

Software Analytics: Opportunities and Challenges

Olga Baysal

School of Computer Science
Carleton University

olga.baysal@carleton.ca

olgabaysal.com

@olgabaysal

Latifa Guerrouj

Département de GL et des TI
École de Technologie Supérieure

latifa.guerrouj@etsmtl.ca

latifaguerrouj.ca

@Latifa_Guerrouj



Development (Big) Data



Mining Software Repositories

Detection and analysis of hidden patterns and trends



Code Tests Build/Config Version Control

Source code information

Usage data

Docs Issues Mailing lists

Other artifacts

Crash repos Field logs

Runtime data

Problem

Development artifacts



Stakeholders



Data Is NOT Actionable

Decisions Drive Development!

Product deadlines
Are we ready to release?

Program defects
How do I fix this bug?

Risks, cost, operation,
planning
When do we release?

Release engineers Developers

Program correctness
How effective is our
test suite?

Managers



QA

Decisions are often based on intuition or experience

Solution – Analytics



Analytics In Industry



Software Analytics



Benefits of Software Analytics



Supporting Development Decisions



Data-driven decision making, fact-based views of projects

Software Artifacts

- Source Code
- Execution Trace
- Development History
- Bug Reports
- Code Reviews
- Developer Activities
- Software Forums
- Software Microblogs

Artifact: Source Code

- Various languages
- Various kinds of systems
- Various scale: small, medium, large
- Various complexities
- Various programming styles

Artifact: Source Code

The image displays two windows side-by-side. The left window is a Notepad application titled "Untitled - Notepad" with a menu bar (File, Edit, Format, View, Help). It contains the following Java code:

```
// The smallest Java program possible
public class HelloWorld {
    public static void main(String[] args) {
        System.out.println("Hello World");
    }
}
```

The right window is the Eclipse IDE titled "Java - BankAccountTests.java - Eclipse SDK". It shows a project named "Banking" with a package "org.eclipse.banking". The package contains a "BankAccount" class and a "BankAccountTests.java" file. The "BankAccount" class has a "balance" attribute and methods "deposit(BigDecimal)", "getBalance()", and "withdraw(BigDecimal)". The "BankAccountTests.java" file contains the following code:

```
package org.eclipse.banking.tests;

import java.math.BigDecimal;

public class BankAccountTests {

    public void test() {
        BankAccount account = new BankAccount();
        account.deposit(new BigDecimal("100"));
        assertEquals("Balance should be 100", account.getBalance(), new BigDecimal("100"));
    }

    public void testWithdraw() {
        BankAccount account = new BankAccount();
        account.deposit(new BigDecimal("100"));
        account.withdraw(new BigDecimal("50"));
        assertEquals("Balance should be 50", account.getBalance(), new BigDecimal("50"));
    }
}
```

Artifact: Source Code

- Where to find code?
 - GitHub: <https://github.com/>
 - Google code: <http://code.google.com/>
 - Many other places online
- How to analyze source code?
 - Program analysis tools
 - WALA: <http://wala.sourceforge.net>
 - JPF: <http://javapathfinder.sourceforge.net/>
 - Soot: <http://sable.github.io/soot/>
 - Clang: <http://clang-analyzer.llvm.org/>

Artifact: Execution Traces

- Information collected when a program is run
- What kind of information is collected?
 - Sequences of methods that are executed
 - State of various variables at various times
 - State of various invariants at various times
 - Which components are loaded at various times

