

Big data. Little Brain.  
Little data. Big brain.  
(with apologies to Dr. Seuss)

Gail C. Murphy

University of British Columbia  
Tasktop Technologies



Unless otherwise indicated on a particular slide, this work is licensed under a [Creative Commons Attribution-Share Alike 2.5 Canada License](https://creativecommons.org/licenses/by-sa/2.5/ca/)



problem



big data\* >> little data      —————> little brain

*transformational  
factors*

*situational  
factors*



result

\*big data refers to much much more data than humans can consume

fact #1:  
software engineering is about solving  
human problems



fact #2:

software developments generate big data over time



fact #3:

human cognition has limits



premise:

our “solutions” often exceed human  
cognition

<cartoon removed due  
to license restrictions>

big data, little brain,  
big brain, little data.

duplicate bug examples to derive “formula”

big data >> little data → little brain

stories about formula properties

transformational & situational factors

take away

duplicate bug problem

## Bug 625989 - Support tabs in title bar on Mac OS X

In Firefox 4, we decided to support tabs in title bar on Windows when the window was in a maximized state.

We'd like to add this capability to the OS X version too, but in a slightly different way: Since OS X has no concept of "maximized" (it only has "zoomed"), it should probably be an explicit switch in the preference panel or the Customize palette.

\*\*\* [Bug 626408](#) has been marked as a duplicate of this bug. \*\*\*

## Bug 626408 - Tabs in title bar on OSX

Status: RESOLVED DUPLICATE of bug 625989

\*\*\* [Bug 645789](#) has been marked as a duplicate of this bug. \*\*\*

## Bug 645789 - Hide title bar in 4.x for OSX

Status: RESOLVED DUPLICATE of bug 625989

how does a developer realize/find a new bug is a duplicate?

# big data → big brain

## 1. duplicate bug search with keywords

I like the change to place tabs in the title bar on Windows and would like to see the same thing implemented on OS~~X~~.  
Right now the title bar just displays duplicate information (e.g. the same text as the current tab), I would prefer to use this screen space for something useful. This change would also make Firefox use less vertical screen space than Safari, instead of more, so that would be a bonus too.

This list is too long for Bugzilla's little mind; the Next/Prev/First/Last buttons won't appear on individual

Status: REOPENED, NEW, ASSIGNED, UNCONFIRMED  
Component: bar Alias: bar Summary: bar

Product: title Component: title  
Whiteboard: bar Content: "bar"

Ali

924 bugs found.

ID	Sev ▲	Pri ▲	OS	Assignee	Status	Resolution	Summary
56162	cri	--	Linu	chofmann	NEW	---	mozilla does not r
136422	cri	--	Mac	nobody	NEW	---	Don't use "Sheets
355647	cri	--	All	nobody	NEW	---	Trying to send me
366457	cri	--	Wind	nobody	NEW	---	Address Book Cor
407098	cri	--	Wind	gavin.sharp	NEW	---	Running "firefox [url]" w
448102	cri	--	Wind	rsherry	NEW	---	Fx crashes when displa AcroRd32.dll@0xef2be) (AcroRd32.exe invoked under firefox.exe sqlite.dll exists in System32 directory)
521585	cri	--	Linu	nobody	NEW	---	Qt: Modal dialogs are broken
570014	cri	--	Wind	nobody	NEW	---	When working with IMAP attachments are sometimes corrupt ( body part will be downloaded on demand " is base64 decoded

bug #56162:

re] bug #136422:  
16 reported in 2000  
hu 32 comments

2x as long as #56162

big data >> little data → big brain

2. duplicate bug search with natural language recommender

mozilla.org  
Bugzilla Bug 297150 View source text should be editable  
Last modified: 200...  
First Last Prev Next No search results available Search page Enter new bug

Bug#: 297150  
alias:   
Product: Firefox  
Component: View Source  
Status: UNCONFIRMED  
Resolution: Nobody's working on this, feel free to close it  
Hardware: PC  
OS: Windows XP  
Version: unspecified  
Priority: --  
Severity: normal  
Target Milestone: ---

Duplicates of Bug 297150

Bug I...	Summary
221668	View Source is broken
291603	View source doesn't show the correct source
245851	View Selection Source automatically converts uppercase HTML tags to lower case
230693	Missing reload within view source window
279614	Should be able to open view source window in a new tab
172817	Allow external source viewer/editor
285560	a view-source windows should show the source of a streamed page

36-50% recall

[Runenson et. Al, ICSE 2007]

[Hiew, MSc Thesis, 2006]

big data >> little data → big brain

3. duplicate bug search with natural language + execution recommender

The screenshot shows a Mozilla Bugzilla page for bug 297150. The page title is "Bugzilla Bug 297150" and the description is "View source text should be editable". The page includes navigation links for "First", "Last", "Prev", "Next", "Search page", and "Enter new bug". The search filters are as follows:

