

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

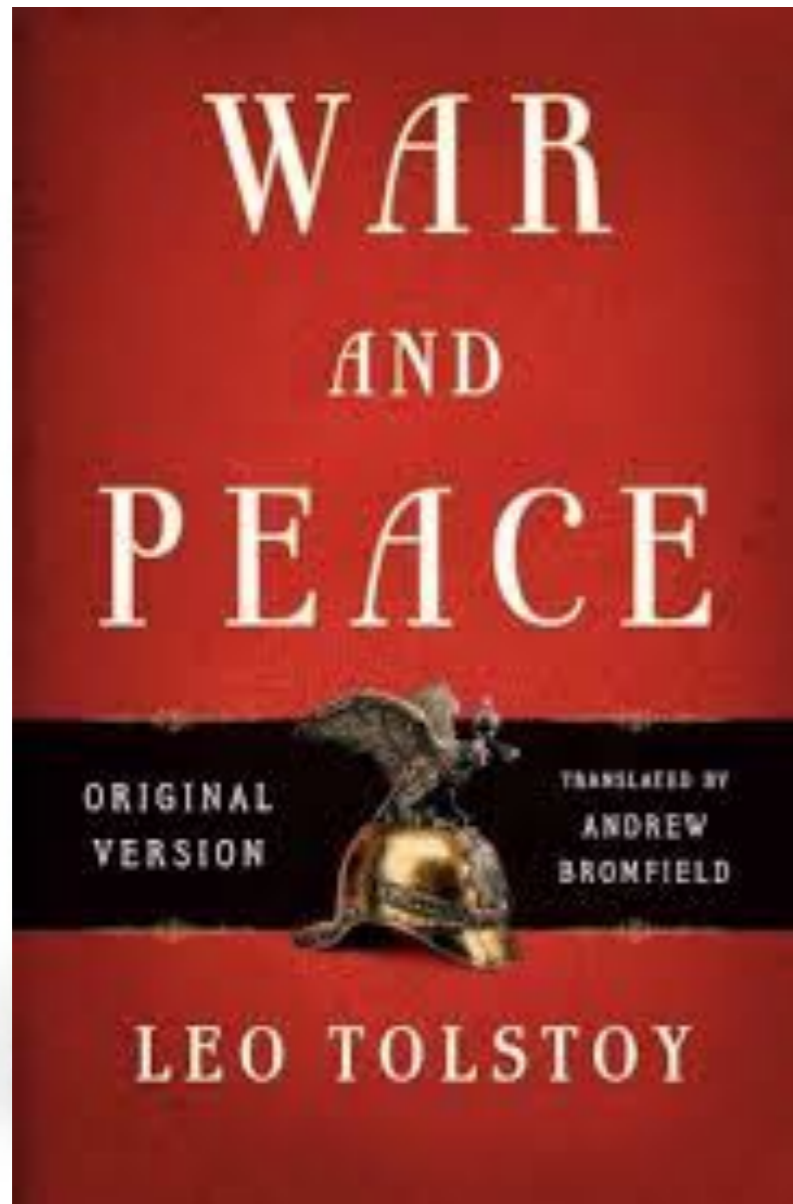
Mircea F. Lungu

SEARCH

JBI

University of Groningen

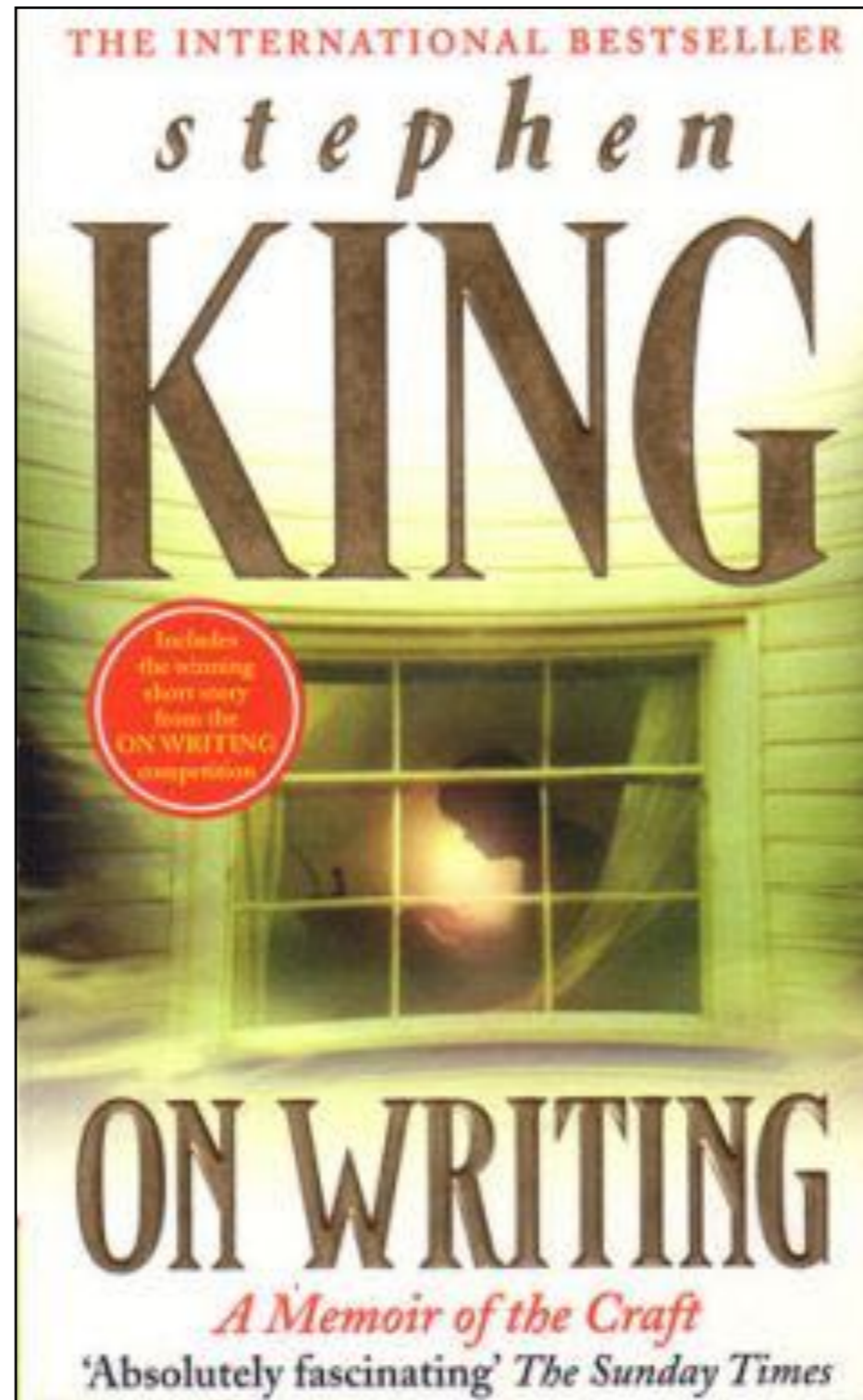
January 2017, SEN Symposium, CWI, Amsterdam



60'000 LOT



60'000 LOC



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



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

The screenshot shows the Eclipse IDE interface. The Package Explorer on the left shows the project structure. The Outline view at the bottom left shows the 'MainFrame' class. The main editor displays the source code for 'MainFrame.java'. The Replay View at the bottom right shows a table of development sessions.

Entity	Description	Date	Author	Revision
ch.usi.inf.pf2.gui.MainFrame.java	change	17/05/2010 12:54:06	Luca	94
ch.usi.inf.pf2.gui.MainFrame.java	change	17/05/2010 12:54:18	Luca	94
ch.usi.inf.pf2.gui.MainFrame.java	change	17/05/2010 12:54:38	Luca	94
ch.usi.inf.pf2.gui.MainFrame.java	change	17/05/2010 12:55:10	Luca	94
ch.usi.inf.pf2.gui.MainFrame.java	change	17/05/2010 12:56:11	Luca	94
ch.usi.inf.pf2.gui.MainFrame.java	change	17/05/2010 13:08:21	Luca	94
ch.usi.inf.pf2.SpreadSheet.java	change	17/05/2010 13:16:35	Luca	72
ch.usi.inf.pf2.gui.MainFrame.java	change	17/05/2010 13:27:12	Luca	94
ch.usi.inf.pf2.gui.MainGUI.java	change	17/05/2010 13:27:12	Luca	92
ch.usi.inf.pf2.gui.JSpreadSheet.java	change	17/05/2010 13:30:38	Luca	100
ch.usi.inf.pf2.gui.JSpreadSheet.java	change	17/05/2010 13:35:03	Luca	100
ch.usi.inf.pf2.gui.JSpreadSheet.java	change	17/05/2010 13:35:42	Luca	100
ch.usi.inf.pf2.gui.JSpreadSheet.java	change	17/05/2010 13:41:29	Luca	100

