

Essentiality of Context in Software Analytics

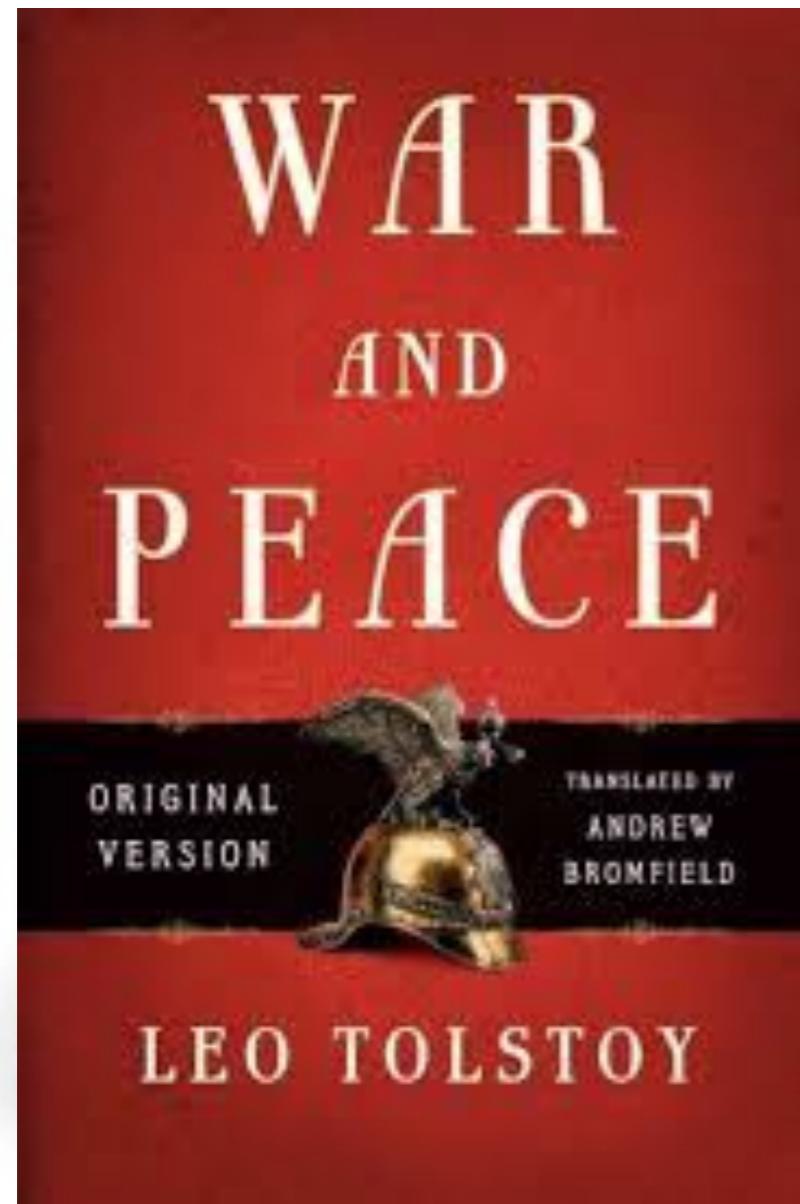
Mircea F. Lungu
SEARCH
JBI
University of Groningen

January 2017, SEN Symposium, CWI, Amsterdam

Three types ~~Essentiality~~ of Context in Software Analytics

Mircea F. Lungu
SEARCH
JBI
University of Groningen

January 2017, SEN Symposium, CWI, Amsterdam

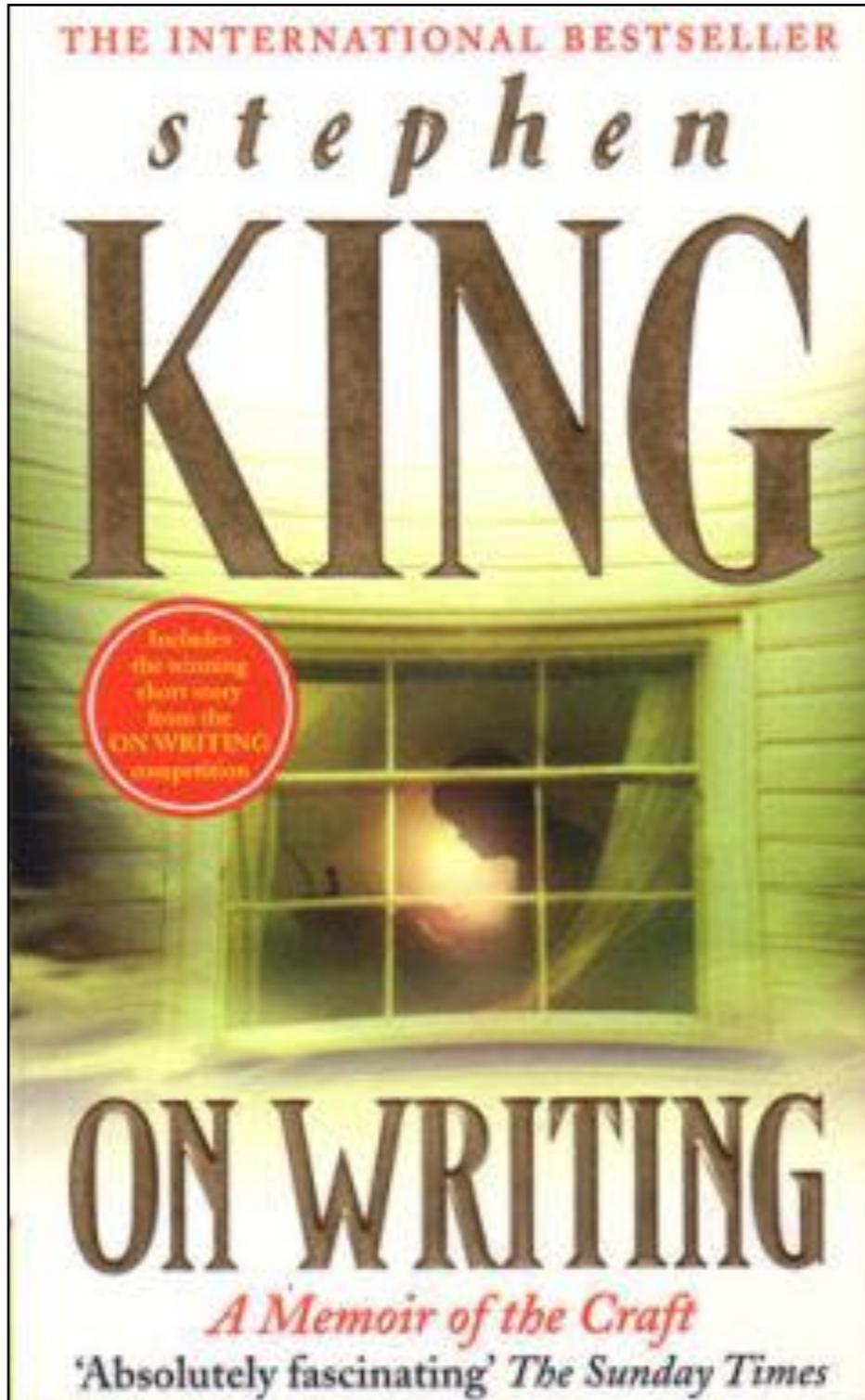


60'000 LOT



60'000 LOC

[Inspired by Andrea Caracciolo]



**Two things above all others:
read a lot and write a lot**

7
**Confessions
of a Book
Editor**



Editor



**Teaching students that
source code is Rwx**

Naming functions...



a strange and difficult art,
a little bit like an
abstracted form of poetry
(S. Wolfram)

Good code detectors and code perfumes!



The codebase of Google has 2 billion LOC.

All in one place

**Remember that time when you
tried to stop using google?
and then you changed your mind?**

Google's future is in **contextual awareness**, machine learning : Sundar Pichai's Founder's Letter - Tech2



Every year Sergey Brin and Larry Page write a letter to their stakeholders, giving an overview of the company and where it is headed. This year, post the re-organisation that made Sundar Pichai CEO of Google and Alphabet became the parent company, the founder's letter was written by Pichai. The letter has some clues to where Google services are headed. Search is going to get more contextual awareness, personal demographic factors, location characteristics and environmental variables will all start changing search results. Search will not just get more personal, but also consider the time and activity.

