,000+ classes





That's

all

folks!