Artifact: Execution Traces

```
nu.fw.jeti.backend.JabberHandler&14716637|nu.fw.jeti.backend.Jabber&12532667|nu.fw.jeti.  
nu.fw.jeti.backend.JabberHandler&14716637|nu.fw.jeti.backend.Jabber&12532667|nu.fw.jeti.  
nu.fw.jeti.backend.JabberHandler&14716637|nu.fw.jeti.backend.Jabber&12532667|nu.fw.jeti.  
nu.fw.jeti.backend.JabberHandler&14716637|nu.fw.jeti.backend.Jabber&12532667|nu.fw.jeti.  
nu.fw.jeti.backend.JabberHandler&14716637|nu.fw.jeti.backend.Jabber&12532667|nu.fw.jeti.  
nu.fw.jeti.backend.JabberHandler&14716637|nu.fw.jeti.backend.Jabber&12532667|nu.fw.jeti.  
nu.fw.jeti.backend.JabberHandler&14716637|nu.fw.jeti.backend.Jabber&12532667|nu.fw.jeti.  
nu.fw.jeti.backend.Jabber&12532667|nu.fw.jeti.ui.ChatWindows&33406487|nu.fw.jeti.ui.Chat  
nu.fw.jeti.ui.ChatWindows&33406487|nu.fw.jeti.ui.ChatWindows&33406487|nu.fw.jeti.ui.Chat  
nu.fw.jeti.backend.JabberHandler&14716637|nu.fw.jeti.backend.Jabber&12532667|nu.fw.jeti.  
nu.fw.jeti.backend.Jabber&12532667|nu.fw.jeti.ui.ChatWindows&33406487|nu.fw.jeti.ui.Chat  
nu.fw.jeti.ui.ChatWindows&33406487|nu.fw.jeti.ui.ChatWindows&33406487|nu.fw.jeti.ui.Chat  
nu.fw.jeti.backend.JabberHandler&14716637|nu.fw.jeti.backend.Jabber&12532667|nu.fw.jeti.  
nu.fw.jeti.backend.JabberHandler&14716637|nu.fw.jeti.backend.Jabber&12532667|nu.fw.jeti.  
nu.fw.jeti.backend.Jabber&12532667|nu.fw.jeti.plugins.drawing.shapes.PictureChat&1507654  
nu.fw.jeti.backend.JabberHandler&14716637|nu.fw.jeti.backend.Jabber&12532667|nu.fw.jeti.  
nu.fw.jeti.backend.Jabber&12532667|nu.fw.jeti.plugins.drawing.shapes.PictureChat&1507654  
nu.fw.jeti.plugins.drawing.shapes.PictureChat$1&16548518|nu.fw.jeti.plugins.drawing.shap  
nu.fw.jeti.plugins.drawing.shapes.PictureHistory&23493887|nu.fw.jeti.plugins.drawing.shap  
nu.fw.jeti.plugins.drawing.shapes.PictureHistory&23493887|nu.fw.jeti.plugins.drawing.ui.  
nu.fw.jeti.plugins.drawing.ui.HistoryPanel&5049504|nu.fw.jeti.plugins.drawing.shapes.act  
nu.fw.jeti.plugins.drawing.shapes.PictureHistory&23493887|nu.fw.jeti.plugins.drawing.shap  
nu.fw.jeti.plugins.drawing.shapes.PictureChat$1&16548518|nu.fw.jeti.plugins.drawing.shap  
nu.fw.jeti.plugins.drawing.shapes.PictureChat$1&7798629|nu.fw.jeti.plugins.drawing.shap  
nu.fw.jeti.plugins.drawing.shapes.PictureHistory&23493887|nu.fw.jeti.plugins.drawing.shap  
nu.fw.jeti.plugins.drawing.shapes.PictureHistory&23493887|nu.fw.jeti.plugins.drawing.shap  
nu.fw.jeti.plugins.drawing.shapes.PictureHistory&23493887|nu.fw.jeti.plugins.drawing.shap  
nu.fw.jeti.plugins.drawing.ui.HistoryPanel&5049504|nu.fw.jeti.plugins.drawing.shapes.act  
nu.fw.jeti.plugins.drawing.ui.HistoryPanel&5049504|nu.fw.jeti.plugins.drawing.shapes.act  
nu.fw.jeti.plugins.drawing.ui.HistoryPanel&5049504|nu.fw.jeti.plugins.drawing.shapes.act  
nu.fw.jeti.plugins.drawing.ui.HistoryPanel&5049504|nu.fw.jeti.plugins.drawing.shapes.act  
nu.fw.jeti.plugins.drawing.ui.HistoryPanel&5049504|nu.fw.jeti.plugins.drawing.shapes.act  
nu.fw.jeti.plugins.drawing.ui.HistoryPanel&5049504|nu.fw.jeti.plugins.drawing.shapes.act
```

Caller | Callee | Method Signature

Artifact: Execution Traces

- How to collect?
 - Insert instrumentation code
 - Execute program
 - Instrumentation code writes a log file
- What tools are available to collect traces?
 - Daikon Chicory: <http://groups.csail.mit.edu/pag/daikon/dist/doc/daikon.html>
 - PIN: <http://software.intel.com/en-us/articles/pin-a-dynamic-binary-instrumentation-tool>
 - Valgrind: <http://valgrind.org/>

Artifact: Development History

- What code is
 - Added
 - Deleted
 - Edited
- When
- By Whom
- For What Reason

Artifact: Development History

The screenshot shows a 'Log Messages' window for the file 'C:\public_html\yrb\index'. The window displays a list of revisions from 48 to 60, all authored by 'victor.stanciu' on June 9, 2011. Revision 60 is the most recent. Below the list is a table of file paths that were changed in the selected revisions. The paths include various PHP files in the application's trunk directory.