**Using context to offer you
the best... ads!**

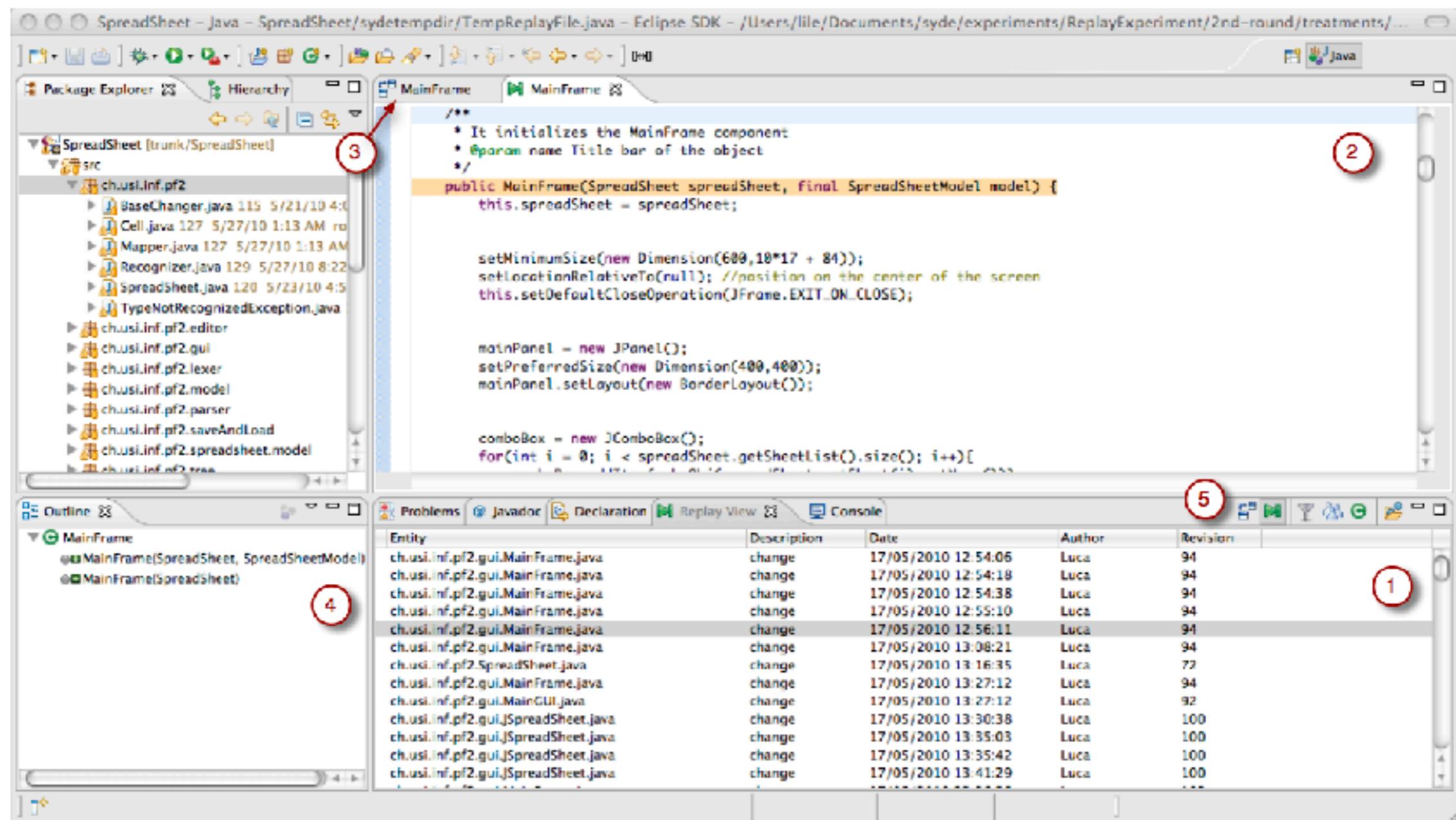


Temporal Context

short term



Replaying past development sessions



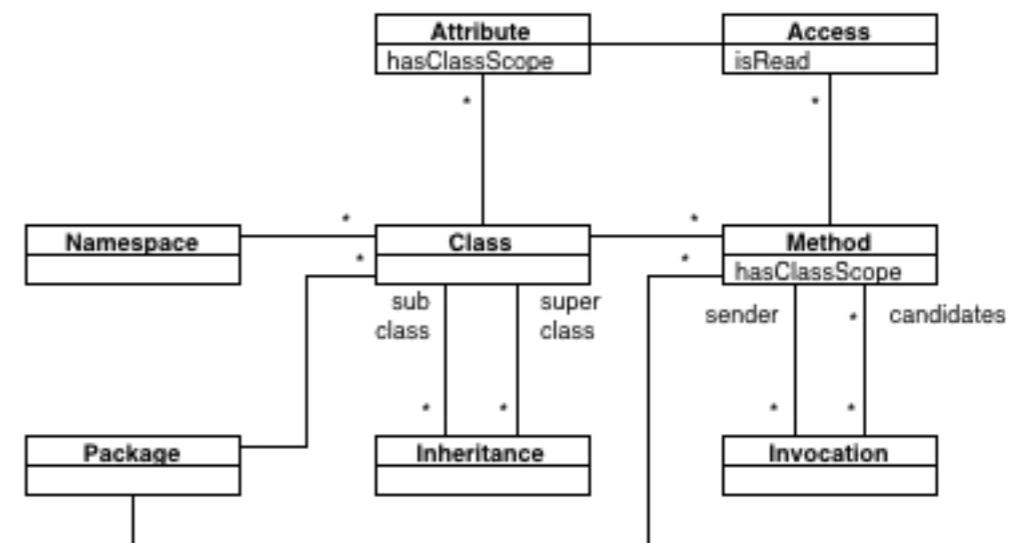
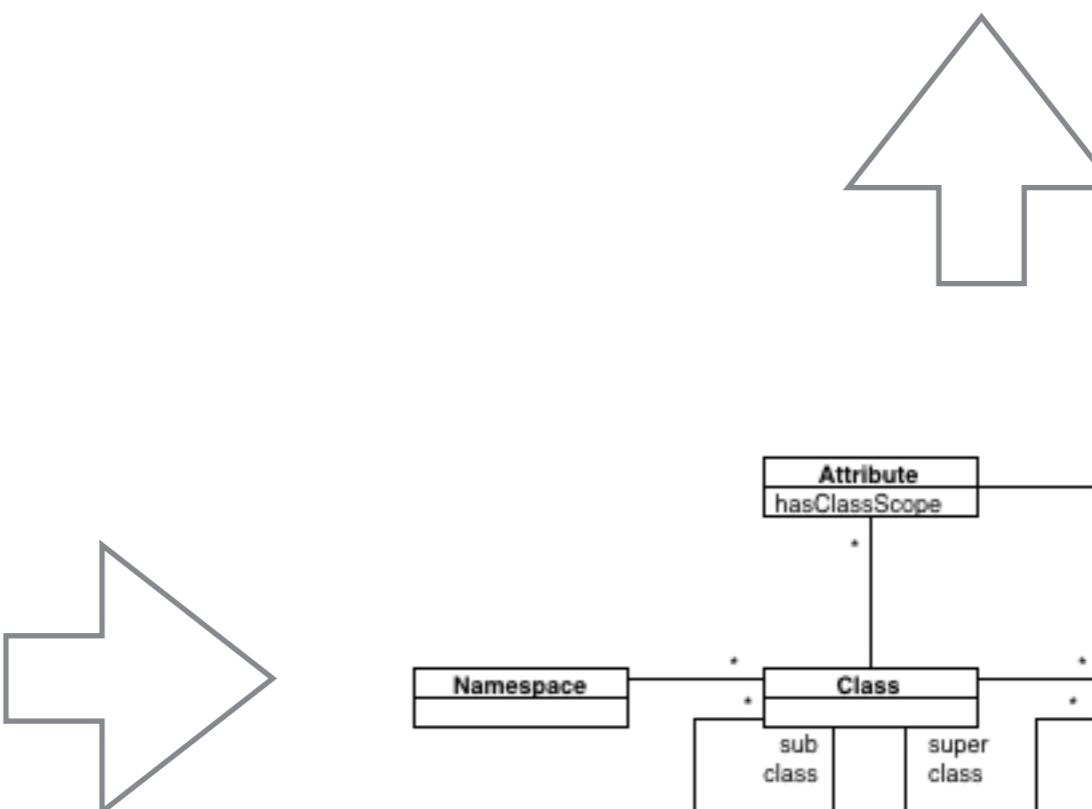
Temporal Context