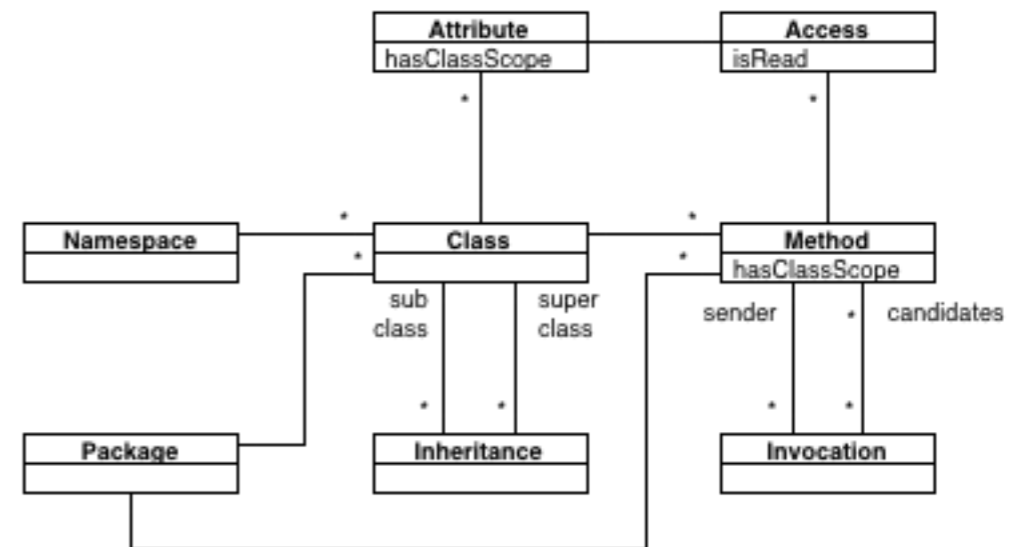
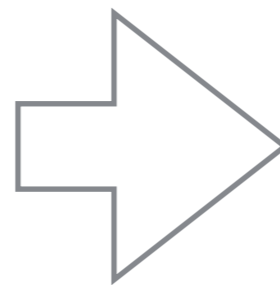
Temporal Context