From: 6/ 3/2011 To: 6/ 9/2011 Messages, authors and paths

Revision	Actions	Author	Date
60		victor.stanciu	3:02:51 PM, Thursday, June 09, 2011
59		victor.stanciu	2:57:21 PM, Thursday, June 09, 2011
58		victor.stanciu	2:47:23 PM, Thursday, June 09, 2011
57		victor.stanciu	2:20:50 PM, Thursday, June 09, 2011
56		victor.stanciu	1:58:41 PM, Thursday, June 09, 2011
55		victor.stanciu	1:50:10 PM, Thursday, June 09, 2011
54		victor.stanciu	1:47:54 PM, Thursday, June 09, 2011
53		victor.stanciu	11:24:04 AM, Thursday, June 09, 2011
52		victor.stanciu	10:57:57 AM, Thursday, June 09, 2011
51		victor.stanciu	6:58:16 PM, Wednesday, June 08, 2011
50		victor.stanciu	5:51:58 PM, Wednesday, June 08, 2011
49		victor.stanciu	5:46:54 PM, Wednesday, June 08, 2011
48		victor.stanciu	4:44:24 PM, Wednesday, June 08, 2011

Action	Path	Copy from path	Revision
	/trunk/web/application/modules/default/controllers/Page.php		
	/trunk/web/application/modules/default/views/login/register.php		
	/trunk/web/application/modules/default/views/modules/footer.php		
	/trunk/web/application/modules/default/views/modules/navigation.php		

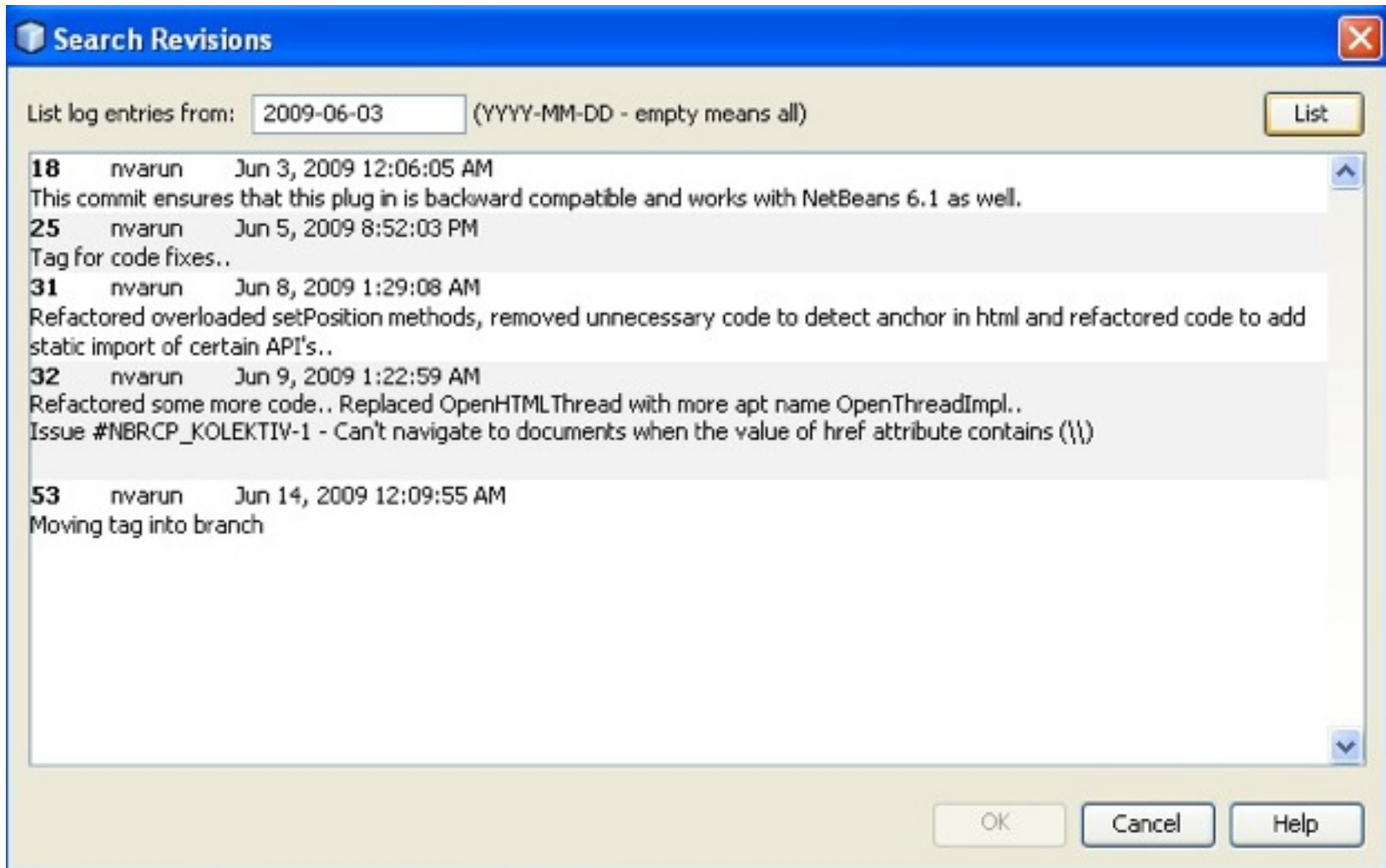
Showing 57 revision(s), from revision 2 to revision 60 - 2 revision(s) selected.