- Bug#: 297150
- alias: [empty]
- Product: Firefox
- Component: View Source
- Status: UNCONFIRMED
- Resolution: Nobody's working on this, feel free to help
- Hardware: PC
- OS: Windows XP
- Version: unspecified
- Priority: --
- Severity: normal
- Target Milestone: ---

At the bottom, there is a table of duplicates for bug 297150:

Bug I...	Summary
221668	View Source is broken
291603	View source doesn't show the correct source
245851	View Selection Source automatically converts uppercase HTML tags to lower case
230693	Missing reload within view source window
279614	Should be able to open view source window in a new tab
172817	Allow external source viewer/editor
285560	a view-source windows should show the source of a streamed page

67-93% recall

[Wang et. Al, ICSE 2008]

big data >> little data → little brain

## 4. duplicate bug search with recommender & summarizer

mozilla.org  
Bugzilla Bug 297150

View source text should be edited

First Last Prev Next No search results available Search page

Bug#: 297150  
alias:   
Product: Firefox  
Component: View Source  
Status: UNCONFIRMED  
Resolution: Nobody's working on this, feel free to help

Hardware: PC  
OS: Windows  
Version: unspecified  
Priority: --  
Severity: normal  
Target Milestone: ---

Bug I...	Summary
221668	View Source is broken
291603	View source doesn't show the correct source
245851	View Selection Source automatically converts uppercase HTML tags to lower case
230693	Missing reload within view source window
279614	Should be able to open view source window in a new tab
172817	Allow external source viewer/editor
285560	a view-source windows should show the source of a streamed page

Summary:

View source is broken.

The second time I use it, the View source window fails to open, but I get an hourglass-and-pointer cursor.

I can't reproduce comment 0 with Mozilla/5.0

interludes

big data >> little data

significant reduction in data  
a human considers

big brain versus little brain

amount of cognitive and physical  
activity required to perform task

big data >> little data → little brain

big data, little brain,  
big brain, little data.

duplicate bug examples to “derive formula”