long term

Software Analytics Platform

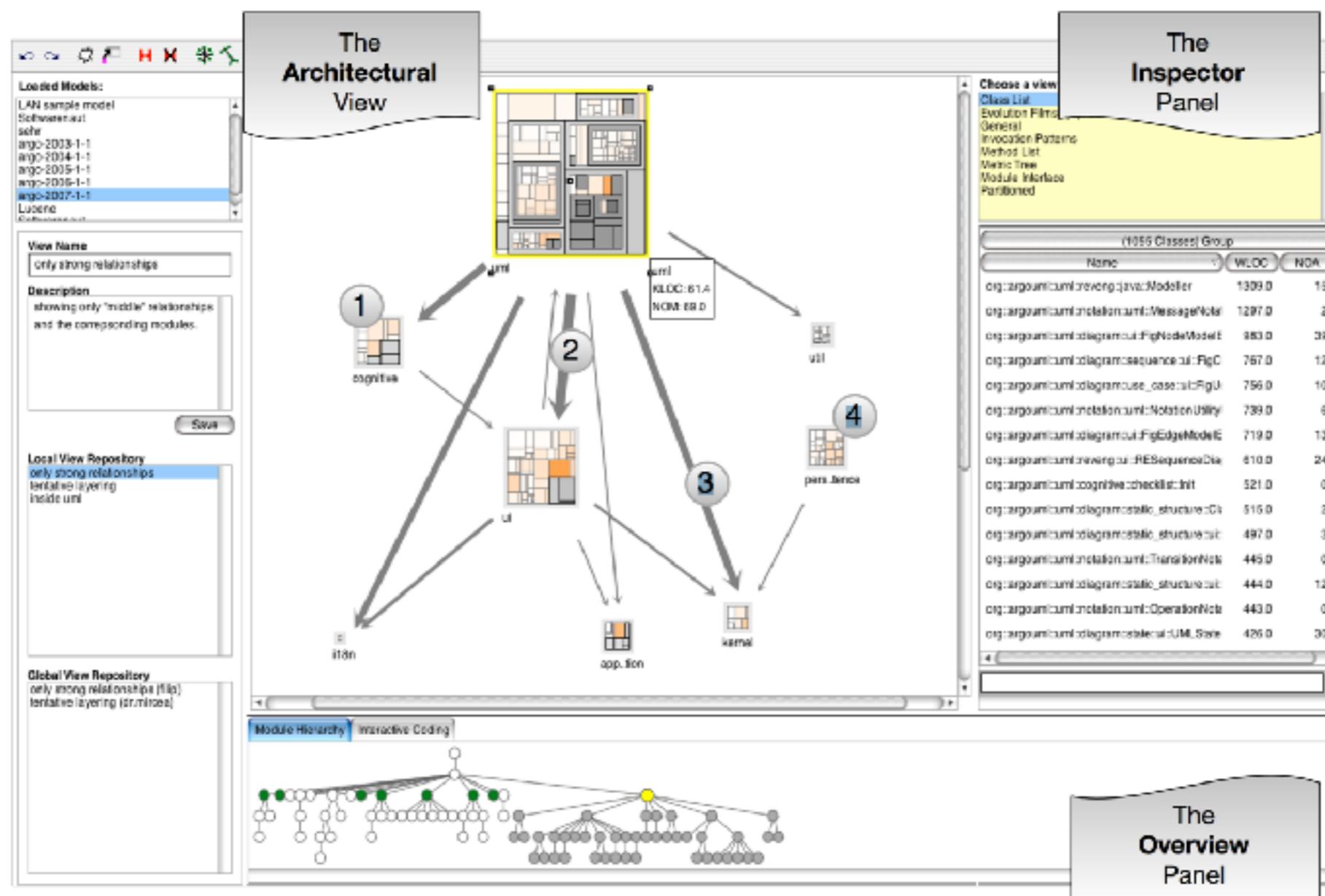


60'000 LOC



Language Independent
Meta-Model

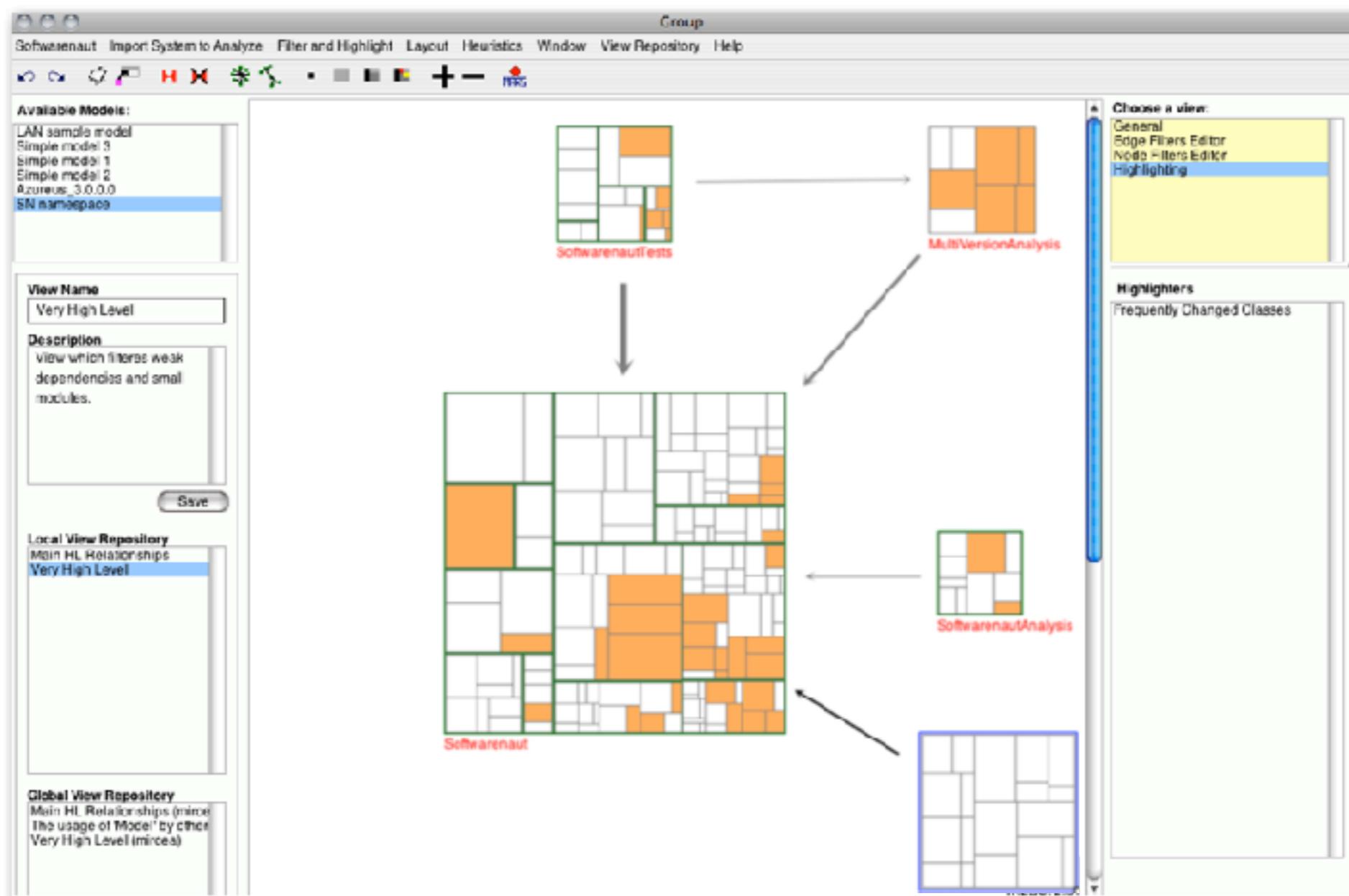
Interactive Architecture Recovery



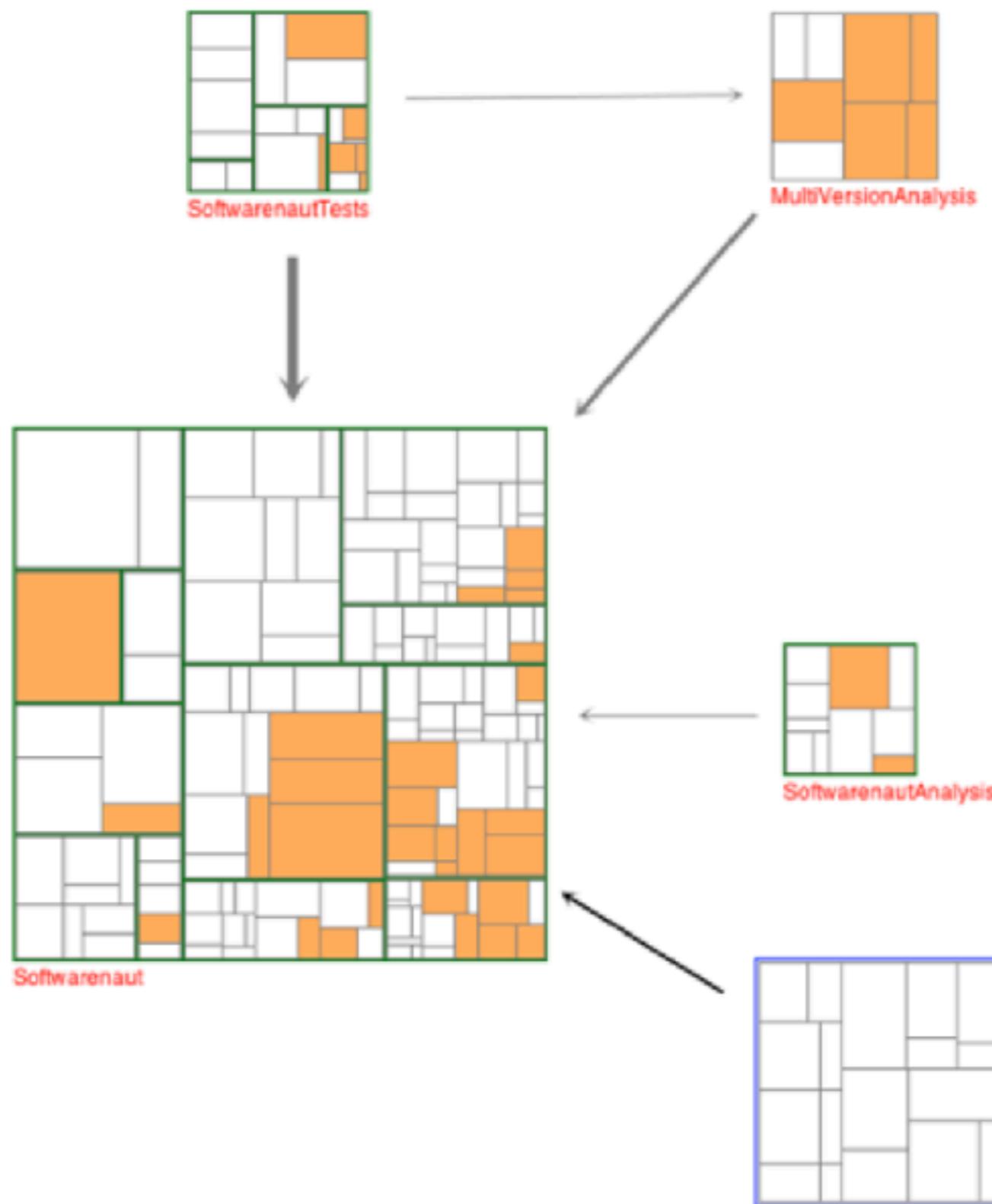
Evolutionary and Collaborative Software Architecture Recovery with Softwarenaut.

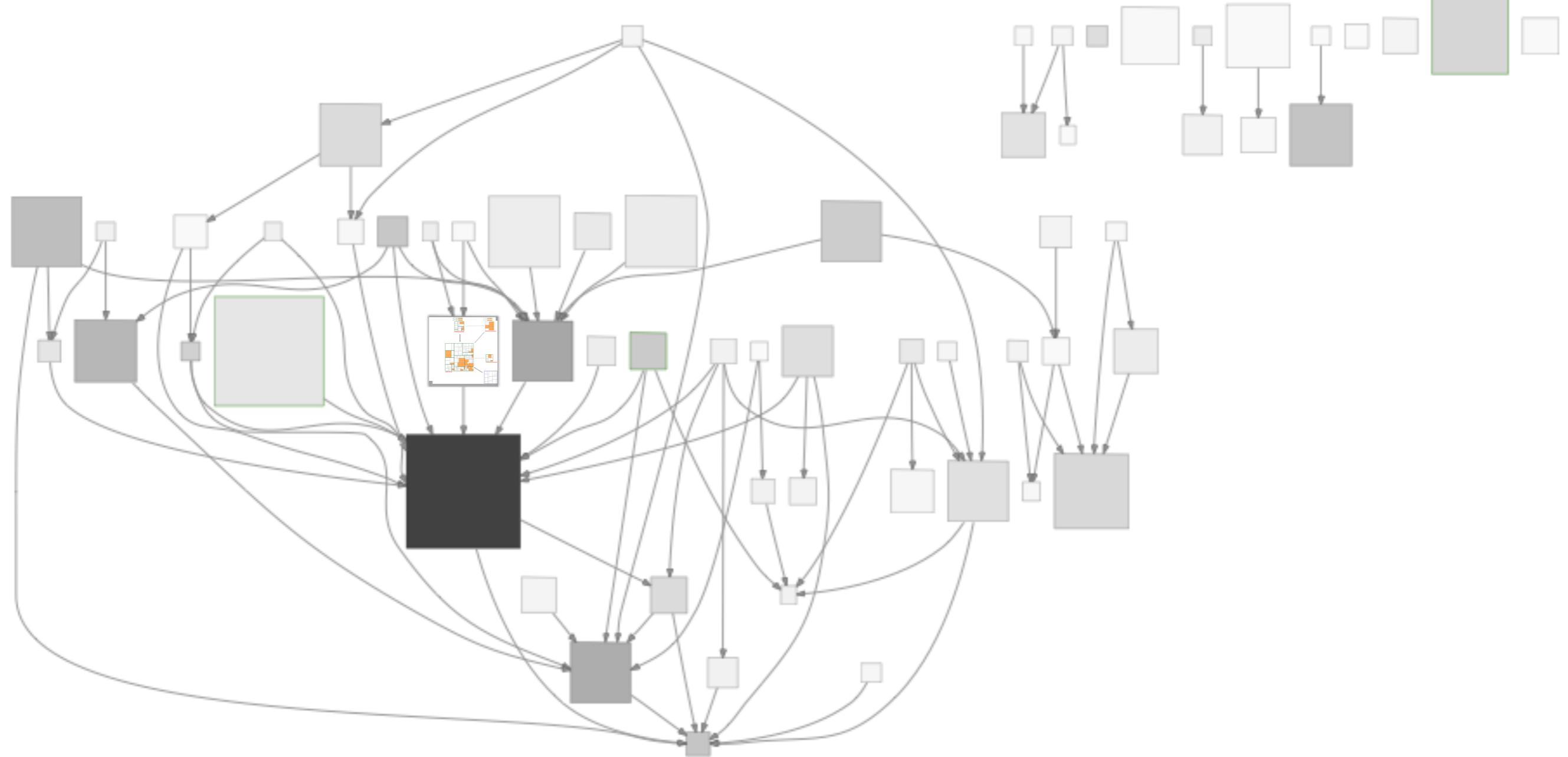
Mircea Lungu, Michele Lanza, and Oscar Nierstrasz. In Science of Computer Programming 79(0) p. 204 - 223, 2014.