long term

Software Analytics Platform

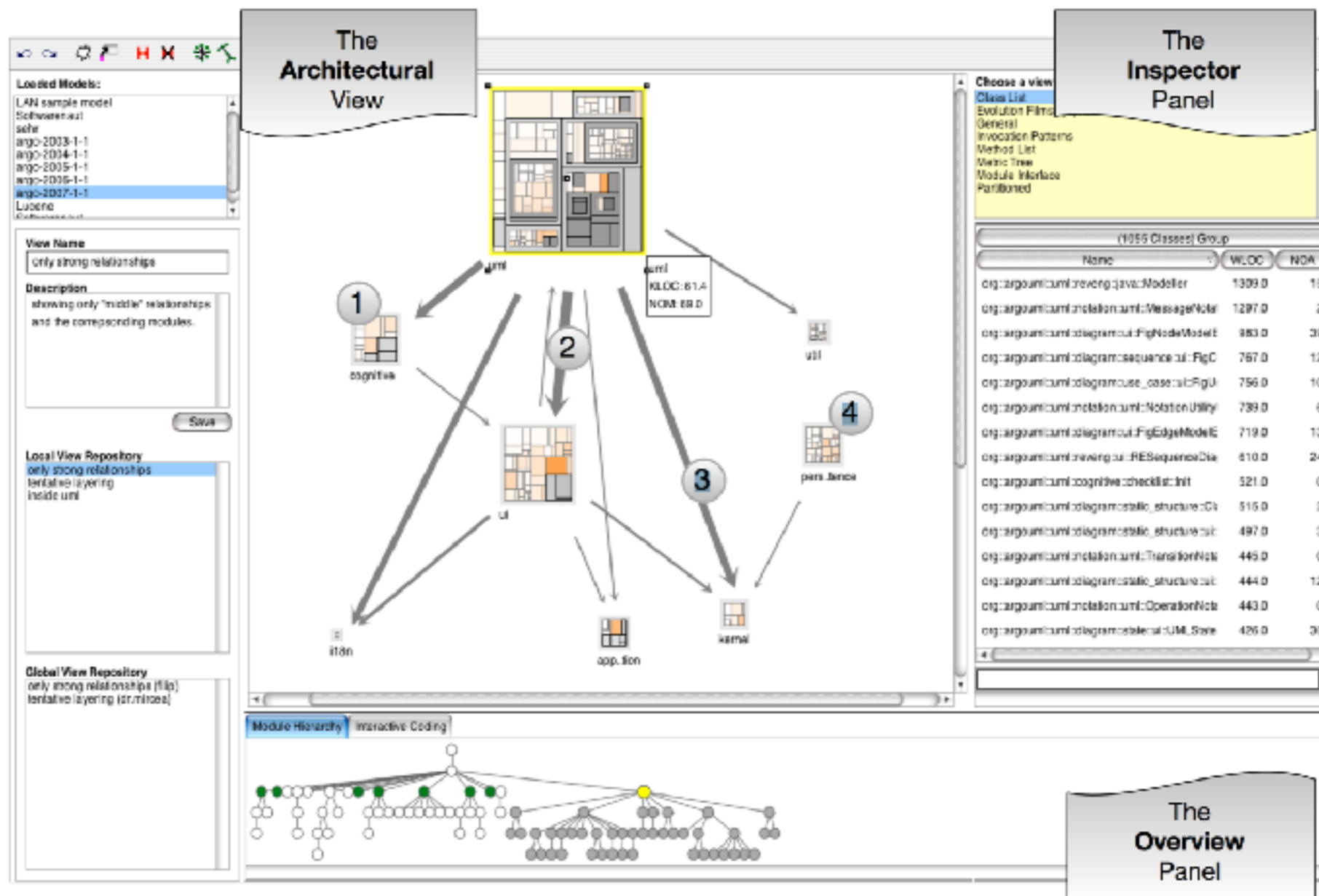


60'000 LOC



Language Independent
Meta-Model

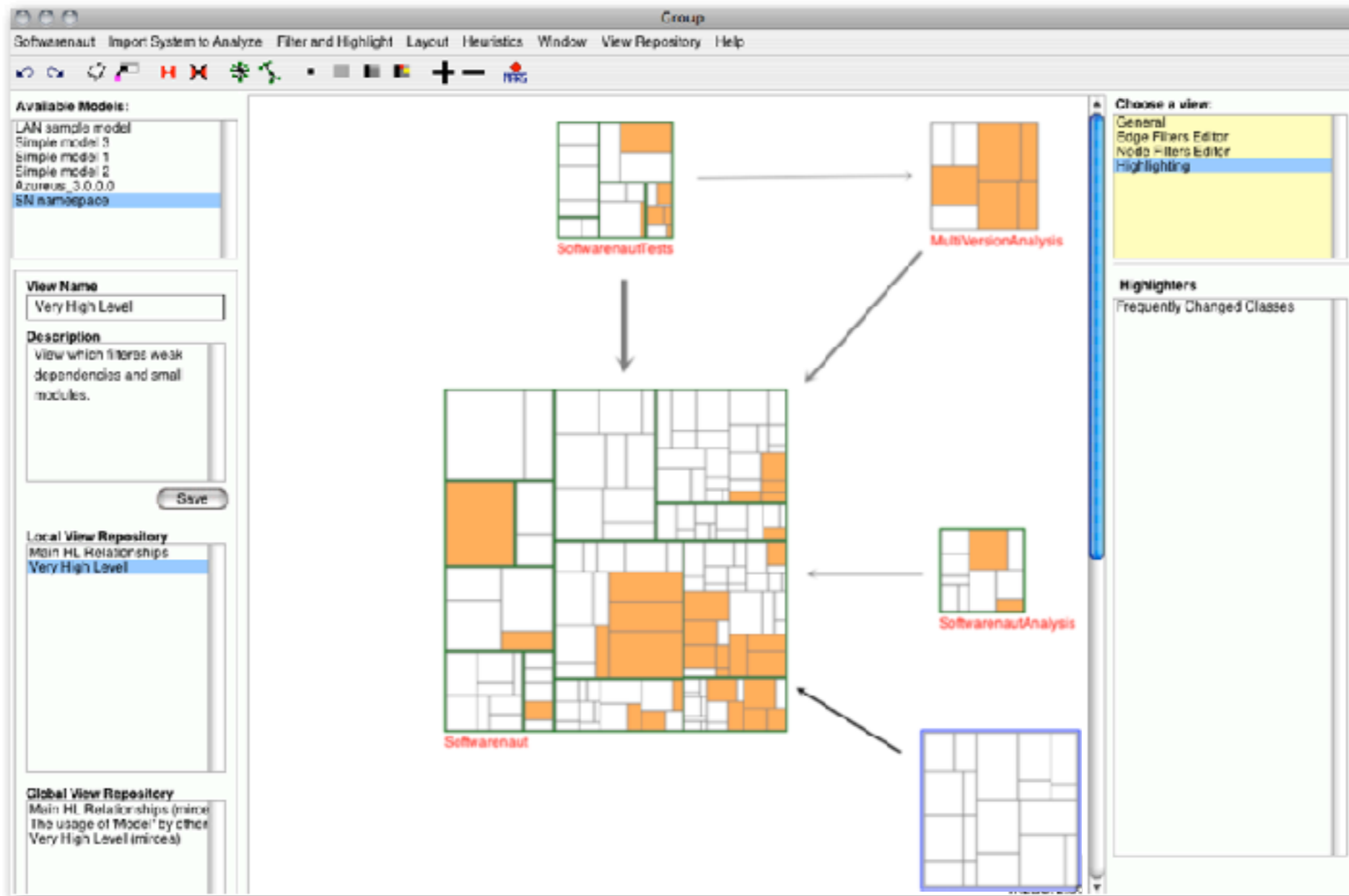
Interactive Architecture Recovery



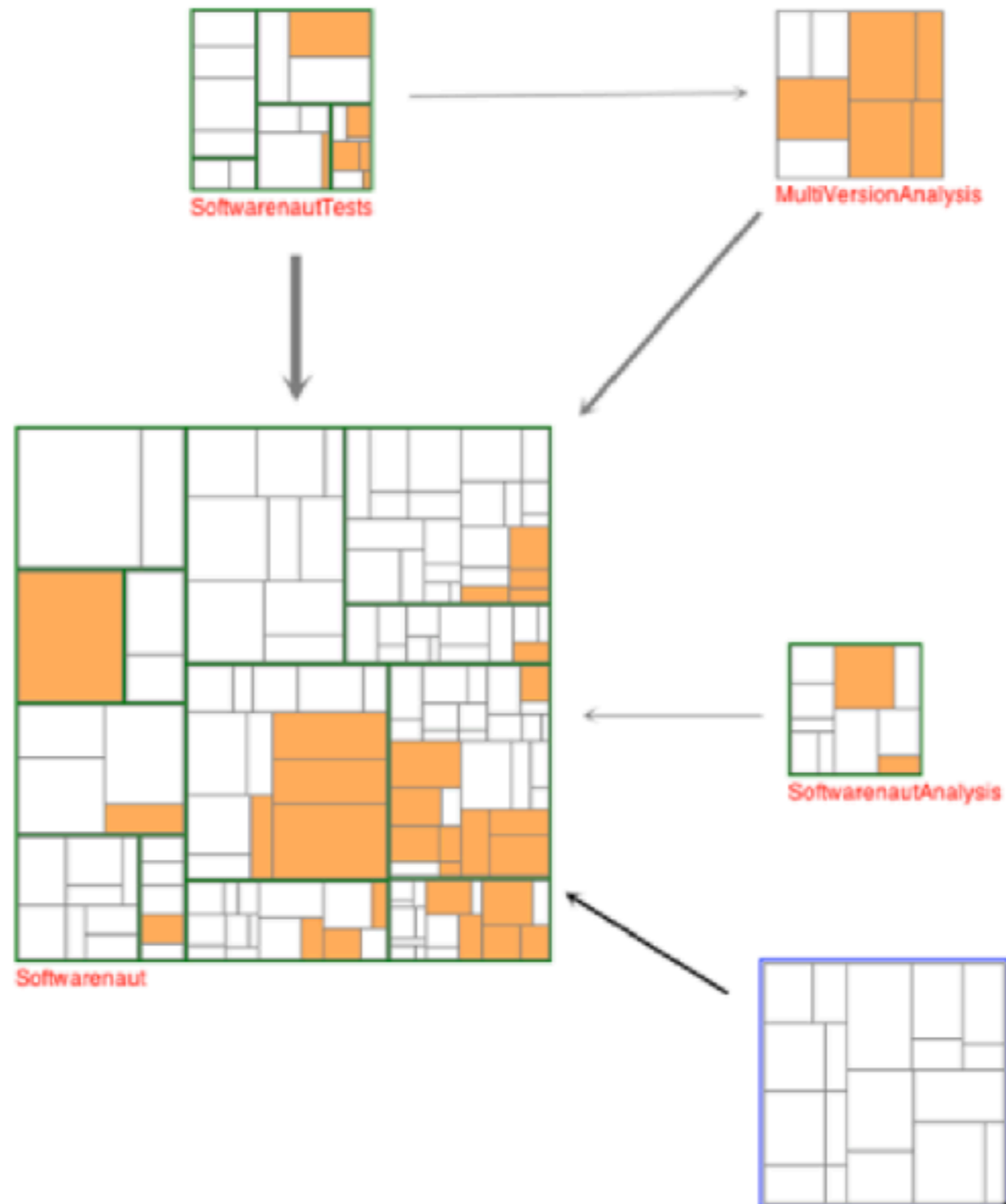
Evolutionary and Collaborative Software Architecture Recovery with Softwareenaut.

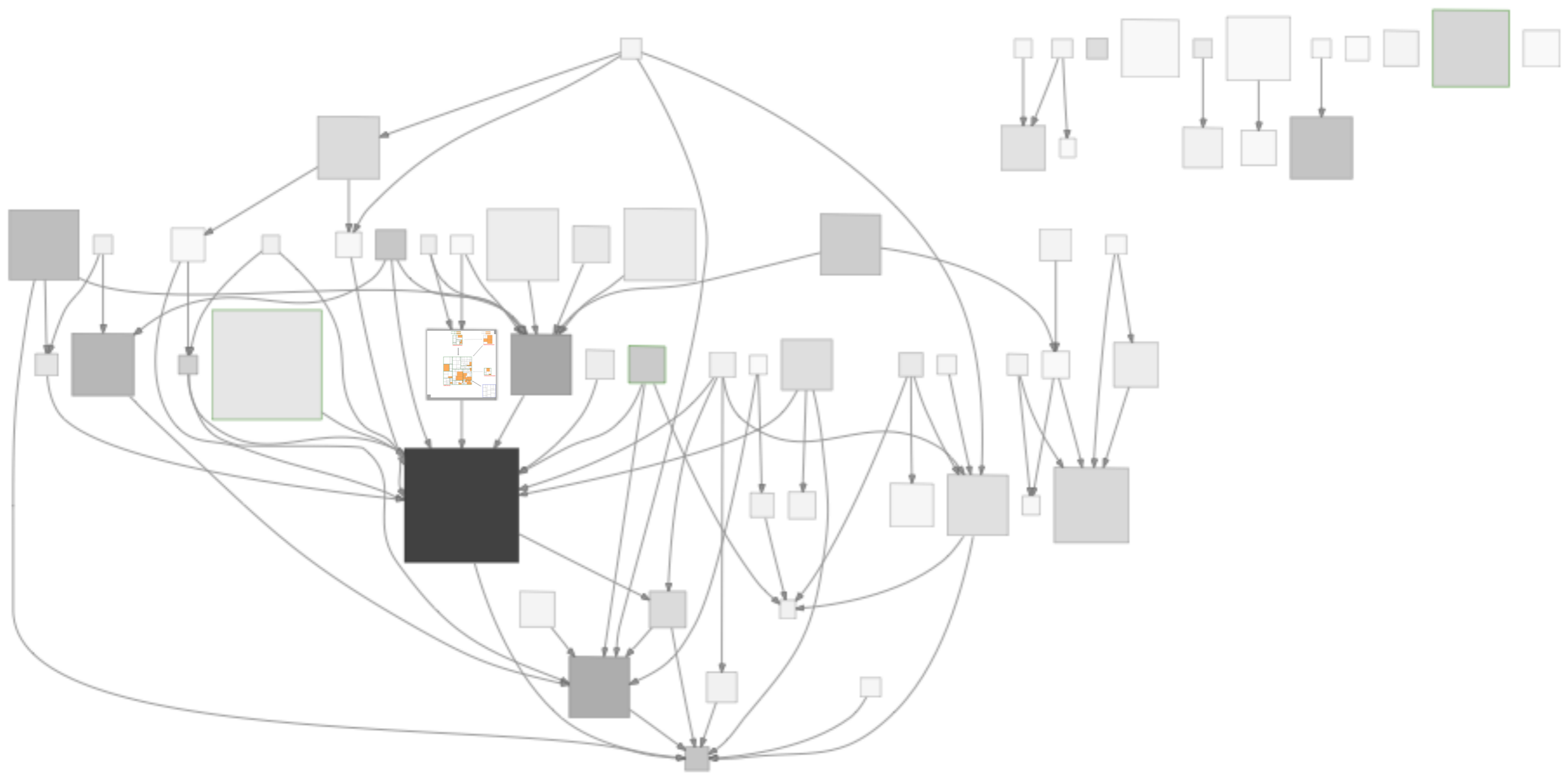
Mircea Lungu, Michele Lanza, and Oscar Nierstrasz. In Science of Computer Programming 79(0) p. 204 - 223, 2014.

Hotspots can focus the analysis process



“No system is an island, entire of itself”