big data >> little data → little brain

stories about formula properties

transformation & situational factors

take away

stories

software reflexion model

hipikat

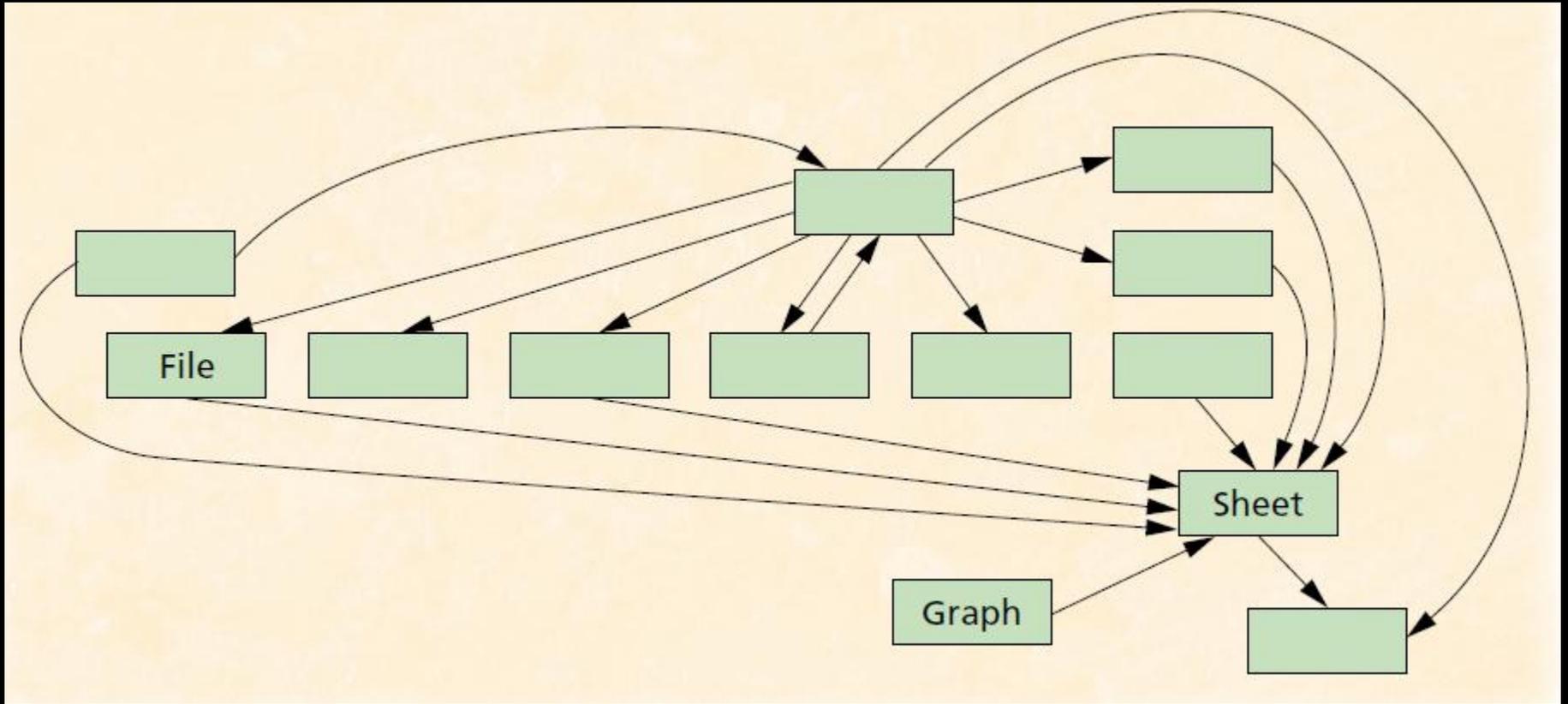
mylyn

spyglass

to get at formula properties

story #1: software reflexion model  
three case studies,  
including one at *Microsoft*

[Murphy, Notkin, Sullivan, FSE 1995]



high-level model of part of Microsoft Excel

[Murphy & Notkin, Computer 1997]



example (not Excel) low-level model  
(Excel : 15,000 functions/77,746 calls)

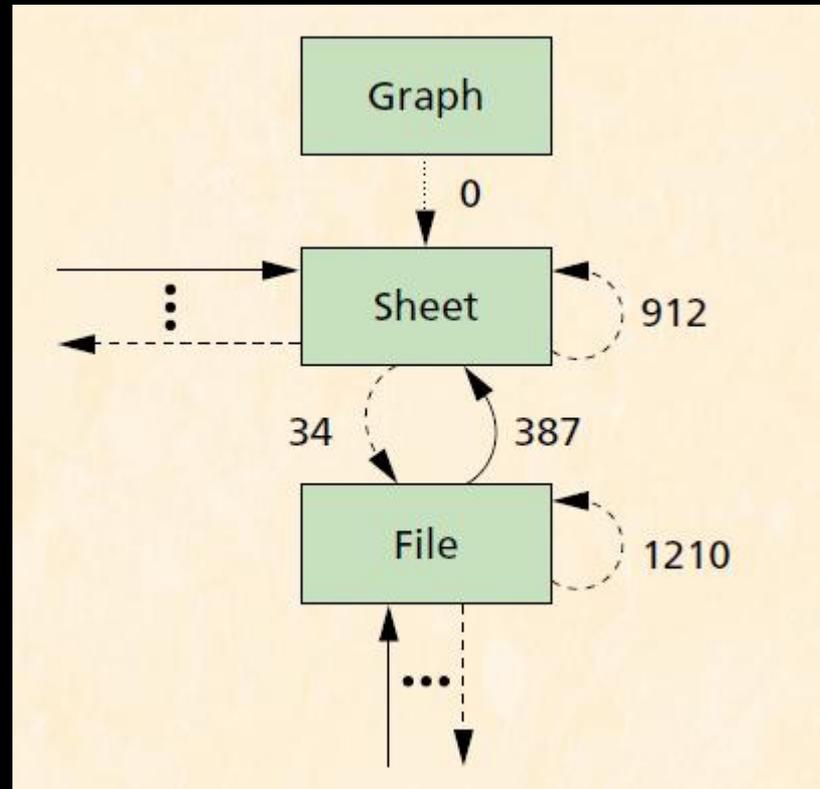
```
[ file = ^shtreal\.c function=foo mapTo=Graph ]
```

```
[ file = ^shtreal\.c mapTo=Sheet ]
```

```
[ file = ^textfl1[ez]\.c$ mapTo=File ]
```

mapping

blue entry is illustrative



reflexion model (partial)

big data >> little data → little brain

big data = low-level model

little data = high-level model

transformational properties

low-cost (albeit manual)

assessable completeness

predictable

high-level  
model

mapping

reflexion

model

mapped  
low-level  
model

predictable

unmapped  
low-level  
model

low-cost because  
mapping can be  
specified simply,  
incrementally and  
partially

easy to assess  
completeness

# story #1: software reflexion models

big data >> little data → little brain

low-cost (manual)

assessable completeness

predictable

story #2: hipikat

wizard-of-oz case study

multiple case study (long sessions)

manual precision/recall

[Cubranic & Murphy, ICSE 2003]

File Edit Source Refactor Navigate Search Project Run Window Help

Package Explorer x TagAction...