Hotspots can focus the analysis process



“No system is an island, entire of itself”

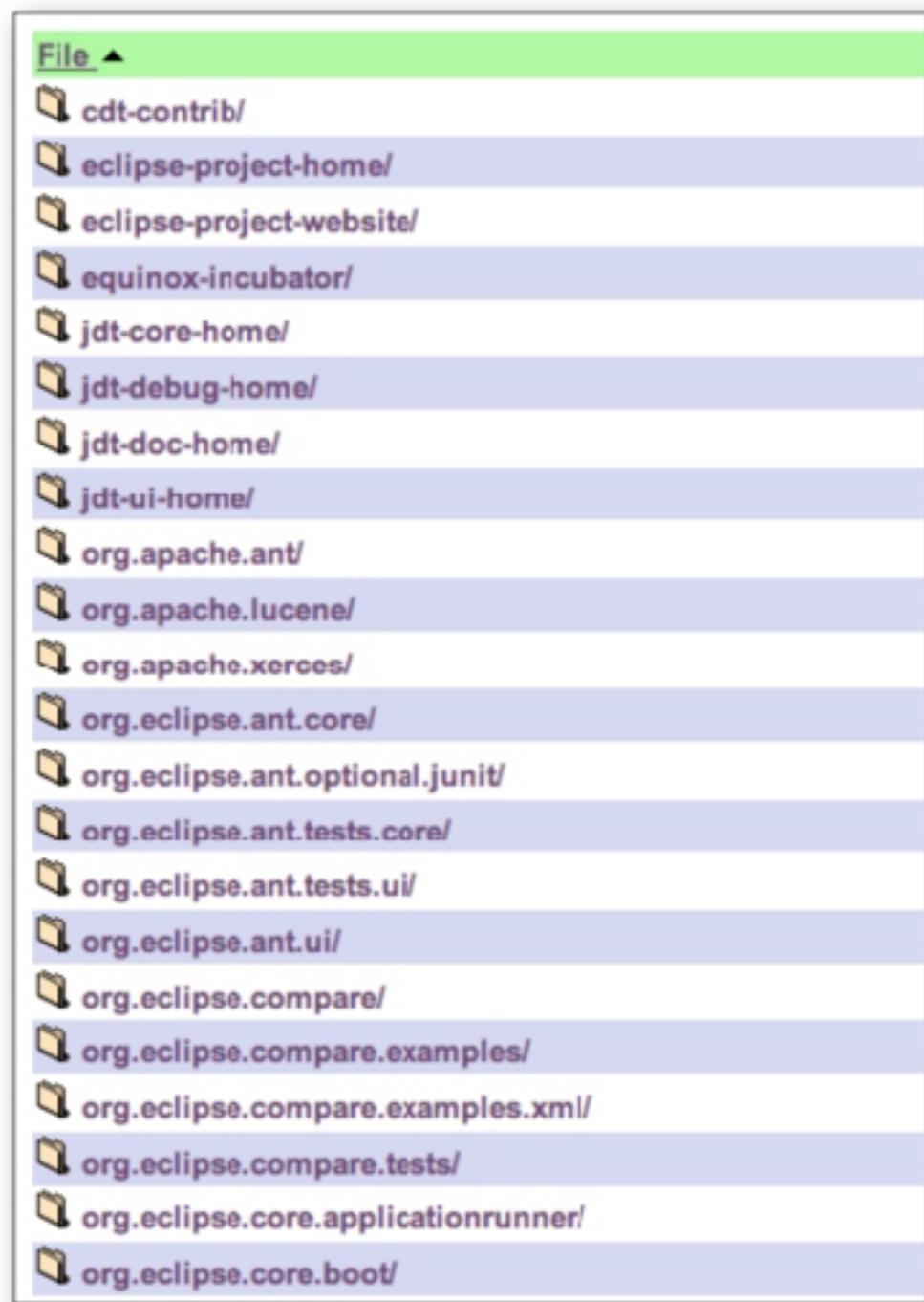




The Systemic Context

The ecosystem: challenges

And some organizations are still using tools like these!



A screenshot of the Cincom Public Repository interface. The title bar says "Published Items on Cincom Public Repository".

The left pane lists "Bundles and Packages":

- Globe
- Globe Development
- Glop** (selected)
- Glop-LensTranslator
- GlopTest
- GlopTools
- GnuCash Utilities
- Goodfellas
- GrapEInfoModel
- Graphs
- GraphsUI
- GreenProject
- Grid
- Grocery
- Hash Analysis Tool
- Hash Analysis Tool Hash Functions
- Hashes
- HotDraw
- HTTP-Support
- HttpStreams 1.0.0
- Hyper
- Hyper Development
- iCal
- IMAP-Support
- Intellika
- Intensional Tools
- InWordsBundle
- IRC-Babelfish-Plugin
- IRC-BottomFeeder-Plugin
- IRC-Client-TypeLess
- IRC-Protocol
- IRC-Support

The right pane shows "Versions" for the selected item:

Version	Type	Date
(0.4.139,aknight)	Development	08/19/2007 16:4520.000
(0.4.138,aknight)	Development	08/16/2007 5:15:23.000
(0.4.137,aknight)	Development	08/09/2007 8:03:27.000
(0.4.136,aknight)	Merged	08/09/2007 3:47:00.000
(0.4.134,1,aknight)	Integrated	08/09/2007 3:45:52.000
(0.4.135,aknight)	Development	08/07/2007 15:4438.000
(0.4.134,aknight)	Development	07/26/2007 14:1005.000
(0.4.133,aknight)	Development	07/23/2007 11:4330.000
(0.4.132,aknight)	Development	07/12/2007 10:2013.000
(0.4.131,aknight)	Development	07/02/2007 10:5716.000
(0.4.130,aknight)	Development	07/01/2007 16:0949.000

Comments for the Glop package:

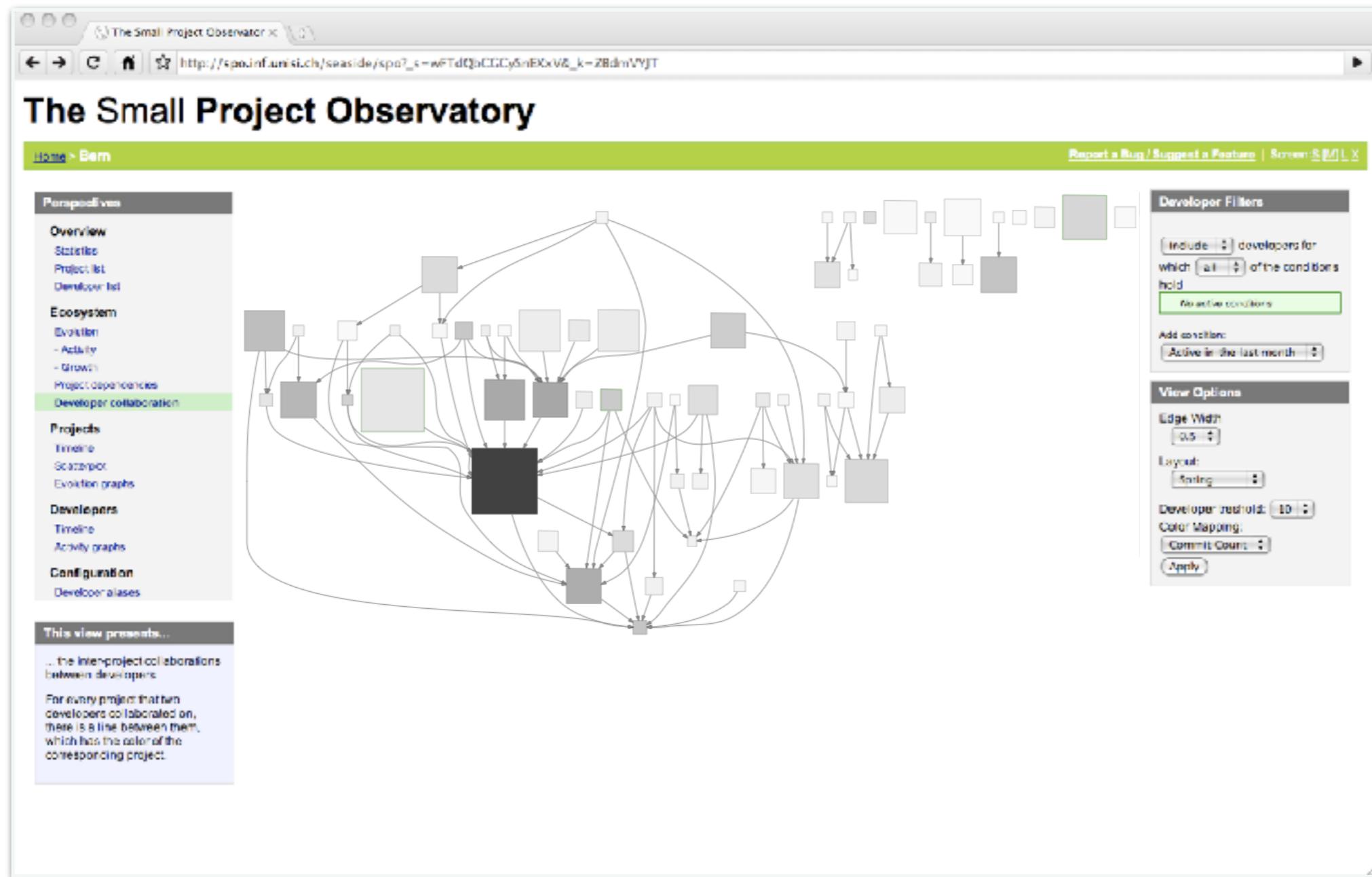
This is the GLOP system, an open-source library (LGPL(S)) for persisting Smalltalk objects in relational databases. For more information see the accompanying files and/or the web sites <http://www.glop.org> or <http://www.sourceforge.net/projects/glop>

The code is Copyright (C) 2000-2003 Alan Knight

This is a Smalltalk library, licensed under the LGPL, and distributed WITHOUT ANY WARRANTY (see below). However, it is not obvious how some of the terms and concepts in the license should be sensibly applied to Smalltalk. The various Smalltalk image models often do not directly correspond to the ideas of programs and libraries that are linked to form executables. We, the authors, would like to clarify our interpretation of the LGPL as it applies to Smalltalk, and what we permit you to do with this code. We will sometimes use the term LGPL(S) to distinguish the use of the LGPL with these clarifications.

This code is intended to be usable as a library, without the intention to

Ecosystem Dashboards



The Small Project Observatory: Visualizing Software Ecosystems. Mircea Lungu, Michele Lanza, Tudor Gîrba, and Romain Robbes. In Science of Computer Programming, Elsevier 75(4) p. 264—275, April 2010.



User:

[...] where is [REDACTED] anymore, why was it deleted?? I use it and now is gone!!! I even had a specialization of it [...]

Competitor Developer:

They have been dropped. A mail went out to this list if anybody still used them and nobody replied. [...] Personally I don't know of any application that uses these dialogs.