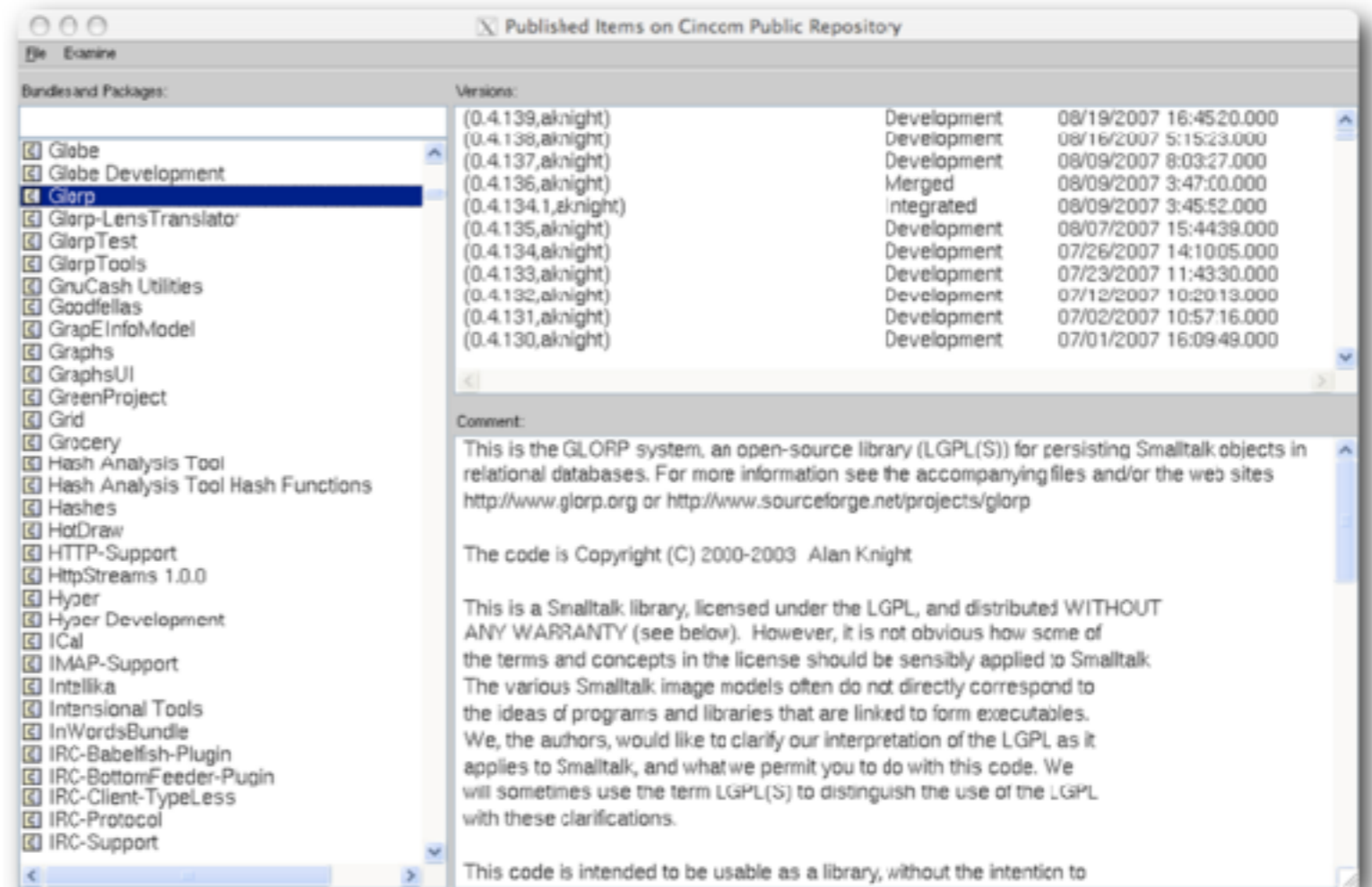
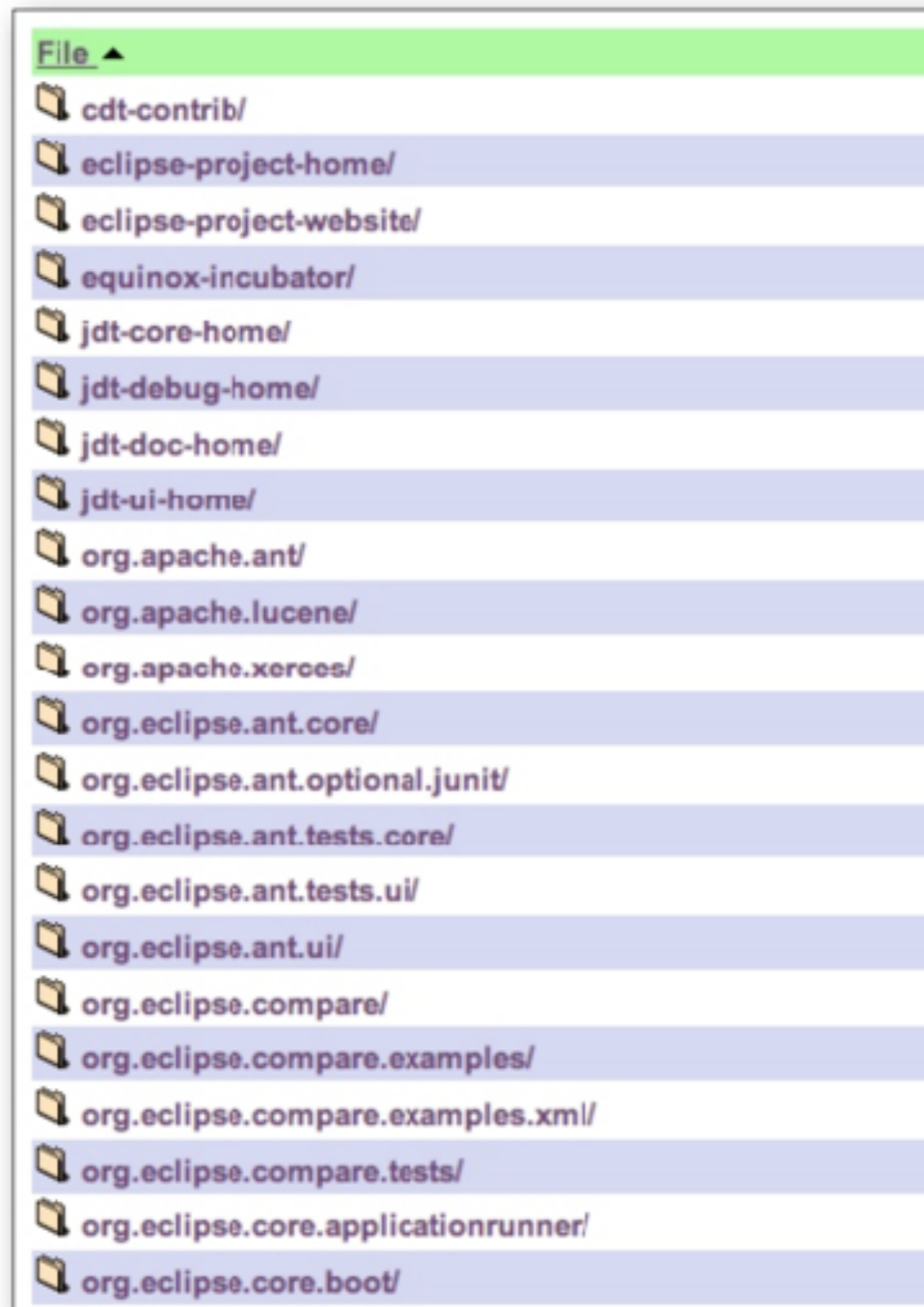




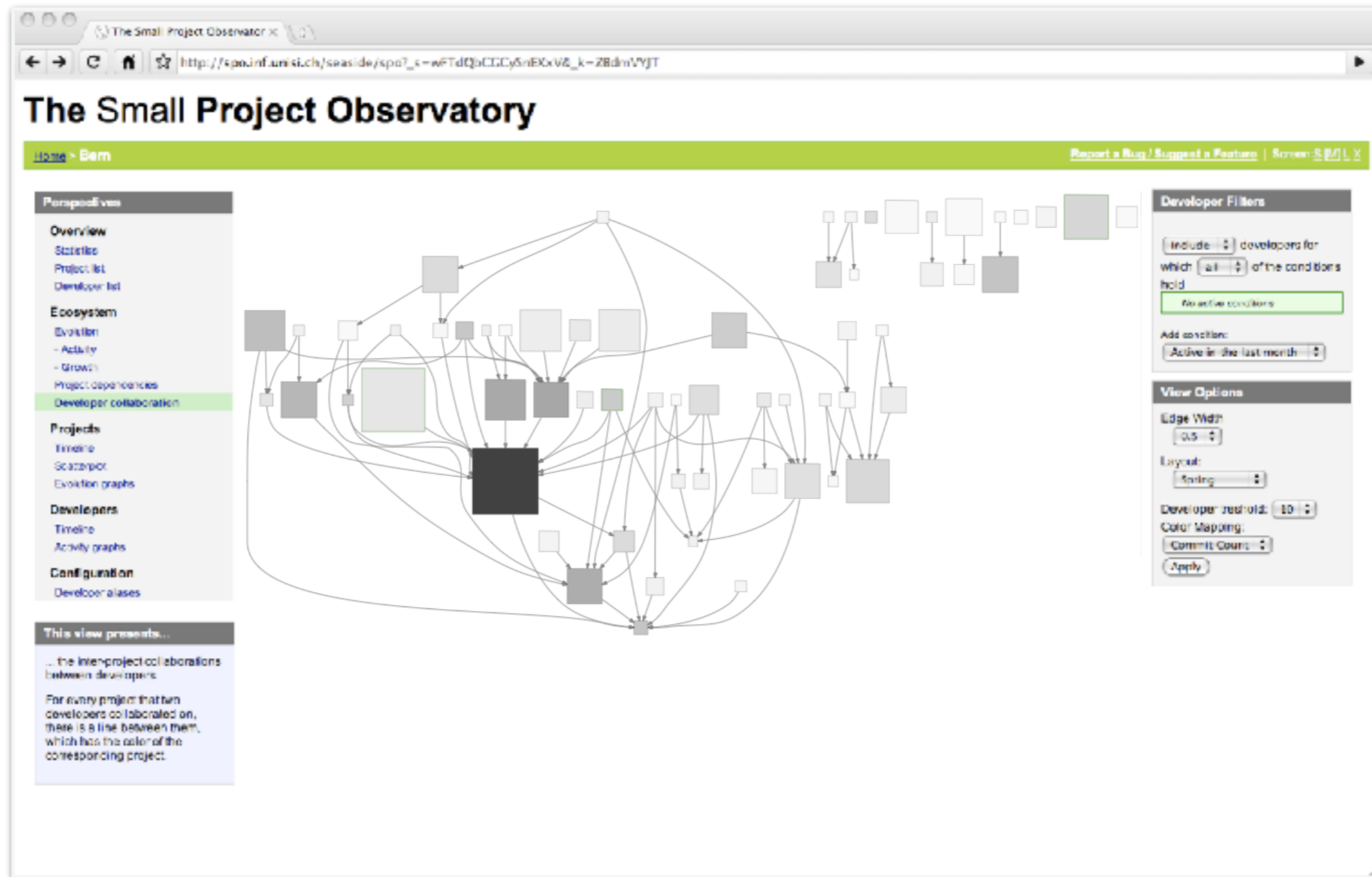
The Systemic Context

The ecosystem: challenges

And some organizations are still using tools like these!