SyncAction  
SyncOutp  
TagAction  
TagInRepl  
Unmanage  
UpdateAct  
org.eclipse.tei  
org.eclipse.tei  
org.eclipse.tei  
org.eclipse.tei  
ECLIPSE\_HOME/pl  
ECLIPSE\_HOME/pl  
icons  
about.html  
build.properties  
plugin.properties  
plugin.xml  
org.eclipse.team.ui  
JRE\_LIB - C:\Prog  
src  
org.eclipse.tei  
Configur...

Package E... Hier

Bug 116 - [CVS Repo View] "Version from Stream" should refresh view

Bug#: 116 Product: Platform  
Version: 2.0 Platform: All  
OS/Version: All Status: RESOLVED  
Severity: normal Priority: P4  
Resolution: FIXED Assigned To: platform-vcn-i:  
Reported By: Kevin\_McGuire@oti.com Component: VCM  
Target Milestone: 2.1 M4 URL:

Summary:  
[CVS Repo View] "Version from Stream" should refresh view (1GENEDO)

Keywords:  
investigate

Description:  
(From Julian J)  
When the user selects the "Version from stream..." action in the Repository view, I'd recommend you refresh the view once the action has been completed.

NOTES:  
KM (6/1/01 11:26:01 AM)

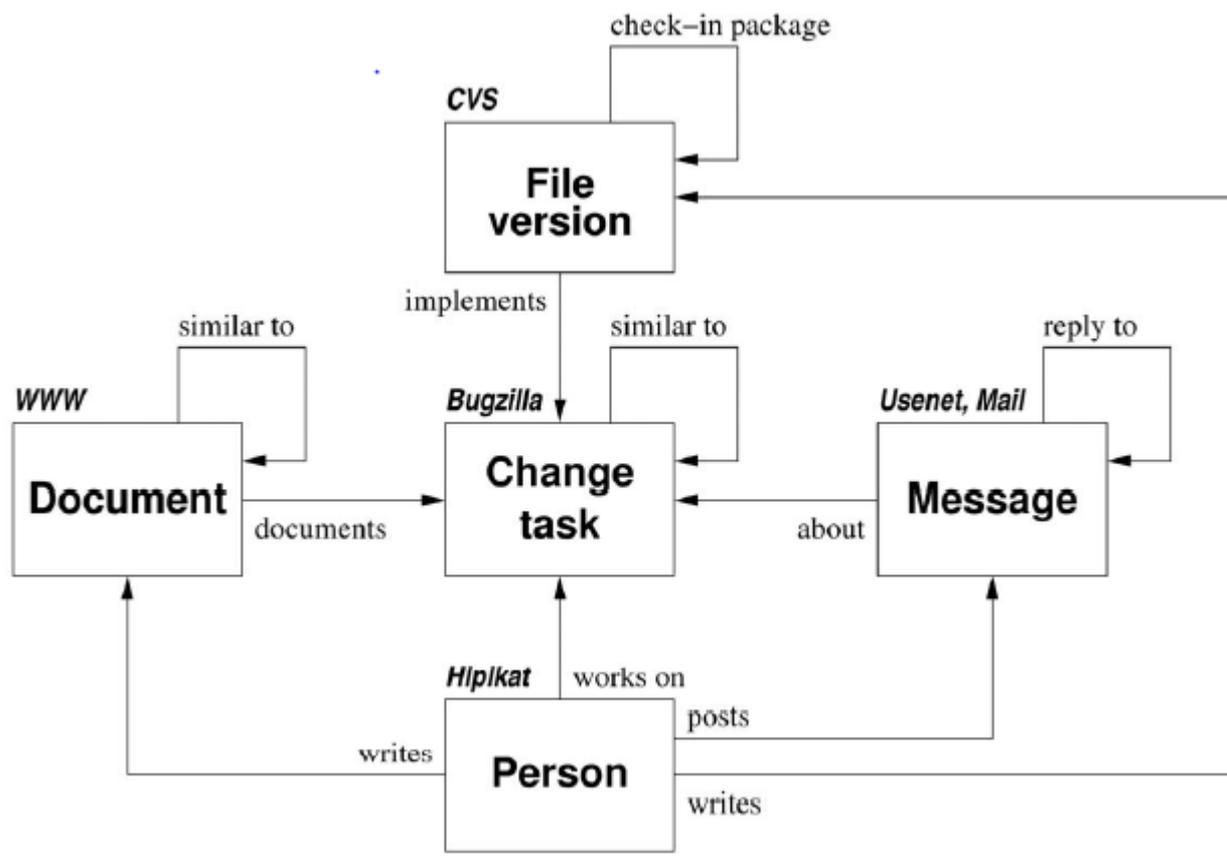
Hipikat (Bug 16161)