UM-1: Strengthening self-esteem
UM-2: Maintaining downstream compatibility

UM-3: Managing resources

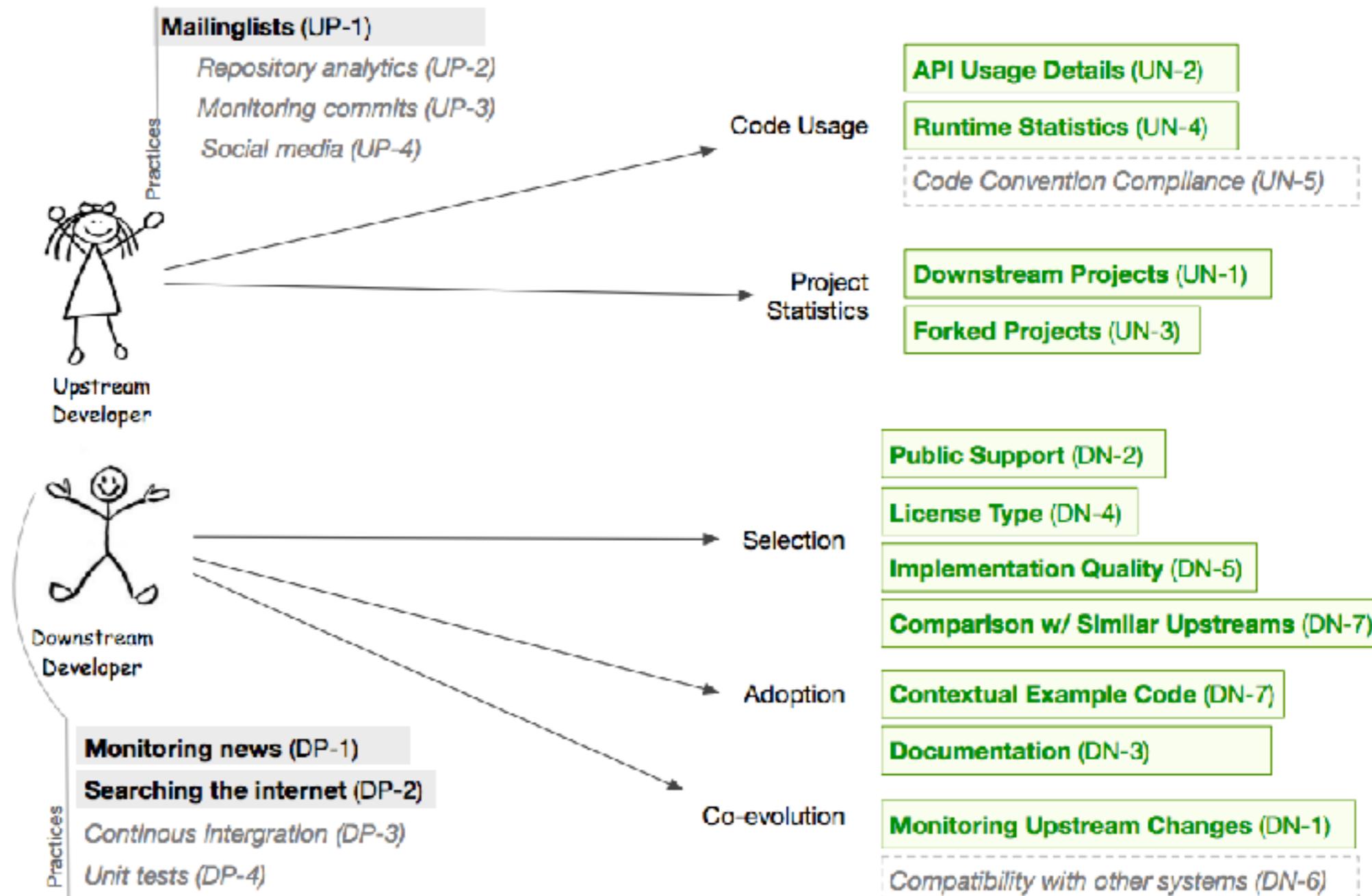
DM-1: API Understanding

DM-2: Keeping up with upstream evolution

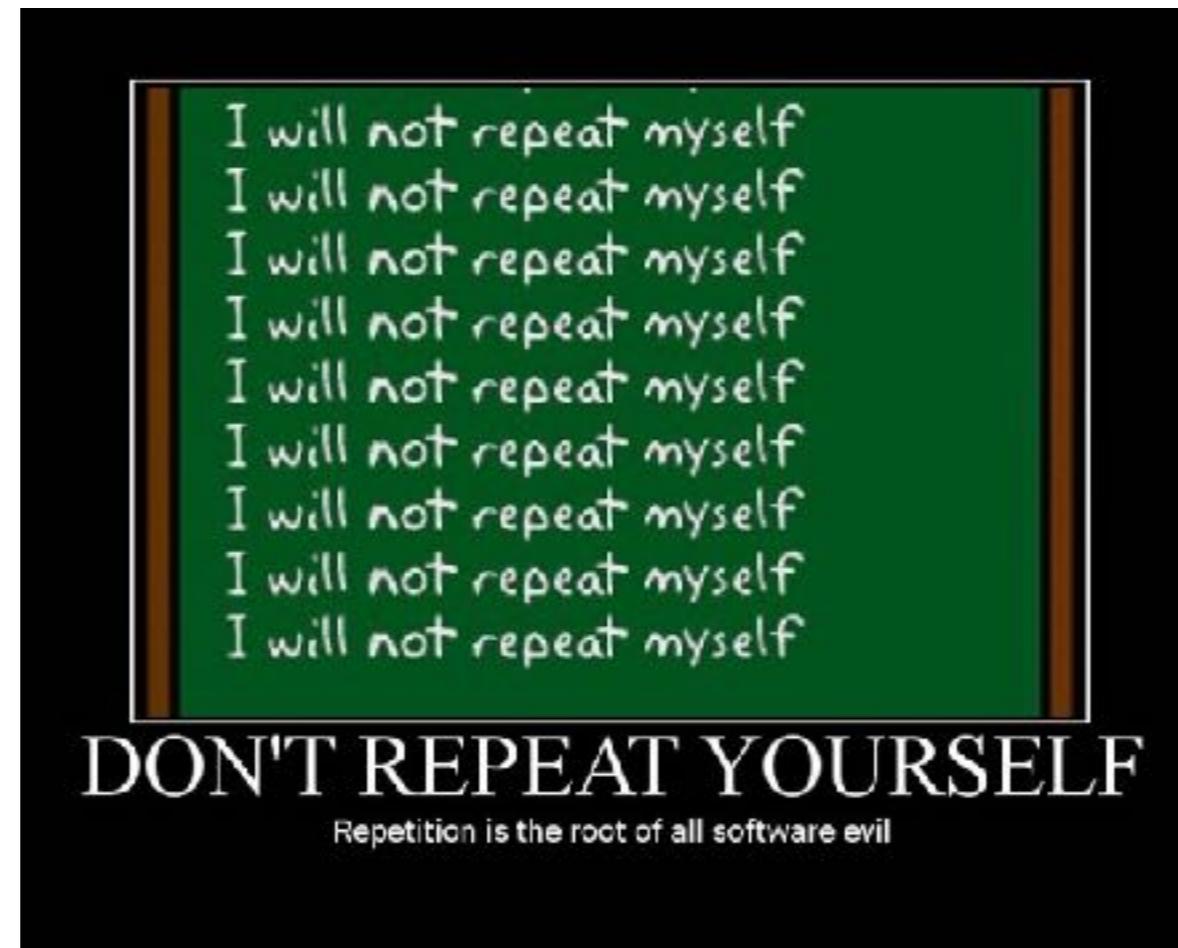
DM-3: Choosing the right upstream

DM-4: Influencing the upstream

DM-5: Estimating the impact of changes



The DRY Principle



But what about DRO?

Smalltalk Ecosystem: 14% of code cloned*!

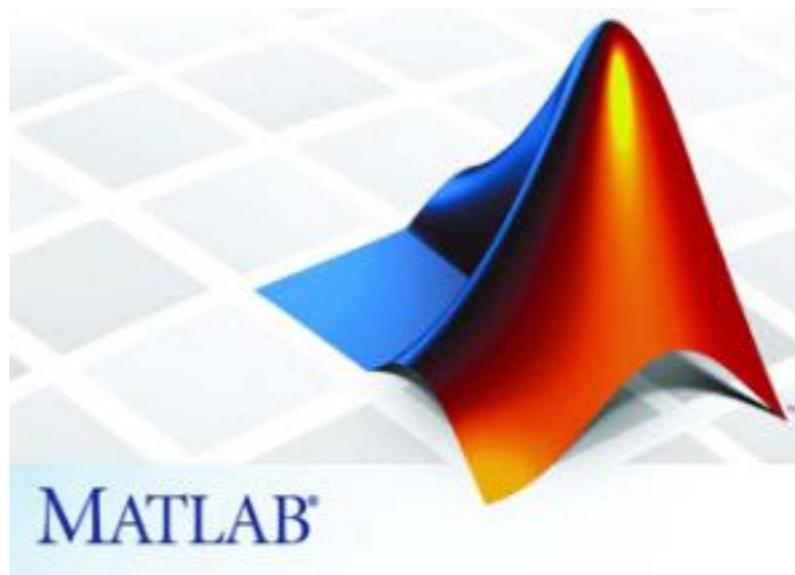
* On How Often Code is Cloned Across Repositories.

Schwarz, Lungu, Robbes. ICSE NIER 2012



designed by freepik.com

Challenge: establishing the right ecosystem norms



back to this later...

The Systemic Context

The ecosystem: opportunities

#1 — usability of our tools++

The screenshot shows a web browser window displaying the Nielsen Norman Group (NN/g) website. The URL in the address bar is <https://www.nngroup.com/articles/alphabetical-sorting-must-mostly-die/>. The page title is "Alphabetical Sorting Must Die". The main navigation menu includes links for HOME, TRAINING, CONSULTING, REPORTS, ARTICLES (which is underlined), and ABOUT NN/G. On the left side, there is a sidebar with sections for Topics (E-commerce, Intranets, Mobile & Tablet, User Testing, Web Usability, See all topics...) and Author (Jakob Nielsen, Don Norman, Bruce "Tog" Tognazzini, See all authors...). The main content area features the article's title, author (Jakob Nielsen), date (October 4, 2010), and summary. The summary text is: "Summary: Ordinal sequences, logical structuring, time lines, or prioritization by importance or frequency are usually better than A-Z listings for presenting options to users." Below the summary, it says: "Sorting a list of options alphabetically has two main benefits: • If users know the name of the thing they want, they can usually find it in the list pretty quickly. • Lazy design teams don't have to work on figuring out a better structure. Because we all know our ABCs anybody can put the". The browser interface includes a search bar and various toolbar icons.