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:

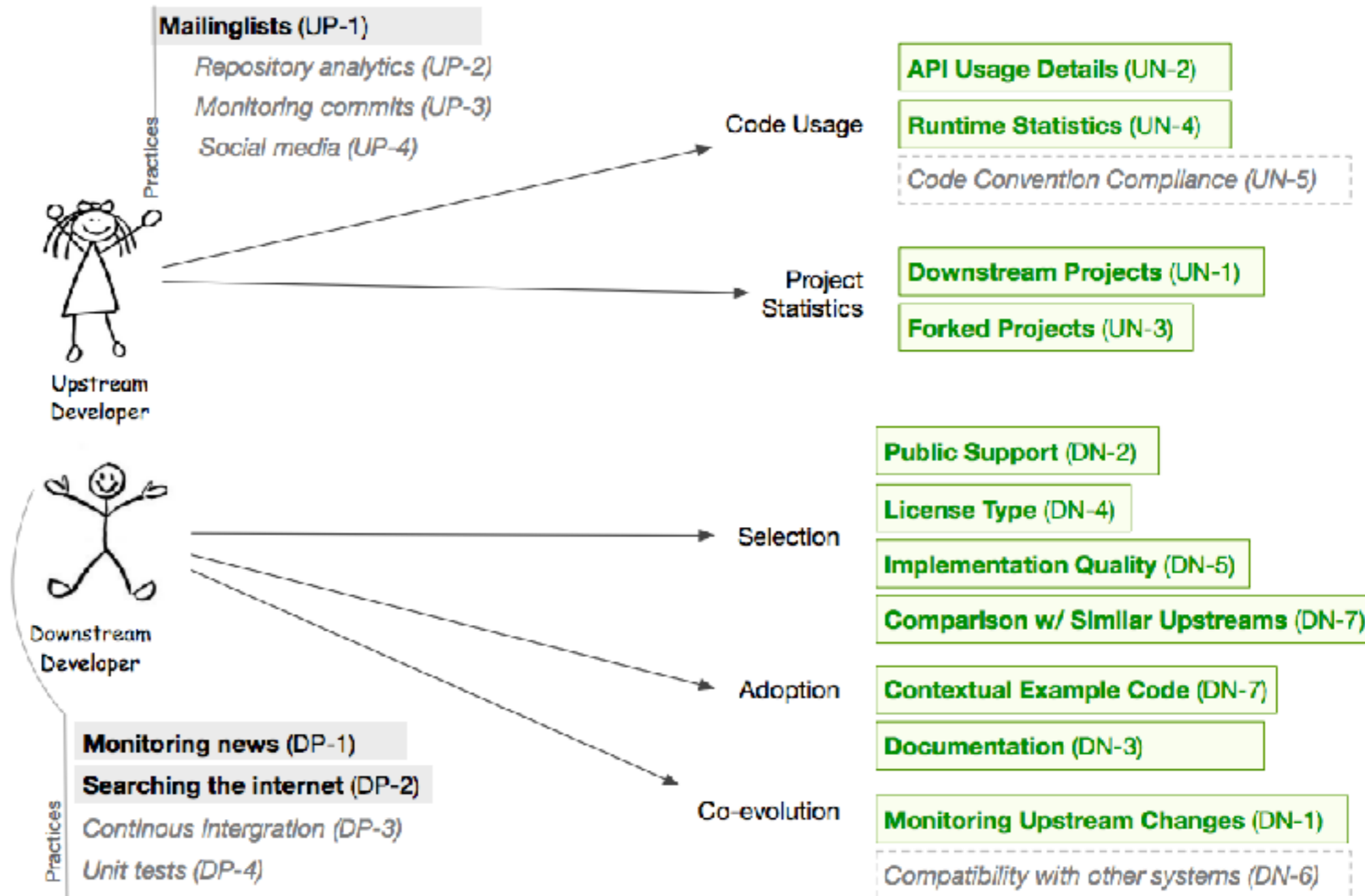
[...] where is **XXXXXXXXXX**, why was it deleted?? I use it and now is gone!!! I even had a specialization of it [...]

Developer:

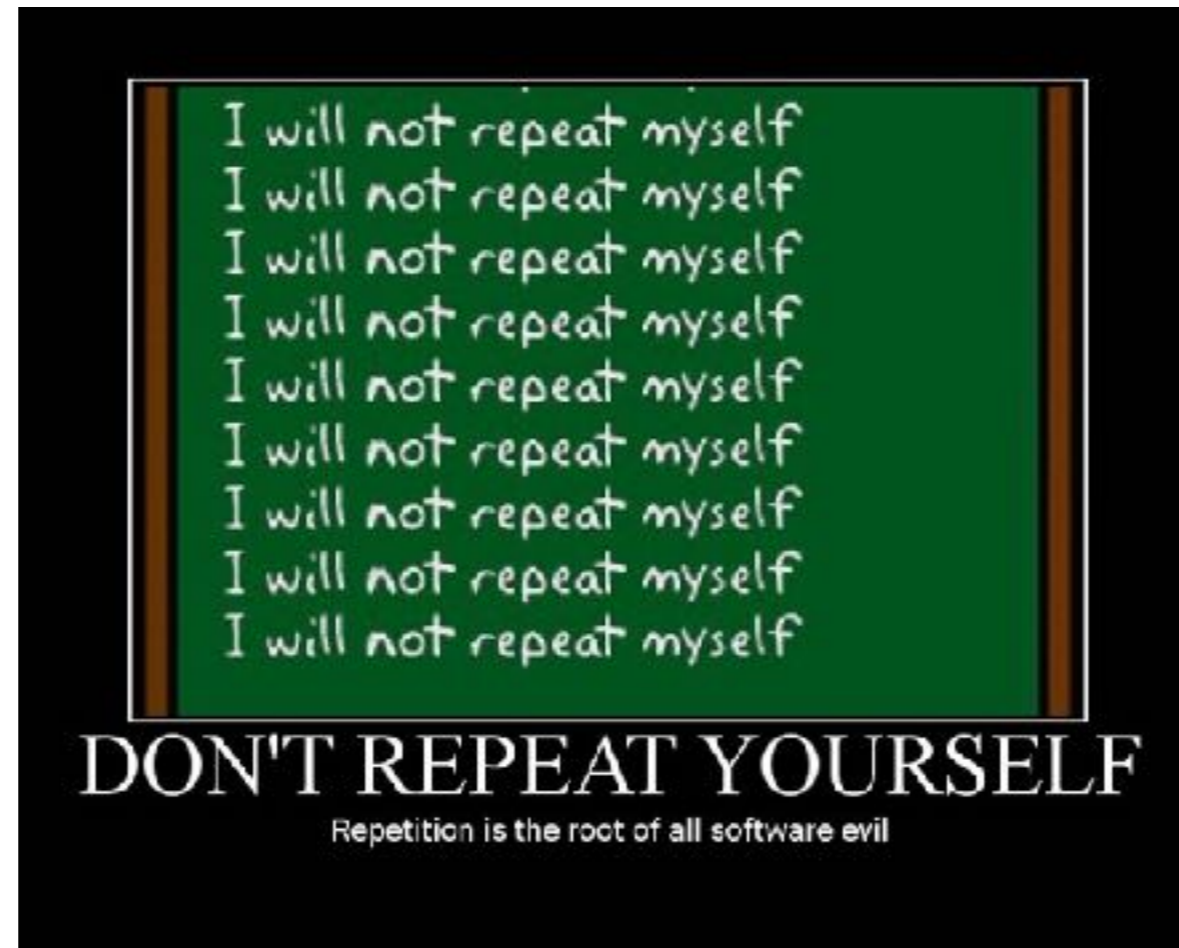
They have been dropped. A mail went out to this list if anybody still used them and nobody replied. [...] Personally I don't know of any application that uses these dialogs.

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



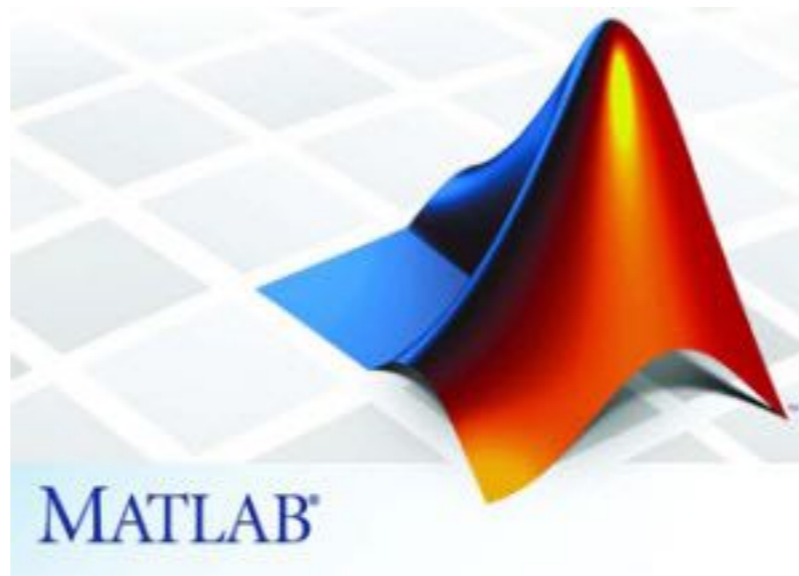
But what about DRO?

Smalltalk Ecosystem: 14% of code cloned*!

* On How Often Code is Cloned Across Repositories.
Schwarz, Lungu, Robbes. ICSE NIER 2012



Challenge: establishing the right ecosystem norms



back to this later...