Type	Name	Reason	Confidenc
cvs	dev.eclipse.org;/home/eclipse/org.eclipse.jdt.ui/ui/org...	Check-in close to bug resolution	Medium lo
cvs	dev.eclipse.org;/home/eclipse/org.eclipse.jdt.ui/ui/org...	Check-in close to bug resolution	Medium lo
bugzilla	Bug 13176 - Team: task "Local addition not under CVS ...	Text similarity	0.328189
bugzilla	Bug 10416 - Improved workflow for "Delete"	Text similarity	0.308314
bugzilla	Bug 10883 - Cannot add new file to CVS	Text similarity	0.31453

(a)

(b)

[Cubranic, Murphy, Booth and Singer, TSE 2005]



---

<b>Location in the IDE</b>	<b>Artifact type</b>
Bug report open in the Bugzilla editor	change task
CVS-managed file open in the Java editor	file version
File in the CVS Repository view	file version
Revision in the CVS Resource History view	file version
CVS-managed file in the workspace Navigator	file version
Item recommended in the Hipikat Results view	<i>item's type</i>
Bugzilla search match in the Search results view	change task
Java class or method in Outline and Hierarchy views	file version

---

+ text

(e.g., exception trace)

big data >> little data → **big brain**

big data = project repositories

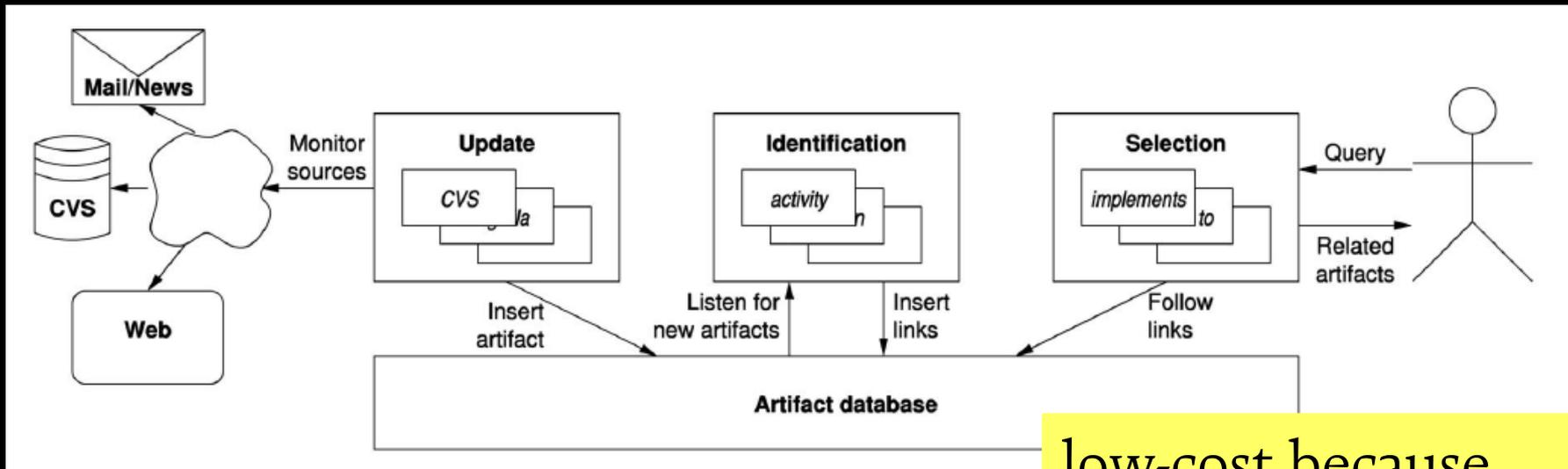
little data = recommendations

transformational properties

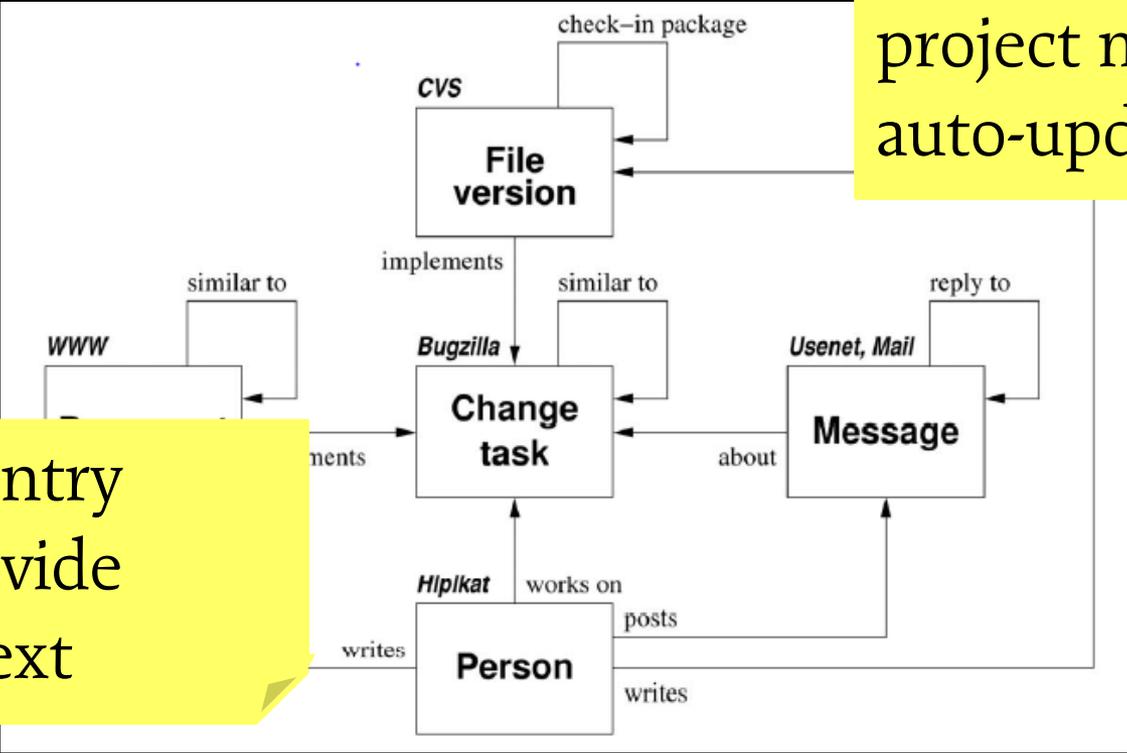
low-cost (automatic)

situational factors

context (automatic)



low-cost because project memory auto-updates



multiple entry points provide auto-context

story #2: hipikat

big data >> little data → **big brain**

low-cost (automatic)

context (automatic)

story #3: mylyn

three field studies

adoption in practice

[Kersten & Murphy, FSE 2006]

demo

big data >> little data → little brain

big data = workspace information

little data = focused workspace information

transformational properties

low-cost (automatic)