Hide unrelated changed paths Stop on copy/rename Include merged revisions

Buttons: Statistics, Help, OK

Buttons: Show All, Next 100, Refresh

Artifact: Development History



The screenshot shows a 'Search Revisions' dialog box with a search filter set to '2009-06-03'. The list of revisions includes:

- 18** nvarun Jun 3, 2009 12:06:05 AM
This commit ensures that this plug in is backward compatible and works with NetBeans 6.1 as well.
- 25** nvarun Jun 5, 2009 8:52:03 PM
Tag for code fixes..
- 31** nvarun Jun 8, 2009 1:29:08 AM
Refactored overloaded setPosition methods, removed unnecessary code to detect anchor in html and refactored code to add static import of certain API's..
- 32** nvarun Jun 9, 2009 1:22:59 AM
Refactored some more code.. Replaced OpenHTMLThread with more apt name OpenThreadImpl..
Issue #NBRCP_KOLEKTIV-1 - Can't navigate to documents when the value of href attribute contains (\\)
- 53** nvarun Jun 14, 2009 12:09:55 AM
Moving tag into branch

Artifact: Development History

- Various tools
 - CVS – Version per file
 - SVN – Version per snapshot
 - Git – Distributed
- Slightly different ways to manage content

Artifact: Bug Reports

- People report errors and issues that they encounter in the field
- These errors include:
 - Description of the bugs
 - Steps to reproduce the bugs
 - Severity level
 - Parts of the system affected by the bug
 - Failure traces

Artifact: Bug Reports

- Various kinds of bug repositories

- BugZilla: <http://www.bugzilla.org/>

Example site:

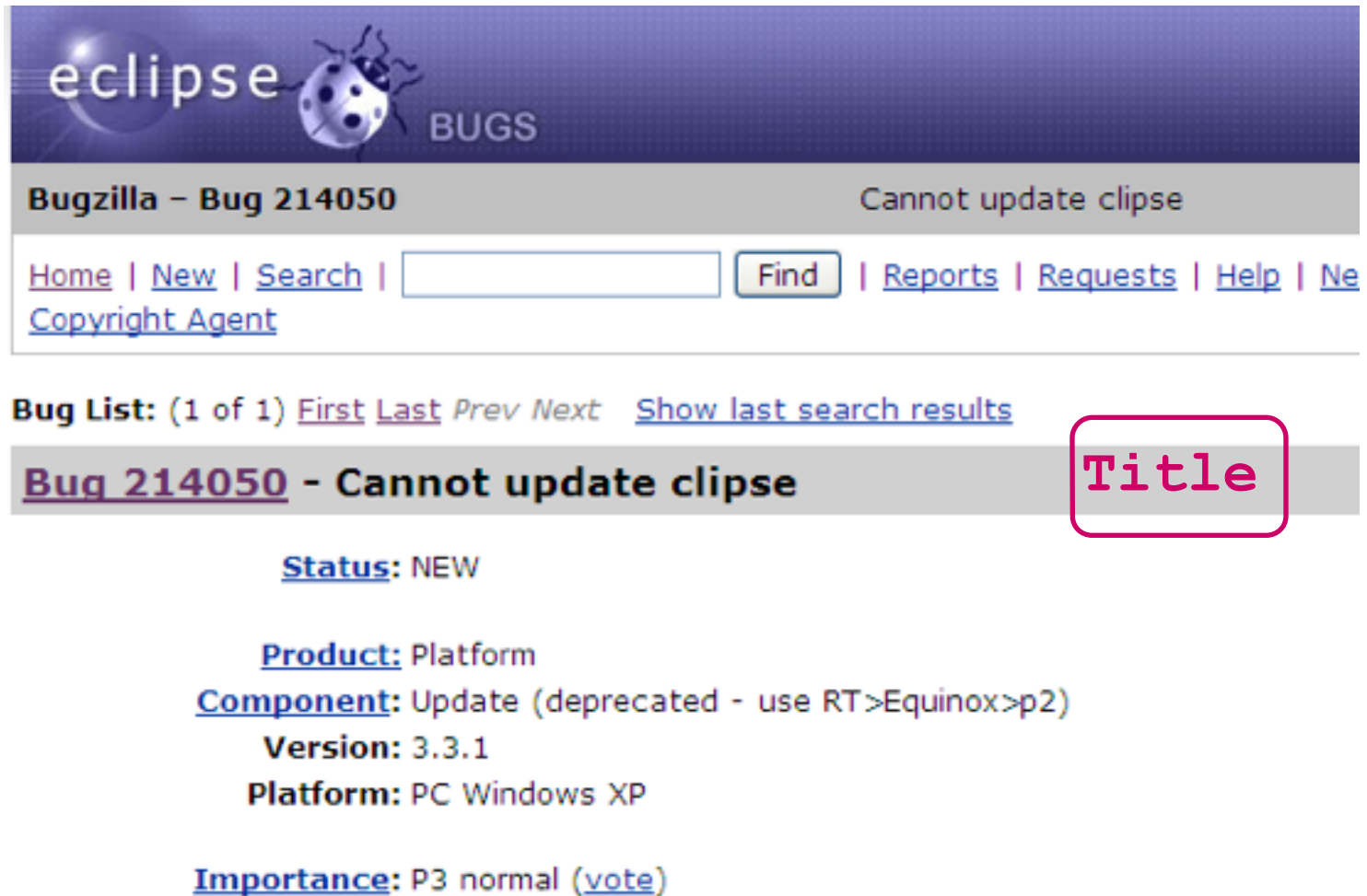
<https://bugzilla.mozilla.org/>

- JIRA: <http://www.atlassian.com/software/jira/>


Example site: <https://issues.apache.org/jira/browse>



Artifact: Bug Reports



The screenshot displays the Eclipse Bugzilla interface. At the top, the Eclipse logo and a ladybug icon are followed by the word 'BUGS'. Below this, the bug title 'Bugzilla - Bug 214050' and the description 'Cannot update clipse' are visible. A navigation bar includes links for 'Home', 'New', 'Search', a search input field, a 'Find' button, 'Reports', 'Requests', 'Help', and 'Ne Copyright Agent'. The 'Bug List' section shows '(1 of 1)' with navigation links 'First', 'Last', 'Prev', 'Next', and 'Show last search results'. The main bug entry is 'Bug 214050 - Cannot update clipse', with the word 'Title' highlighted in a red box. Below the title, the following details are listed: 'Status: NEW', 'Product: Platform', 'Component: Update (deprecated - use RT>Equinox>p2)', 'Version: 3.3.1', 'Platform: PC Windows XP', and 'Importance: P3 normal (vote)'.

eclipse  BUGS

Bugzilla - Bug 214050 Cannot update clipse

[Home](#) | [New](#) | [Search](#) | | [Reports](#) | [Requests](#) | [Help](#) | [Ne Copyright Agent](#)

Bug List: (1 of 1) [First](#) [Last](#) [Prev](#) [Next](#) [Show last search results](#)

Bug 214050 - Cannot update clipse Title

Status: NEW

Product: Platform

Component: Update (deprecated - use RT>Equinox>p2)

Version: 3.3.1

Platform: PC Windows XP

Importance: P3 normal ([vote](#))

Artifact: Bug Reports

Yair Eshel 2008-01-01 07:55:49 EST

Detailed
Description

Build ID: M20071023-1652

Steps To Reproduce:

- 1.Update eclipse 3.3.1.1 from the help menu
- 2.Mark with V Eclipse RCP Patch 1 for 3.3.1.1 3.3.1.1_v20071204_3311, on <http://ftp.osuosl.org/pub/eclipse/eclipse/updates/3.3/site.xml>
- 3.Next until error

More information:

Update operation has failed

Error retrieving

"plugins/com.ibm.icu36.data.update_3.6.1.v20071204_2007j.jar". [Serve
HTTP response code: "403 Forbidden" for URL:

<http://ftp.osuosl.org/pub/eclipse/eclipse/updates/3.3/plugins/com.ibm>

Server returned HTTP response code: "403 Forbidden" for URL:

<http://ftp.osuosl.org/pub/eclipse/eclipse/updates/3.3/plugins/com.ibm>

Running on winxp

Artifact: Code Reviews

ANDROID

open source project

Bro

Search for status:open

Subject	Status	Owner	Project	Branch
QEMU1 changes to make AndroidEmu usable in QEMU2		Yurii Zubrytskyi	platform/external/qemu	emu-master-dev
test comment - do not merge		Bruce J Beare	platform/build	master
MIPS32: int java.lang.*.numberOfLeadingZeros		Chris Larsen	platform/art	master
Add sconsole to system/extras.		Elliott Hughes	platform/system/extras	master
adb/base: fix adb push of Unicode filenames on Win32		Spencer Low	platform/system/core	master
Fix GNU/POSIX basename headers.		Josh Gao	platform/bionic	master

Artifact: Code Reviews

Change [179541](#) - Needs Code-Review

QEMU1 changes to make AndroidEmu usable in QEMU2

Cleaning up the qemu1 build and AndroidEmu to make it easily accessible from Qemu2:

- build has to define paths starting from the root
- AndroidEmu should not include anything from outside world
- Don't define some functions when building for Qemu2 temporary workaround, until we decide which implementation wins

Change-Id: [I819e990a9514233aa5e028093cfdaad9cce5774c](#)

Artifact: Code Reviews

History	Expand All
Yurii Zubrytskyi	Uploaded patch set 1.
Yurii Zubrytskyi	Patch Set 1: Code-Review+1
Vince Harron Patch Set 1: Code-Review+1 (1 comment) android/looper-qemu.cpp Line 19: nit: other places in the code, each namespace gets its own line	
Yurii Zubrytskyi	Uploaded patch set 2.
Yurii Zubrytskyi Patch Set 1: (1 comment) android/looper-qemu.cpp Line 19: fixed	
Yurii Zubrytskyi	Patch Set 2: Code-Review+1 Verified+1 Ping :)

Artifact: Developer Activities

- Developers form a social network
 - Developers work on various projects
 - Projects have various types, programming languages and developers
 - Developers *follow updates from* various other developers and projects
- *A heterogeneous* social network is formed

Artifact: Developer Activities



- **GitHub's CEO, Tom Preston-Werner:**

“We like the ideas of social networking. We think that developers work more effectively when they work together. So let’s take the ideas of a social network and add on top of that code hosting, and let’s create a site that makes it easy to share and collaborate on code”.

Artifact: Developer Activities



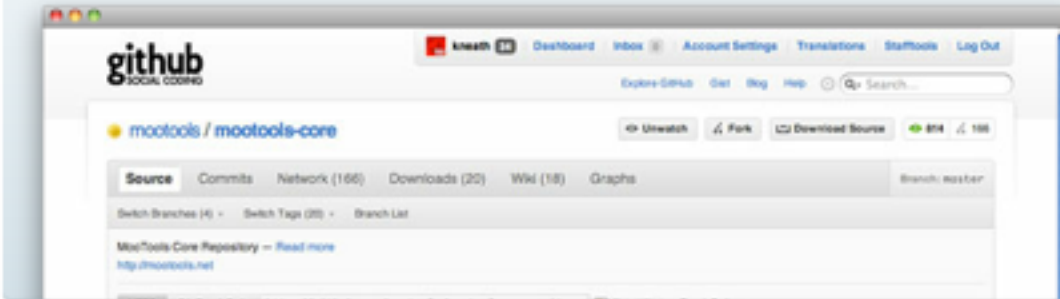
[About GitHub](#)

[Our Team](#)

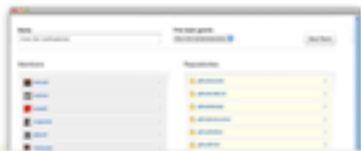
[Press](#)

[Jobs](#)

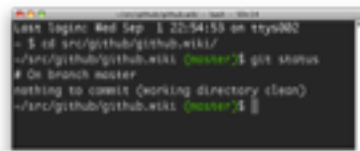
There are more than **12M** people collaborating right now on GitHub on over **31M** projects using a powerful collaborative development workflow.



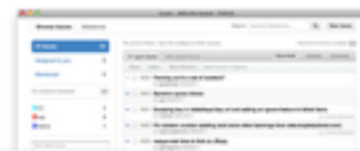
Every repository on GitHub comes with the tools needed to manage your project. Open to the community for public projects – secured for private projects.