The Systemic Context

The ecosystem: opportunities

#1 — usability of our tools++

The image shows a screenshot of a web browser displaying an article on the Nielsen Norman Group website. The browser's address bar shows the URL: <https://www.nngroup.com/articles/alphabetical-sorting-must-mostly-...>. The page header includes the NN/g logo and the text 'Nielsen Norman Group Evidence-Based User Experience Research, Training, and Consulting'. A search bar is visible on the right. The navigation menu includes 'HOME', 'TRAINING', 'CONSULTING', 'REPORTS', 'ARTICLES', and 'ABOUT NN/G'. The 'ARTICLES' menu item is highlighted with an orange underline. On the left side, there is a 'Topics' section with links for 'E-commerce', 'Intranets', 'Mobile & Tablet', 'User Testing', 'Web Usability', and 'See all topics...'. Below that is an 'Author' section with links for 'Jakob Nielsen', 'Don Norman', 'Bruce "Tog" Tognazzini', and 'See all authors...'. The main content area features the article title 'Alphabetical Sorting Must Die' in a large, bold font, which is highlighted with a thick orange border. Below the title, it says 'by JAKOB NIELSEN on October 4, 2010' and 'Topics: Information Architecture Navigation'. A red-bordered box highlights the following summary text: '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, the text reads 'Sorting a list of options alphabetically has two main benefits:' followed by a bulleted list: '• If users know the name of the thing they want, they can usually find it in the list pretty quickly.' and '• Lazy design teams don't have to work on figuring out a better structure. Because we all know our ABCs anybody can put the'.

Alphabetical Sorting Must Die

by JAKOB NIELSEN on October 4, 2010

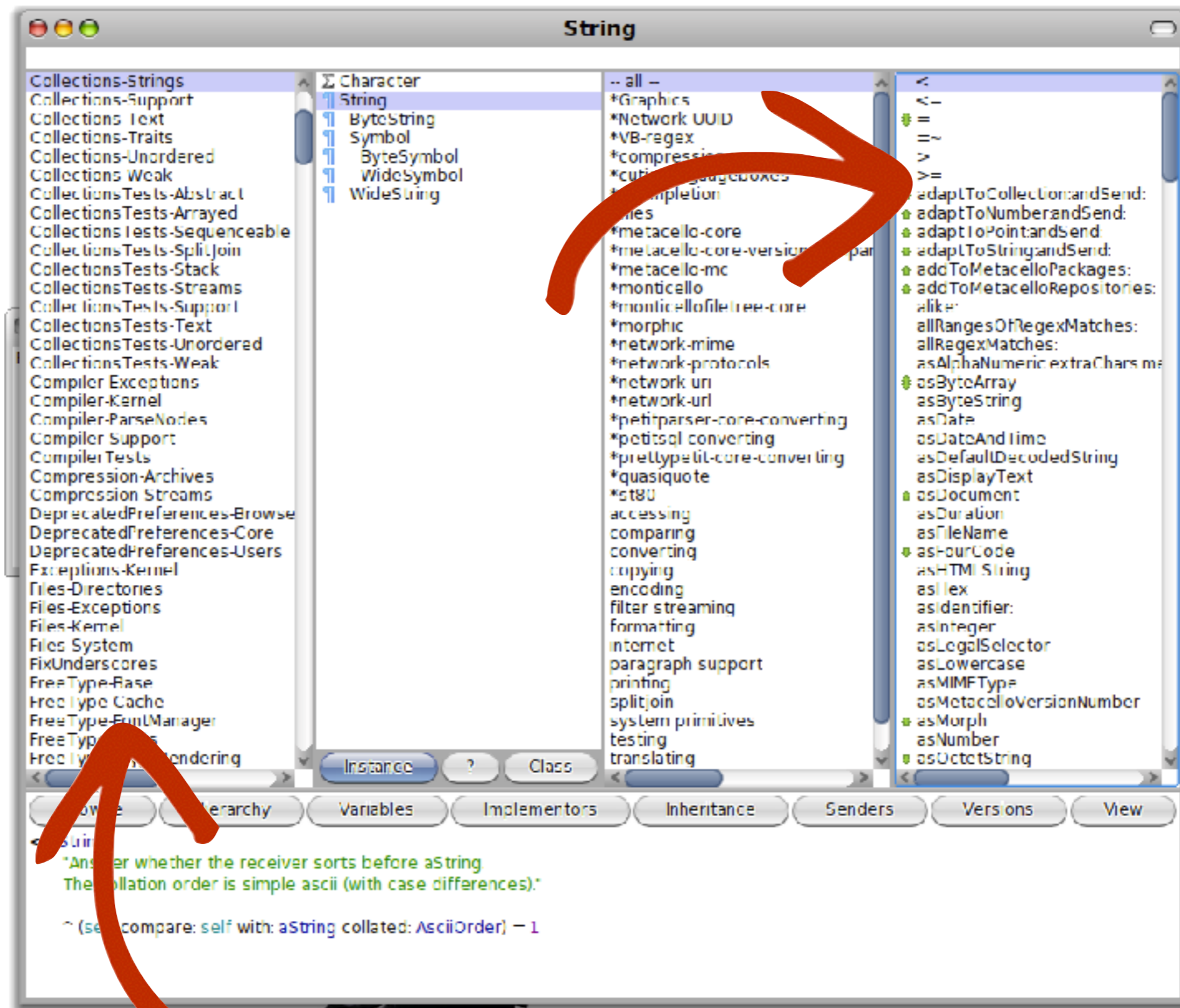
Topics: Information Architecture Navigation

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



- 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](#)
 - [AbstractMarshalerImpl](#)
 - [AbstractMethodError](#)
 - [AbstractOwnableSynchronizer](#)
 - [AbstractPreferences](#)
 - [AbstractProc...](#)
 - [AbstractQu...](#)
 - [AbstractQu...Synchronizer](#)
 - [Abstract...Synchronizer](#)
 - [Abstract...Engine](#)

Overview Package **Class** Use Tree Deprecated Index Help

PREV CLASS NEXT CLASS SUMMARY: NESTED | FIELD | CONSTS | METHOD

FRAMES NO FRAMES DETAIL: FIELD | CONSTS | METHOD

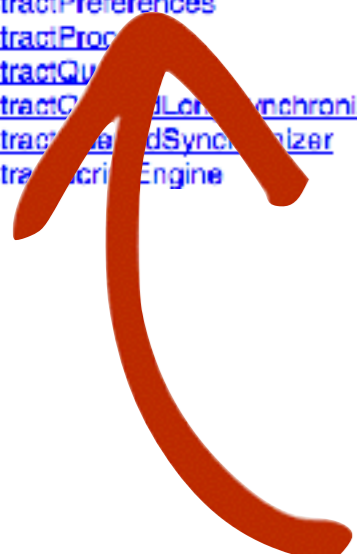
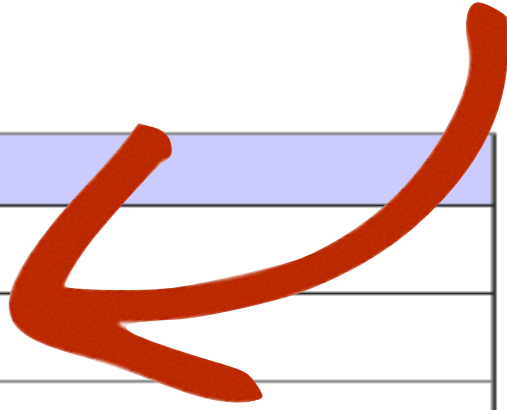
java.lang **Class String**