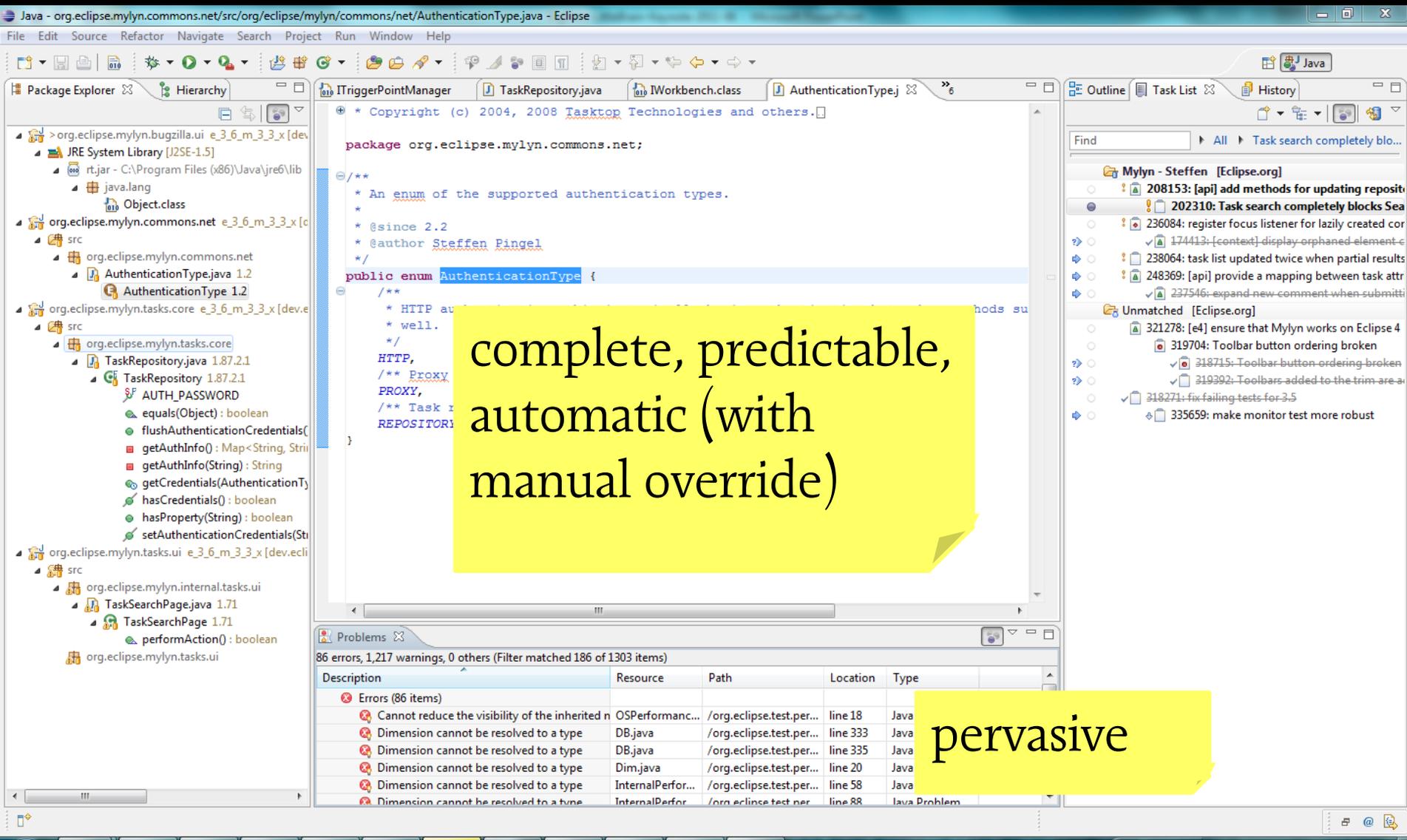
assessable completeness

predictable

situational factors

pervasive

fits in workflow



```

 * Copyright (c) 2004, 2008 Tasktop Technologies and others.

package org.eclipse.myllyn.commons.net;

/**
 * An enum of the supported authentication types.
 *
 * @since 2.2
 * @author Steffen Pingel
 */
public enum AuthenticationType {

    /**
     * HTTP authentication
     * well.
     */
    HTTP,
    /** Proxy
     * authentication
     */
    PROXY,
    /** Task repository
     * authentication
     */
    REPOSITORY
}

```

complete, predictable,  
automatic (with  
manual override)

86 errors, 1,217 warnings, 0 others (Filter matched 186 of 1303 items)

Description	Resource	Path	Location	Type
<b>Errors (86 items)</b>				
Cannot reduce the visibility of the inherited n	OSPerformanc...	/org.eclipse.test.per...	line 18	Java
Dimension cannot be resolved to a type	DB.java	/org.eclipse.test.per...	line 333	Java
Dimension cannot be resolved to a type	DB.java	/org.eclipse.test.per...	line 335	Java
Dimension cannot be resolved to a type	Dim.java	/org.eclipse.test.per...	line 20	Java
Dimension cannot be resolved to a type	InternalPerfor...	/org.eclipse.test.per...	line 58	Java
Dimension cannot be resolved to a type	InternalPerfor	/org.eclipse.test.per...	line 88	Java Problem

pervasive

```

private void initContextChangeSets () {
    * For testing.
    public void clearActiveChangeSets () {
        activeChangeSets.clear ();
    }
    public IResource[] getResource...
    public void cont...
    public void cont...
    public List<Cont...
    private ITask ge...
        List<ITask>...
        if (activeTa...

```

folding/ &  
content assist

type filter text

- Eclipse Application
- Equinox OSGi Framework
- Java Applet
- Java Application
- JUnit
- JUnit Plug-in Test
- Context Plug-in Test
- Test Active Context (plug-in)
- Context Test
- Test Active Context

Name: Test Active Context (plug-in)

JUnit Main Arguments

Subclasses of TestCase in the active task context automatically added to suite.

workflow

test

change sets

History Task List

TaskListToolTipHandler.java

Revisi...	Tags	Revisi...	Author	Comment
1.23		4/24/07 3...	relves	NEW - bug 181038: [api] Pul...
1.20	R_2_0_M2	3/30/07 5...	relves	ASSIGNED - bug 178678: Act...
1.19		3/30/07 1...	relves	NEW - bug 166990: [multi m...
1.18		3/23/07 4...	relves	NEW - bug 176297: aet rid o...