N N/g Nielsen Norman Group

Evidence-Based User Experience Research,
Training, and Consulting

Search

HOME TRAINING CONSULTING REPORTS ARTICLES ABOUT NN/G

Topics

E-commerce
Intranets
Mobile & Tablet
User Testing
Web Usability
[See all topics...](#)

Author

Jakob Nielsen
Don Norman
Bruce "Tog" Tognazzini
[See all authors...](#)

Recent Articles

Alphabetical Sorting Must Die

by [JAKOB NIELSEN](#) on October 4, 2010

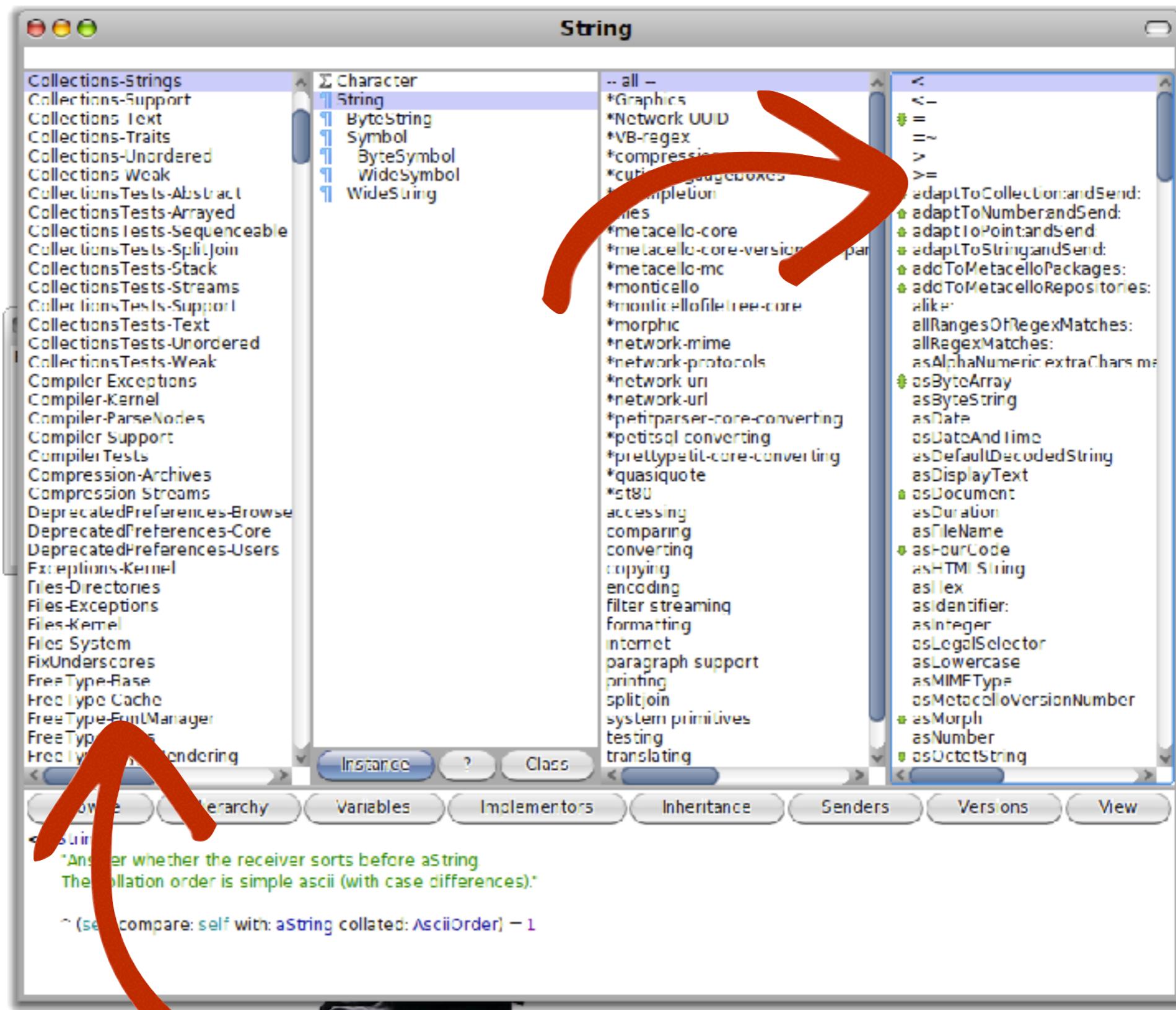
Topics: [Information Architecture](#) [Navigation](#)

Summary: Ordinal sequences, logical structuring, time lines, or prioritization by importance or frequency are usually better than A-Z listings for presenting options to users.

Sorting a list of options alphabetically has two main benefits:

- If users know the name of the thing they want, they can usually find it in the list pretty quickly.
- Lazy design teams don't have to work on figuring out a better structure. Because we all know our ABCs anybody can put the

The Refactoring Browser



Java™ Platform
Standard Ed. 6[All Classes](#)

Packages

[java.applet](#)
[java.awt](#)
[java.awt.color](#)
[java.awt.datatransfer](#)
[java.awt.dnd](#)
[java.awt.event](#)
[java.awt.font](#)[All Classes](#)[AbstractAction](#)
[AbstractAnnotationValueVisitor6](#)[AbstractBorder](#)
[AbstractButton](#)
[AbstractCellEditor](#)
[AbstractCollection](#)
[AbstractColorChooserPanel](#)
[AbstractDocument](#)
[AbstractDocument.AttributeContext](#)
[AbstractDocument.Content](#)
[AbstractDocument.ElementEdit](#)
[AbstractElementVisitor6](#)
[AbstractExecutorService](#)
[AbstractInterruptibleChannel](#)
[AbstractLayoutCache](#)
[AbstractLayoutCache.NodeDimensions](#)
[AbstractList](#)
[AbstractListModel](#)
[AbstractMap](#)
[AbstractMap.SimpleEntry](#)
[AbstractMap.SimpleImmutableEntry](#)
[AbstractMarshallerImpl](#)
[AbstractMethodError](#)
[AbstractOwnableSynchronizer](#)
[AbstractPreferences](#)
[AbstractProc](#)
[AbstractQu](#)
[AbstractURLConnectionSynchronizer](#)
[AbstractReadSyncnizer](#)
[AbstractScriptEngine](#)[Overview](#) [Package](#) [Class](#) [Use](#) [Tree](#) [Deprecated](#) [Index](#) [Help](#)[PREV CLASS](#) [NEXT CLASS](#)[SUMMARY](#) [NESTED](#) [FIELD](#) [CONSTR](#) [METHOD](#)[FRAMES](#) [NO FRAMES](#)[DETAIL](#) [FIELD](#) [CONSTR](#) [METHOD](#)

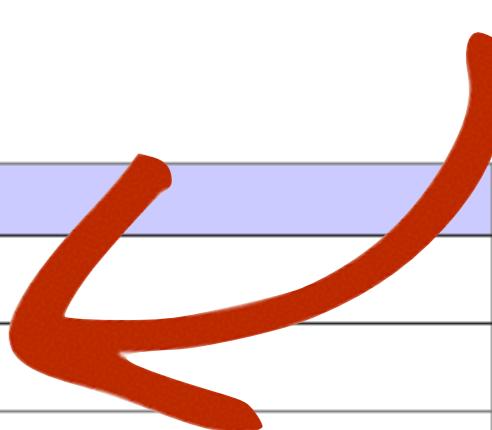
java.lang

Class String

[java.lang.Object](#)
└ [java.lang.String](#)**All Implemented Interfaces:**[Serializable](#), [CharSequence](#), [Comparable<String>](#)

Method Summary

<code>char</code>	<code>charAt(int index)</code>	Returns the <code>char</code> value at the specified index.
<code>int</code>	<code>codePointAt(int index)</code>	Returns the character (Unicode code point) at the specified index.
<code>int</code>	<code>codePointBefore(int index)</code>	Returns the character (Unicode code point) before the specified index.
<code>int</code>	<code>codePointCount(int beginIndex, int endIndex)</code>	Returns the number of Unicode code points in the specified text range of this <code>String</code> .
<code>int</code>	<code>compareTo(String anotherString)</code>	Compares two strings lexicographically.
<code>int</code>	<code>compareToIgnoreCase(String str)</code>	Compares two strings lexicographically, ignoring case differences.
<code>String</code>	<code>concat(String str)</code>	Concatenates the specified string to the end of this string.
<code>boolean</code>	<code>contains(CharSequence s)</code>	Returns true if and only if this string contains the specified sequence of char values.
<code>boolean</code>	<code>contentEquals(CharSequence cs)</code>	Compares this string to the specified <code>CharSequence</code> .
<code>boolean</code>	<code>contentEquals(StringBuffer sb)</code>	Compares this string to the specified <code>StringBuffer</code> .
<code>static String</code>	<code>copyValueOf(char[] data)</code>	Returns a <code>String</code> that represents the character sequence in the array specified.
<code>static String</code>	<code>copyValueOf(char[] data, int offset, int count)</code>	Returns a <code>String</code> that represents the character sequence in the array specified.
<code>boolean</code>	<code>endsWith(String suffix)</code>	Tests if this string ends with the specified suffix.
<code>boolean</code>	<code>equals(Object anObject)</code>	Compares this string to the specified object.



[Collapse All](#) [Export \(0\)](#) [Print](#)

- [MSDN Library](#)
- [.NET Development](#)
- [.NET Framework 4.5](#)
- [.NET Framework Class Library](#)
- [System](#)

String Class

- [String Constructor](#)
- [String Fields](#)
- [String Methods](#)
- [String Operators](#)
- [String Properties](#)