java.lang.Object
└ java.lang.String

All Implemented Interfaces:
Serializable, CharSequence, Comparable<String>

Method Summary

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



- 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

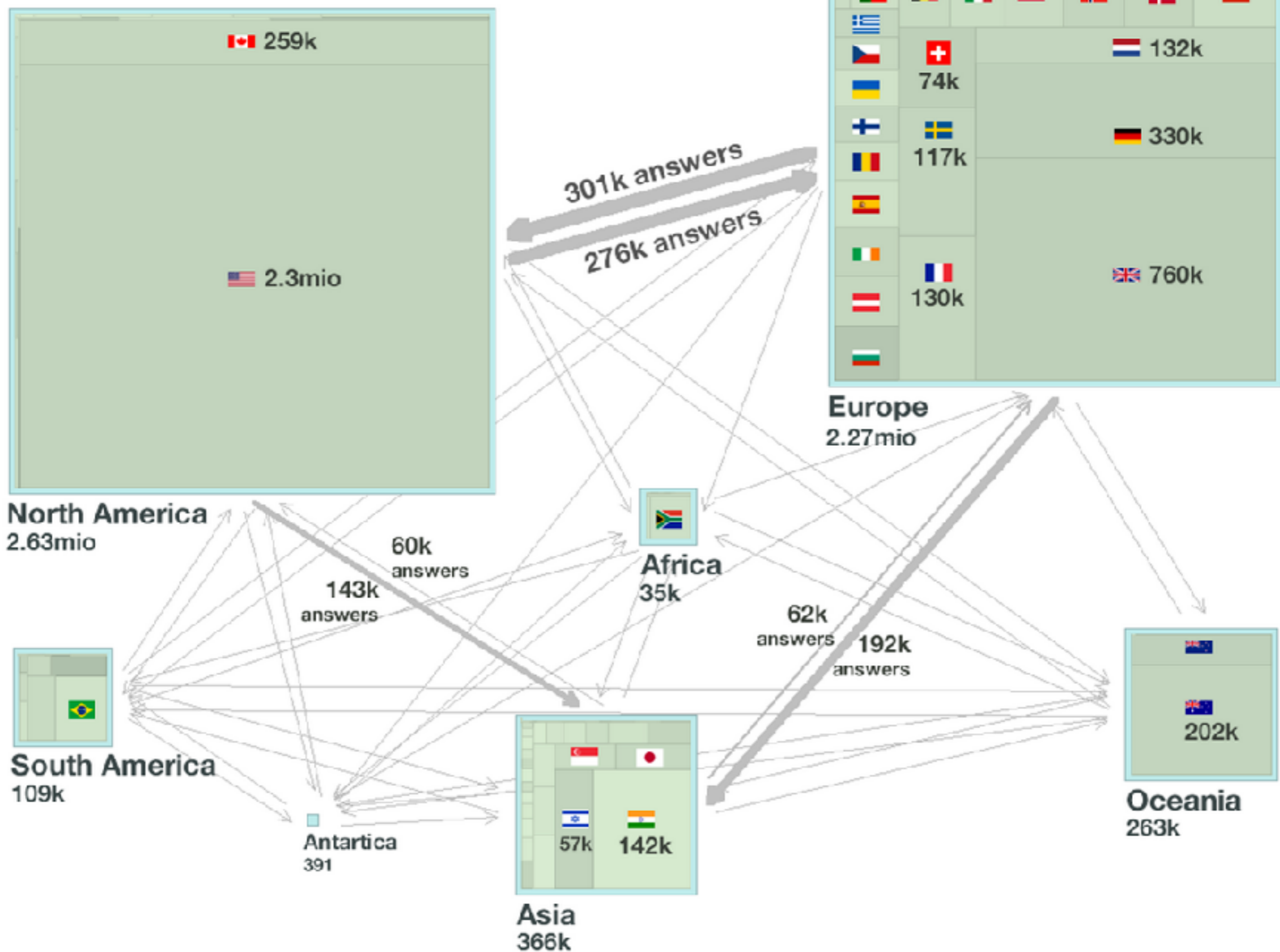
.NET Framework 4.5 | Other Versions ▾ | 37 out of 51 rated this helpful – Rate this topic

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, Boolean, 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



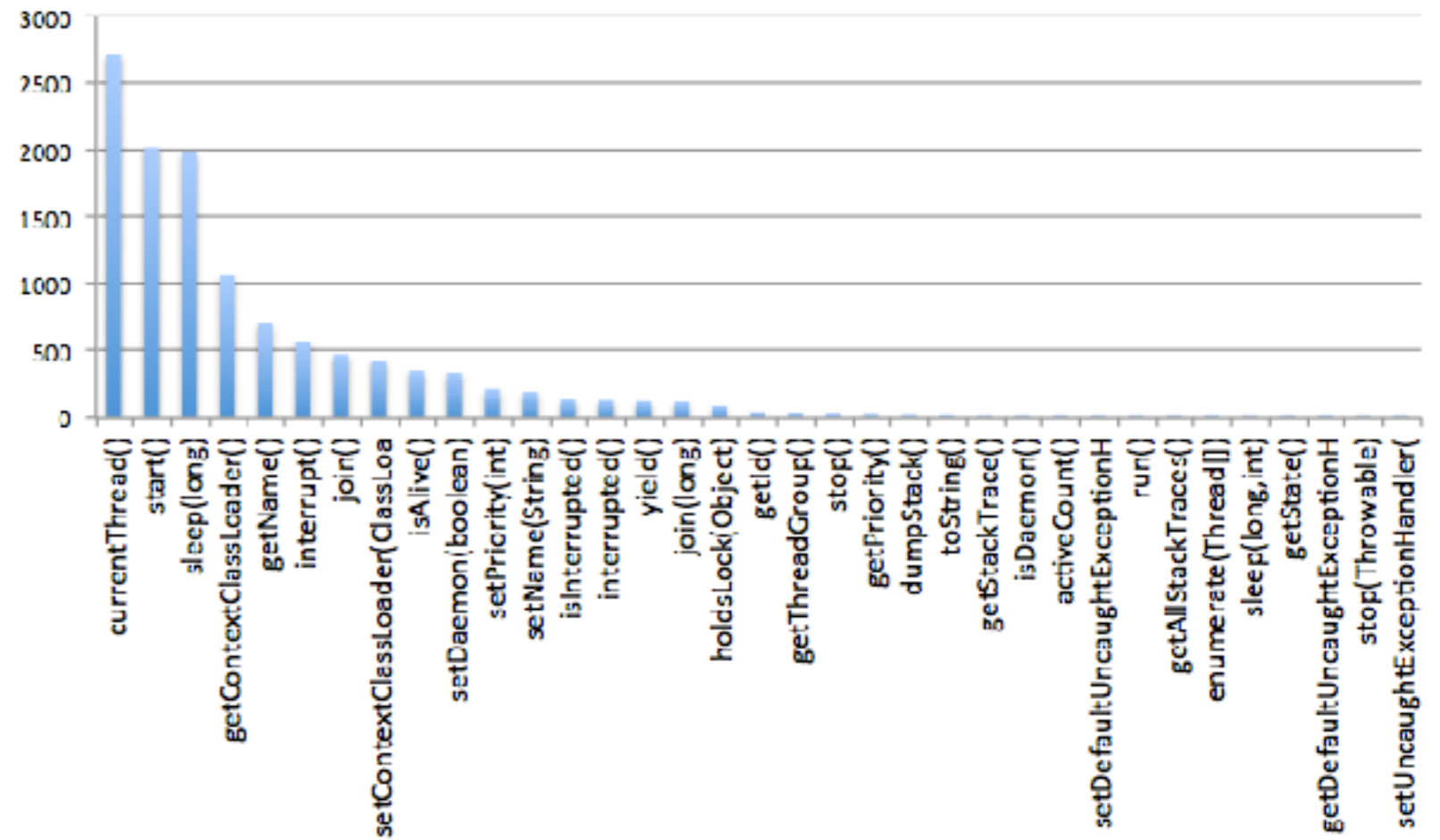




Usage of java.lang.Thread in the ecosystem



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





Augmenting Javadoc in Chrome

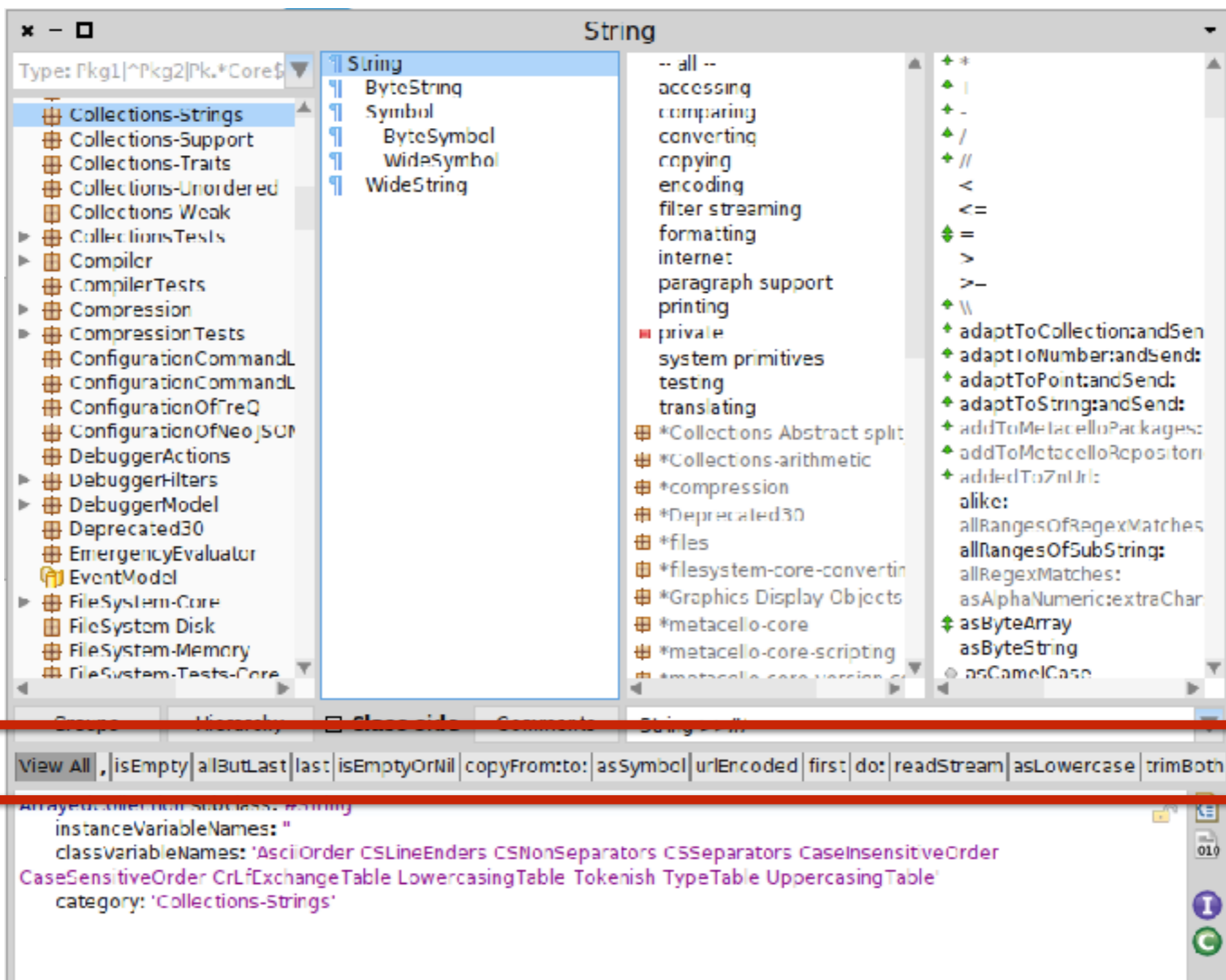
The screenshot shows the Oracle Java Platform SE 7 API documentation for the `String` class. The page includes navigation tabs for Overview, Package, Class, Use, Tree, Deprecated, Index, and Help. Below the class name, there is a table of statistics for the Javadoc comments, highlighted with a red border. The table has two columns: 'Modifier and Type' and 'Method and Description'.