NEW - [bug 166990](#): [multi monitor] if Eclipse is on secondary monitor to left of primary task list tooltips show on primary

story #3: mylyn

big data >> little data → little brain

low-cost (automatic)

assessable completeness

predictable

pervasive

fits in workflow

story #4: spyglass

longitudinal case study  
controlled lab study

[Viriyakattiyaporn & Murphy, CASCON 2010]

```
Map axisStateMap = new HashMap();
```

```
// draw the top axes
```

```
double yMin = dataArea.getMinY();
```

```
double yMax = dataArea.getMaxY() - this.axisOffset.calculateTop
```

B

A

### Rationale

If you wanted to see **call relationship** between `draw` in `[CategoryAxis]` and `drawAxes` in `[CategoryPlot]`,

you could have used Open Call Hierarchy via

**Keyboard Shortcut:** Select class/method > Ctrl + Alt + H

**Menu:** Context Menu: Select class/method > Right Click > Open Call Hierarchy

Open Call Hierarchy opens up the selected method in the Call Hierarchy view showing its callers or callees.

Rationale

```
drawRangeGridlines(Graphics2D, Rectangle2D)
```

```
drawRangeMarkers(Graphics2D, Rectangle2D,
```

```
getDataRange(ValueAxis)
```

```
getDomainAxisIndex(CategoryAxis)
```

### Recommended Functionality

Name	Shortcut Key	Fre...	Created
<input checked="" type="checkbox"/> Open Call Hierarchy	Select class/method > Ctrl + Alt + H	0	02/19/09 03:4
<input type="checkbox"/> Open Type Hierarchy	Select class/method > F4	0	05/29/09 11:2
<input checked="" type="checkbox"/> Work Item Editor Links	N/A	0	05/29/09 11:2

[Click here to go to Eclipse Help page.](#)

Rate This

Tools Recommendation

Writable

Smart Insert

<No Current work>



big data >> little data → **big brain**

big data = user interactions in workspace

little data = recommended commands

transformational properties

low-cost (automatic)

situational factors

(mostly) fits in workflow

(mostly) fits in workflow

B

A

### Rationale

If you wanted to see **call relationship** between `draw` in `[CategoryAxis]` and `drawAxes` in `[CategoryPlot]`.

you could have used Open Call Hierarchy via

**Keyboard Shortcut:** Select class/method > `Ctrl + Alt + H`

**Menu:** Context Menu: Select class/method > Right Click > Open Call Hierarchy

Open Call Hierarchy opens up the selected method in the Call Hierarchy view showing its callers or callees.

Rationale

- drawRangeGridlines(Graphics2D, Rectangle2D)
- drawRangeMarkers(Graphics2D, Rectangle2D)
- getDataRange(ValueAxis)
- getDomainAxisIndex(CategoryAxis)

### Recommended Functionality

Name	Shortcut Key	Fre...	Created
Open Call Hierarchy	Select class/method > <code>Ctrl + Alt + H</code>	0	02/19/09 03:4
Open Type Hierarchy	Select class/method > <code>F4</code>	0	05/29/09 11:2
Work Item Editor Links	N/A	0	05/29/09 11:2

[Click here to go to Eclipse Help page.](#)

Rate This

Tools Recommendations

<No Current Work>

automatic

story #4: spyglass

big data >> little data → **big brain**

low-cost (automatic)

(mostly) fits in workflow

lessons

lessons about transformational factors

big data >> little data

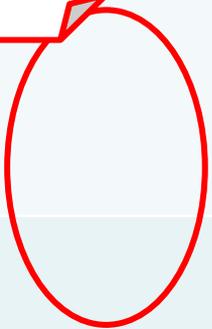
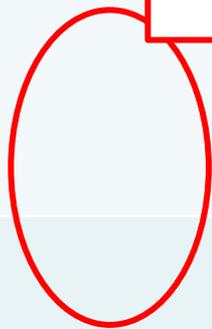
low-cost (manual or automatic)

assessable completeness

predictable

	reflexion model	hipikat	mylyn	spyglass
low-cost	√	√	√	√
assessable completeness	√			
predictable	√		√	

need sufficiently high precision/recall?



lessons about situational factors

little data  $\longrightarrow$  little brain

context (automatic)

pervasive

workflow

big brain

	reflexion model	provide intent automatically?	how to handle specialized tools?	spyglass
context (automatic)				
pervasive				
workflow			✓	~





meghan allen  
 john anvik  
 elisa baniassad  
 wesley coelho  
 davor cubranic  
 brian de alwis  
 rob elves  
 thomas fritz  
 jan hannemann  
 lyndon hiew  
 reid holmes  
 mik kersten  
 shawn minto  
 sarah rastkar  
 martin robillard  
 izzet safer  
 david shepherd  
 ducky sherwood  
 apple  
 viriyakkiyaporn  
 annie ying  
 trevor young  
 robert walker  
*and others!*

take away



problem

big data >> little data  $\longrightarrow$  little brain

*transformational  
factors*

*situational  
factors*

result