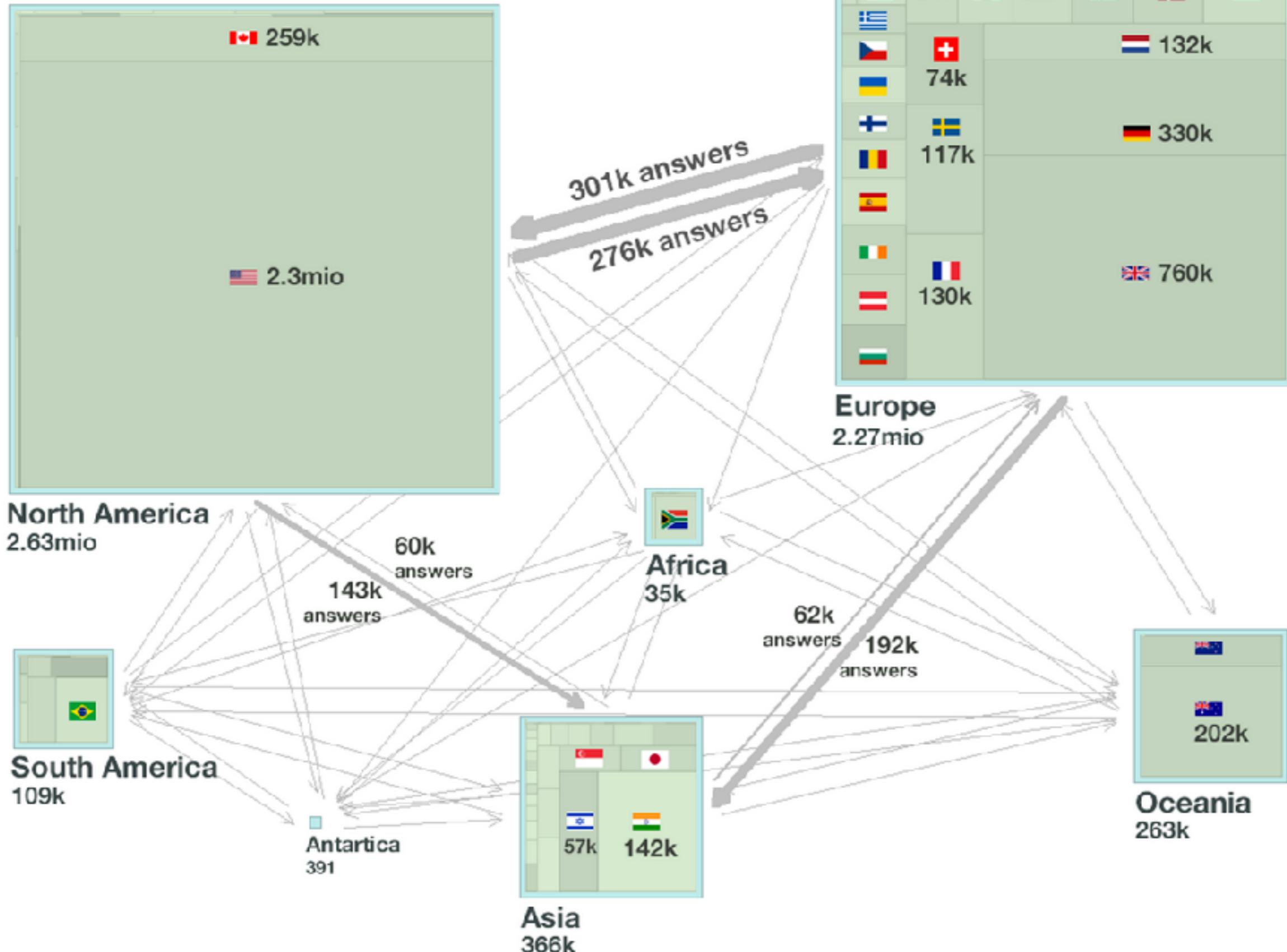
String Class

.NET Framework 4.5 | [Other Versions](#) | 37 out of 51 rated this helpful – [Rate this topic](#)

Show: Inherited Protected

	Name	Description
	Clone	Returns a reference to this instance of String.
	Compare(String, String)	Compares two specified String objects and returns an Integer that indicates their relative position in the sort order.
	Compare(String, String, Boolean)	Compares two specified String objects, ignoring or honoring their case, and returns an Integer that indicates their relative position in the sort order.
	Compare(String, String, StringComparison)	Compares two specified String objects using the specified rules, and returns an integer that indicates their relative position in the sort order.
	Compare(String, String, CultureInfo)	Compares two specified String objects, ignoring or honoring their case, and using culture-specific information to influence the comparison, and returns an integer that indicates their relative position in the sort order.
	Compare(String, String, CultureInfo, CompareOptions)	Compares two specified String objects using the specified comparison options and culture-specific information to influence the comparison, and returns an Integer that indicates the relationship of the two strings to each other in the sort order.
	Compare(String, Int32, String)	Compares substrings of two specified String objects and returns an

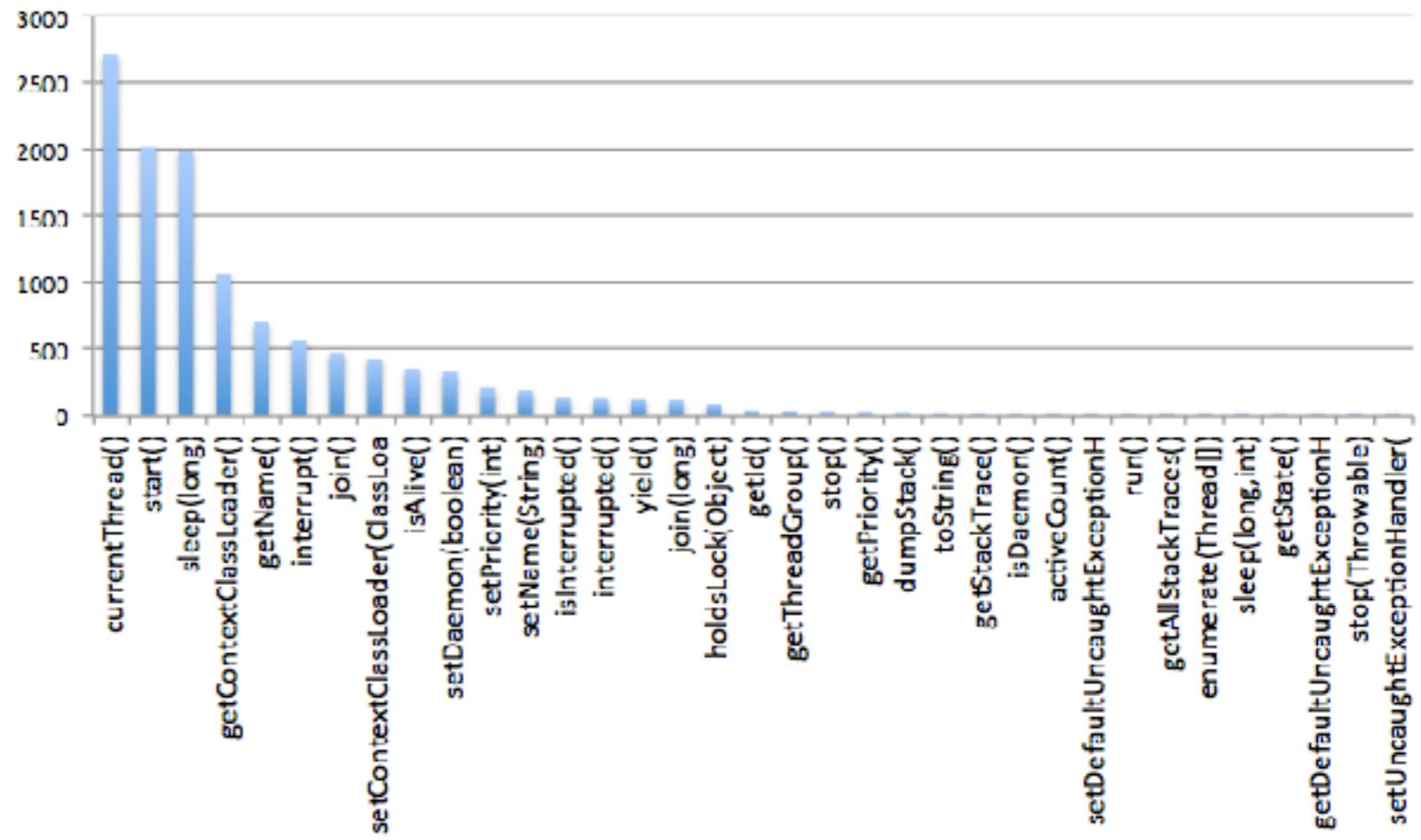






[**Pangea**](#) streamlines analyzing
multiple systems with Moose.

Usage of `java.lang.Thread` in the ecosystem





Augmenting JavaDoc in Chrome

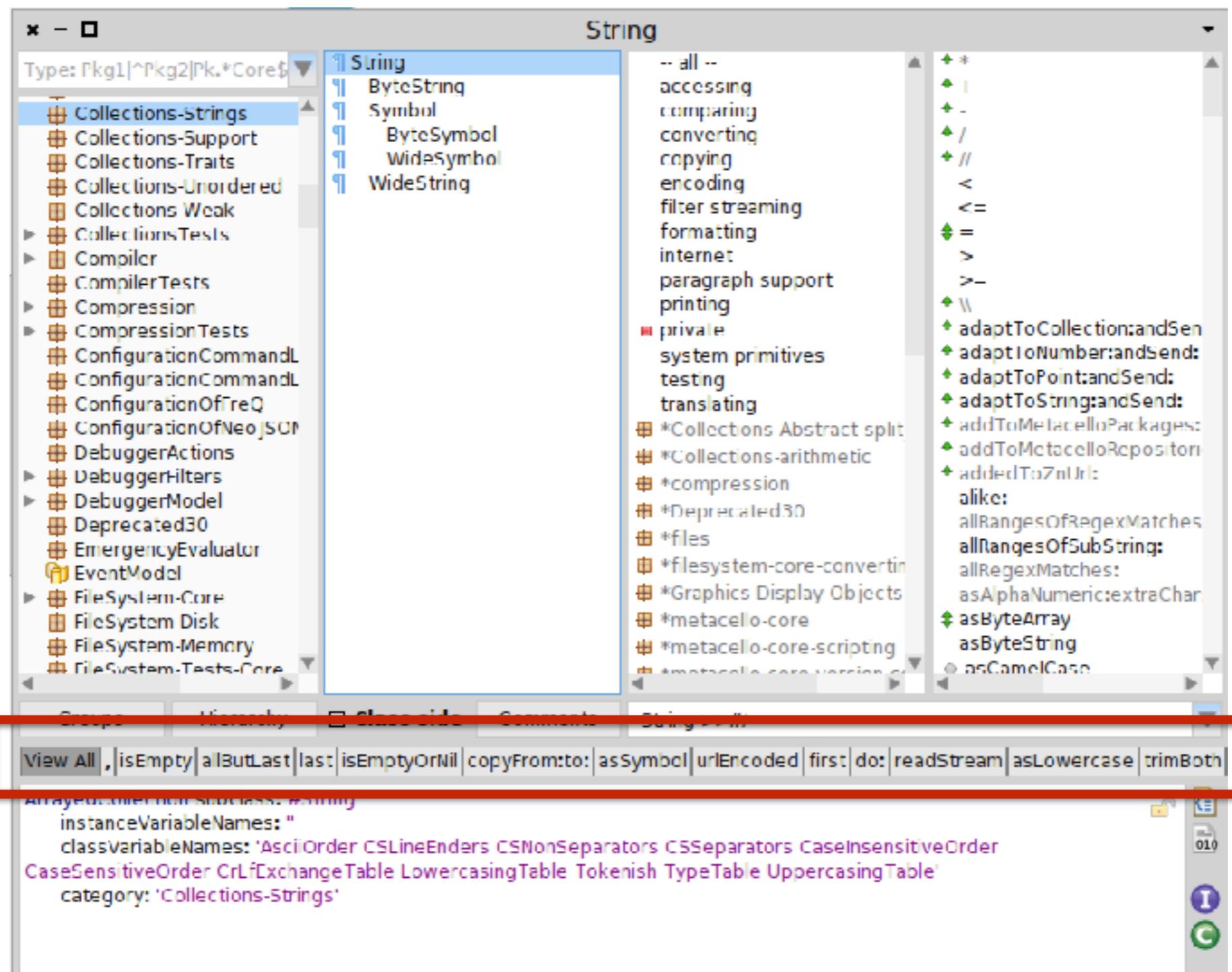
The screenshot shows a JavaDoc page for the `String` class in the `java.lang` package. The page includes navigation links like Overview, Package, Class, Use, Tree, Deprecated, Index, Help, and links for Prev Class, Next Class, Frames, No Frames, and All Classes. The summary section shows nested classes, fields, constructors, and methods. A red box highlights a sidebar titled "jatis - Javadoc statistics" which contains a table of methods categorized by their return type and parameters.