Collaboration
manage large teams with ease



Git Powered Wikis
clone, push, branch and tag it



Integrated Issue Tracking
bug tracking tied to commits



Code Review
living discussions about code

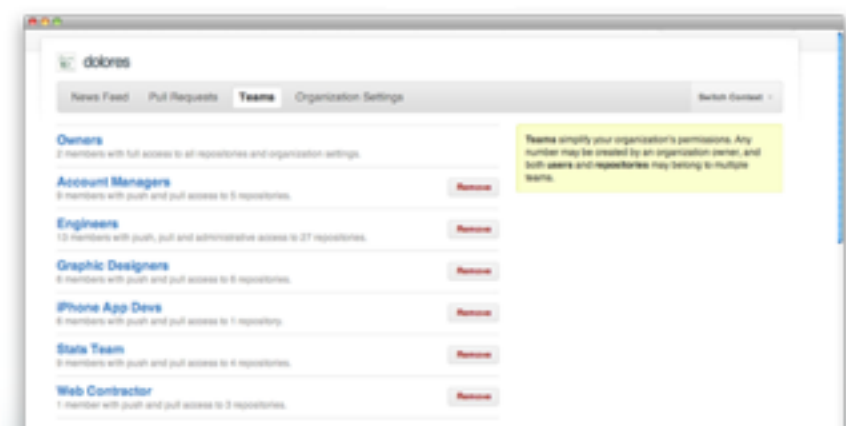
Manage Teams with Organizations

Whether you're running an open source project or a Fortune 500 company, Organizations simplify team management.

With **teams** you can give your developers as much or as little power as they need, from the ability to create projects on behalf of your organization to read-only access on existing projects.

Team permissions: Read-only, read-write, and admin-level access.

Best of all: create **as many** teams with as many members as you need.





Keith Mitchell
kemitche

Lyft
 San Francisco Bay Area
 Joined on Jun 15, 2011

23 Followers
22 Starred
0 Following

[Contributions](#) [Repositories](#) [Public activity](#) [Follow](#)

Popular repositories

l2cs DEPRECATED - Python library to rewrite Luc...	7 ★
reddittweeter Tweeter for reddit JSON feeds	2 ★
indextank-service The API, BackOffice, Storefront, and Nebulize...	2 ★
indextank-engine Indexing engine for IndexTank	1 ★
jenkins-tools Tools used by/with/alongside a jenkins continu...	1 ★

Repositories contributed to

reddit/reddit the code that powers reddit.com	9,770 ★
reddit/reddit-plugin-thebutton The Button.	101 ★
jlafon/PynamoDB A pythonic interface to Amazon's DynamoDB	134 ★
praw-dev/praw PRAW, an acronym for "Python Reddit API W...	869 ★

Public contributions

Summary of pull requests, issues opened, and commits. [Learn how we count contributions.](#) Less More



reddit

San Francisco, CA

<https://www.reddit.com/r/redditdev>

Repositories

People 13

Filters ▾

Find a repository...

reddit

the code that powers reddit.com

Updated 6 hours ago

Python ★ 9,770 1,988

gevent

↳ forked from [gevent/gevent](#)

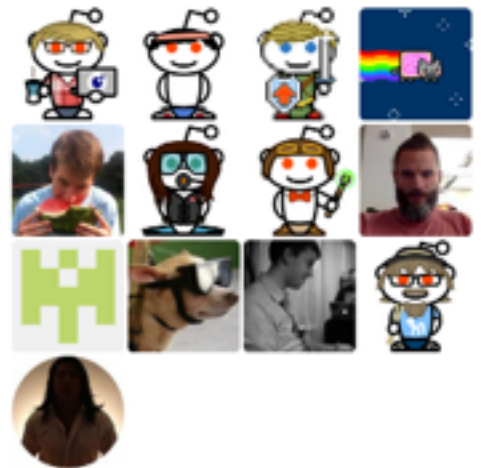
Coroutine-based concurrency library for Python

Updated 6 hours ago

Python ★ 0 523

People

13 >





Jesper Nøhr (jespern)

<http://noehr.org/>

San Francisco, CA

Member since June 2008

Overview

Snippets

Followers 245

Following 44

Teams 7



django-piston



hg-7




git-7

Artifact: Software Forums

- Developers ask and answer questions
- About various topics
- In various threads, some of which are very long
- Stored in various sites
 - StackOverflow: <http://stackoverflow.com/>
 - SoftwareTripsAndTricks: <http://www.softwaretipsandtricks.com/forum/>

Artifact: Software Forums



stackoverflow

Questions Tags Users Badges Unanswered

Top Questions

interesting 280 featured hot week month

0 votes 0 answers 1 view [Check_mk agent based checks for monitoring each jenkins job.](#)
[jenkins](#) [health-monitoring](#) just now [Maverick143](#) 33

0 votes 1 answer 58 views [MimeType is always application/octet-stream](#)
[c#](#) [asp.net](#) [flash](#) [mime-types](#) [uploadify](#) 2s ago [TomTom](#) 24.5k

0 votes 0 answers 1 view [Sys.Webforms.PageRequestManagerServerErrorException server error 503](#)
[c#](#) [asp.net](#) [windows-server-2003](#) [runtime-error](#) 5s ago [vini](#) 482

Artifact: Software Forums

C++ multiple c++ files - ld: symbol(s) not found for architecture x86_64

CAREERS 2.0
by stackoverflow



+



Have projects on BitBucket?
Import them easily to your profile



Hello World!