Modifier and Type	Method and Description
boolean	<code>equals(Object anObject)</code> Compares this string to the specified object.
int	<code>length()</code> Returns the length of this string.
boolean	<code>startsWith(String prefix)</code> Tests if this string starts with the specified prefix.
char	<code>charAt(int index)</code> Returns the char value at the specified index.
String	<code>trim()</code> Returns a copy of the string, with leading and trailing whitespace omitted.
boolean	<code>equalsIgnoreCase(String anotherString)</code> Compares this string to another string, ignoring case considerations.
String	<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 and evolving 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, testing, and understanding code is very important. Static analysis and information in source code helps developers understand

#3 — understanding our organization++

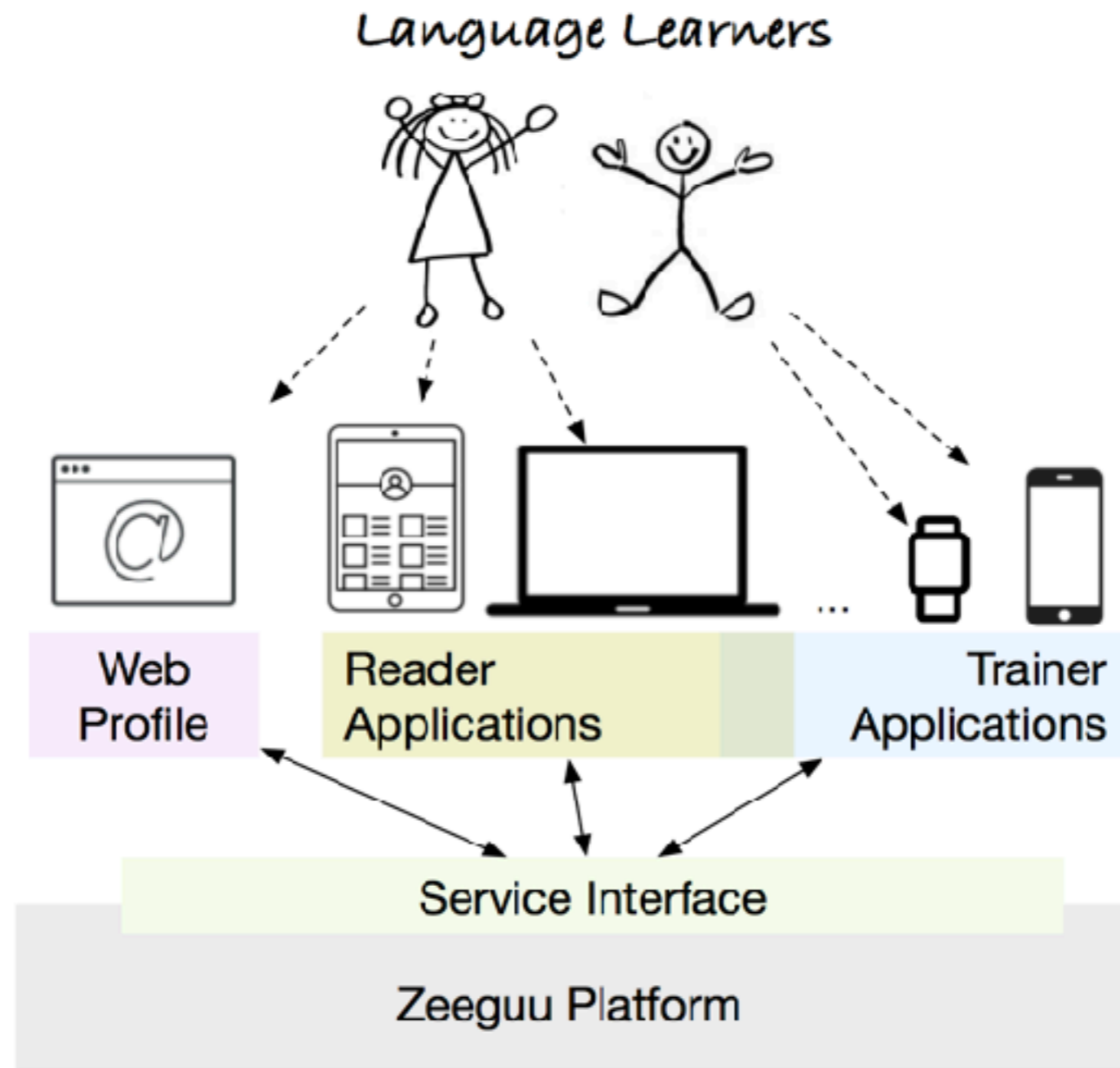
Most Popular Bug-Fixes

	Instances	Projects
↳ \longrightarrow <i>if(T != null)</i>	3,718	316
↳ \longrightarrow <i>if(T == null) return;</i>	1051	190
↳ \longrightarrow <i>if(T == null) return null;</i>	243	80
↳ \longrightarrow <i>if(T == null) throw new T();</i>	207	75
↳ \longrightarrow <i>if(T == null) T=T;</i>	157	67
↳ \longrightarrow <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

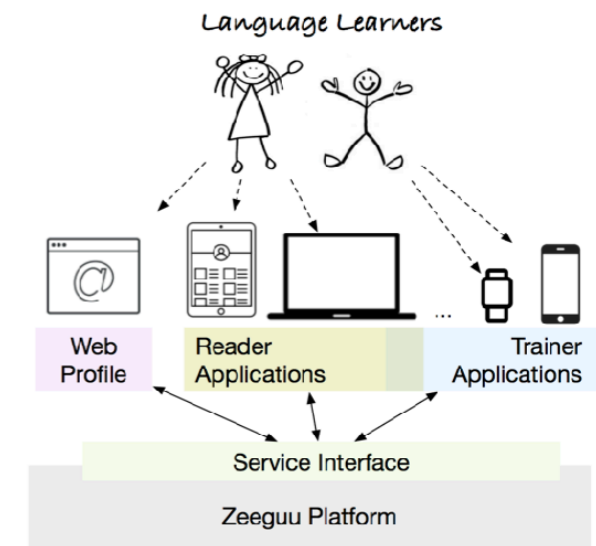
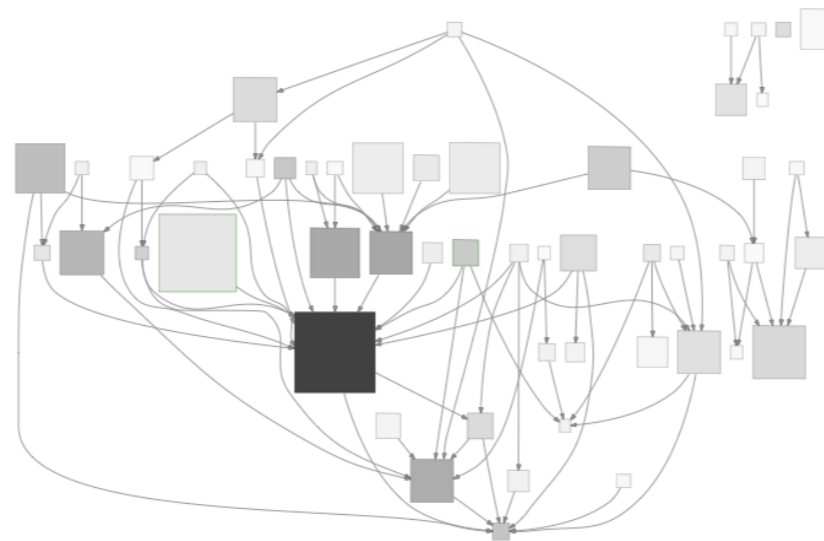
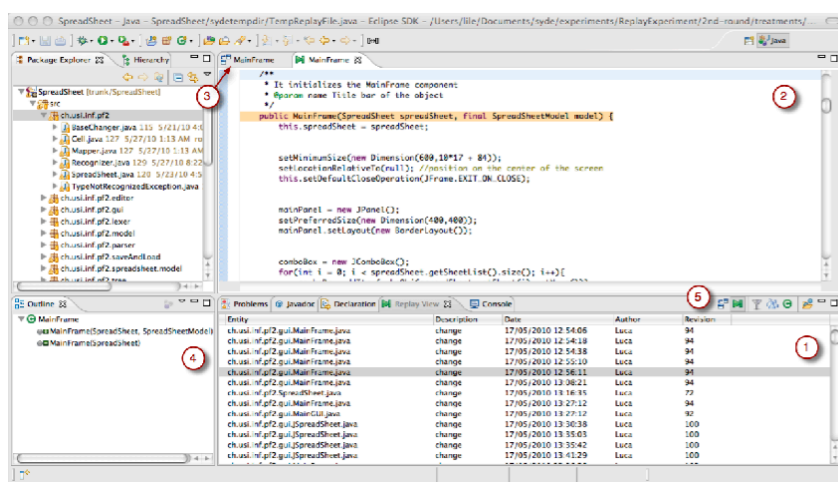


Code Reading Recommender

based on subjective complexity



Context in Software Analytics



Temporal

(eco)Systemic

Personal

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