Modifier and Type	Method and Description
<code>boolean</code>	<code>equals(Object anObject)</code> Compares this string to the specified object.
<code>int</code>	<code>length()</code> Returns the length of this string.
<code>boolean</code>	<code>startsWith(String prefix)</code> Tests if this string starts with the specified prefix.
<code>char</code>	<code>charAt(int index)</code> Returns the char value at the specified index.
<code>String</code>	<code>trim()</code> Returns a copy of the string, with leading and trailing whitespace omitted.
<code>boolean</code>	<code>equalsIgnoreCase(String anotherString)</code> Compares this String to another String, ignoring case considerations.
<code>String</code>	<code>substring(int beginIndex)</code> Returns a new string that is a substring of this string.

Overthrowing the Tyranny of Alphabetical Ordering in Documentation Systems. Boris Spasojević, Mircea Lungu, and Oscar Nierstrasz. In 2014 IEEE International Conference on Software Maintenance and Evolution (ERA Track), p. 511-515, September 2014.



Augmenting The Refactoring Browser



Overthrowing the Tyranny of Alphabetical Ordering in Documentation Systems. Boris Spasojević, Mircea Lungu, and Oscar Nierstrasz. In 2014 IEEE International Conference on Software Maintenance and Evolution (ERA Track), p. 511-515, September 2014.

#2 — intelligence of our tools++

Onward! 2014 – October 20–24, 2014, Portland, OR, USA

Mining the Ecosystem to Improve Type Inference For Dynamically Typed Languages

Boris Spasojević, Mircea Lungu, Oscar Nierstrasz

Software Composition Group
University of Bern
Switzerland

{spasojev,lungu,oscar}@iam.unibe.ch

Abstract

Dynamically typed languages lack information about the types of variables in the source code. Developers care about this information as it supports program comprehension. Basic type inference techniques are helpful, but may yield many false positives or negatives.

We propose to mine information from the software ecosys-

1. Introduction

Software developers spend more time on maintaining existing software than writing new code. Maintenance consumes over 70 percent of the total life-cycle of a software product [4]. This means that supporting and understanding code is very important. Static analysis in source code helps developers understand

#3 — understanding our organization++

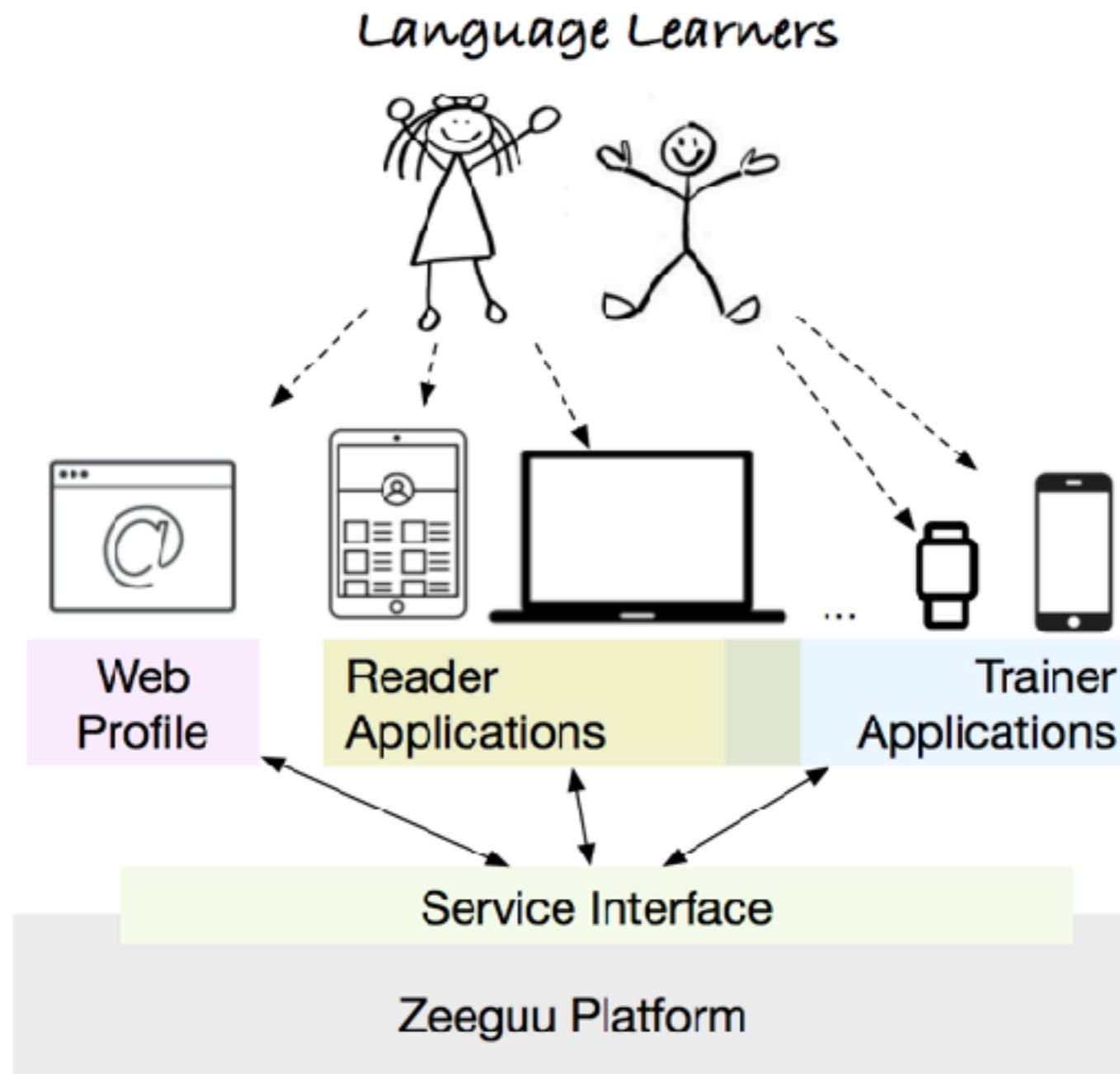
Most Popular Bug-Fixes

	Instances	Projects
↳ → <i>if(T != null)</i>	3,718	316
↳ → <i>if(T == null) return;</i>	1051	190
↳ → <i>if(T == null) return null;</i>	243	80
↳ → <i>if(T == null) throw new T();</i>	207	75
↳ → <i>if(T == null) T=T;</i>	157	67
↳ → <i>if(T == null) continue;</i>	82	34
Total	5,172	348

Mining frequent bug-fix code changes. Haidar Osman, Mircea Lungu, and Oscar Nierstrasz. In Software Maintenance, Reengineering and Reverse Engineering (CSMR-WCRE), 2014 Software Evolution Week - IEEE Conference on, p. 343-347, February 2014

The Personal Context

Accelerating Second Language Acquisition

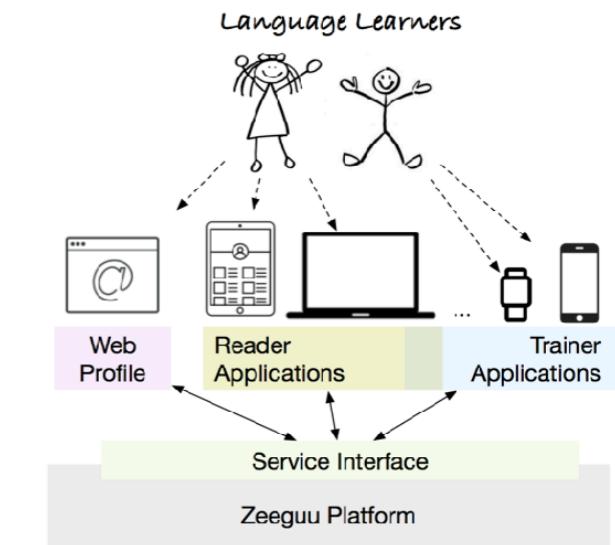
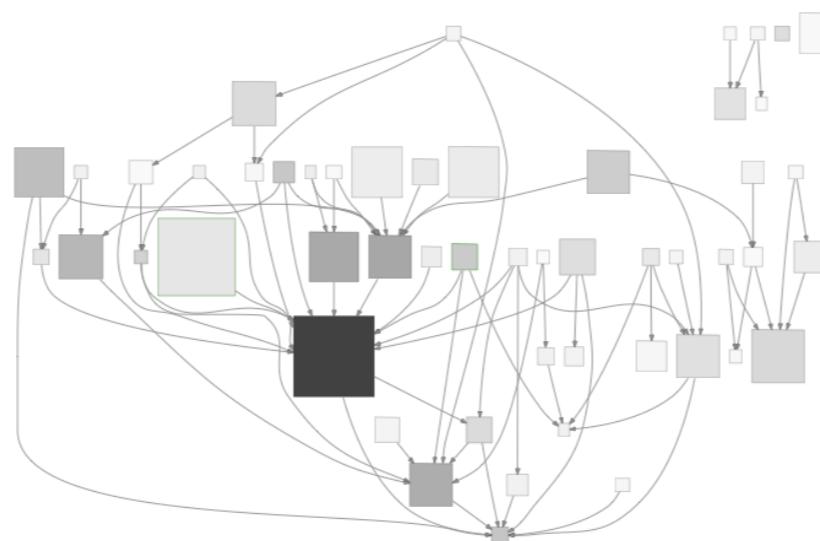
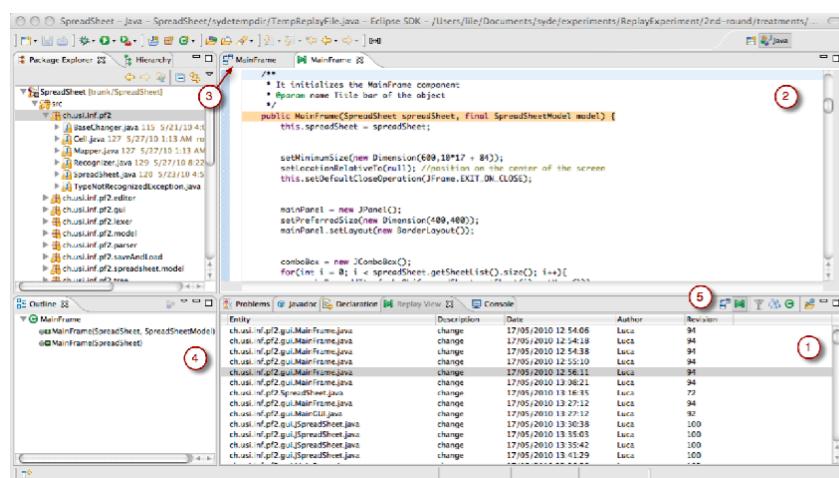


Code Reading Recommender

based on subjective complexity



Context in Software Analytics



Temporal (eco)Systemic Personal

@mircealungu
<https://mircealungu.github.io>