This is a collaboratively edited question and answer site for **professional and enthusiast programmers**. It's 100% free, no registration required.

[about](#) » [faq](#) »

0
▲
▼
☆

When I put my Stack.cpp into Stack.h it works just fine but, when I separate Stack.h, .cpp files it gives this error. I have also a main.cpp file which does nothing but includes AlgebraicExpression.h I use this command to compile : "g++ -o main main.cpp AlgebraicExpression.cpp Stack.cpp"

```
Undefined symbols for architecture x86_64:
  "Stack<char>::pop()", referenced from:
  infix2postfix(char*) in ccKgnm.o
  "Stack<char>::top()", referenced from:
  infix2postfix(char*) in ccKgnm.o
  "Stack<char>::push(char)", referenced from:
  infix2postfix(char*) in ccKgnm.o
  "Stack<char>::size()", referenced from:
  infix2postfix(char*) in ccKgnm.o
  "Stack<char>::Stack()", referenced from:
  infix2postfix(char*) in ccKgnm.o
  "Stack<double>::pop()", referenced from:
  evaluatePostfix(char*) in ccKgnm.o
  "Stack<double>::top()", referenced from:
  evaluatePostfix(char*) in ccKgnm.o
  "Stack<double>::push(double)", referenced from:
  evaluatePostfix(char*) in ccKgnm.o
  "Stack<double>::Stack()", referenced from:
  evaluatePostfix(char*) in ccKgnm.o
ld: symbol(s) not found for architecture x86_64
```

AlgebraicExpression.h

tagged

[c++](#) × 169186

[templates](#) × 12930

[linker](#) × 3820

[compiler-errors](#) × 3343

asked **today**

viewed **13 times**

Artifact: Software Microblogs

- Developers microblog too
- Developers microblog about various activities:
 - Advertisements
 - Code and tools
 - News
 - Q&A
 - Events
 - Opinions
 - Tips
 - Etc.



Artifact: Software Microblogs

Results for #csharp 

Tweets [Top](#) / [All](#)

 **Kunal Chowdhury** @kunal2383 1h
What's the difference between the "ref" and "out" keywords in C#? - shar.es/4aRyl #dotNet #CSharp #InterviewQuestion
[Expand](#)

 **Rozalia Mihailescu** @rozaliamih 1h
Wow! 70% off Programming Courses w/ Paul Deitel: #iOS6, #Android, #Csharp, #Java, #Javascript, C++ ow.ly/gqrvU #udemy
[Expand](#)

 **Robert W. McBean** @robertmcbean 1h
please join my linkedin network @ linkedin.com/in/robertwmcbe...
#csharp #it #software #yyc #web
[Expand](#)

 **Lucian Mihailescu** @rointercer 1h
Wow! STILL ON! 70% off Programming Courses w/ Paul Deitel: #iOS6, #Android, #Csharp, #Java, #Javascript, C++ ow.ly/gv1pn #udemy
[Expand](#)

Opportunities



- Help developers/managers to understand their projects, cope with their evolution, and support them during their decision-making.



- Extract relevant & insightful information, analyze it, and transform to decisions for the future.



- Find trends, anticipate issues, and bring awareness on weaknesses or conditions for making future decisions.

Proactive

- Make proactive decisions using proactive analytics: predictive modelling, data mining, machine learning, statistical analysis, etc.



Opportunities

actionableinsight

data > insight > action

- Leading tech. companies need insights to create actionable tools, increase quality, efficiency, services and risk management.



- Organizations apply analytics to create opportunities for growth, innovation and competitive advantage.



- Data analytics identify patterns, trends and opportunities for improvement, enabling to spot which initiatives work, which fail, and to adjust accordingly.

Challenges

- SE data without explicit format.
- SE data is plentiful.
- Acting on results from data analysis is not easy.
- Analytic tools? May be but should meet the need and be easy to use.
- Adoption of software analytics into software development processes.
- Development and integration of analytics tools in practical settings.