Code Search & Reuse

Software Mining
Lecture 5 - 24th May ‘07

Rahul Premraj

What is Reuse?

Always Reuse?

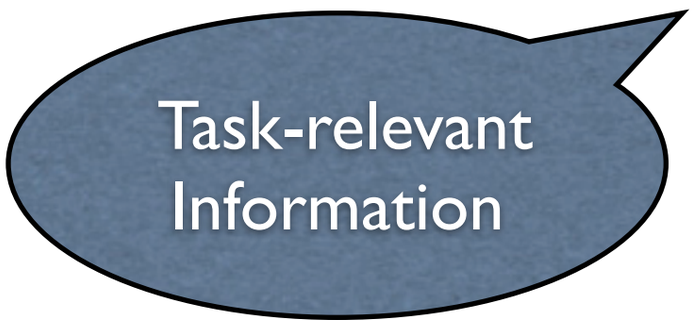
Always Reuse?

But...

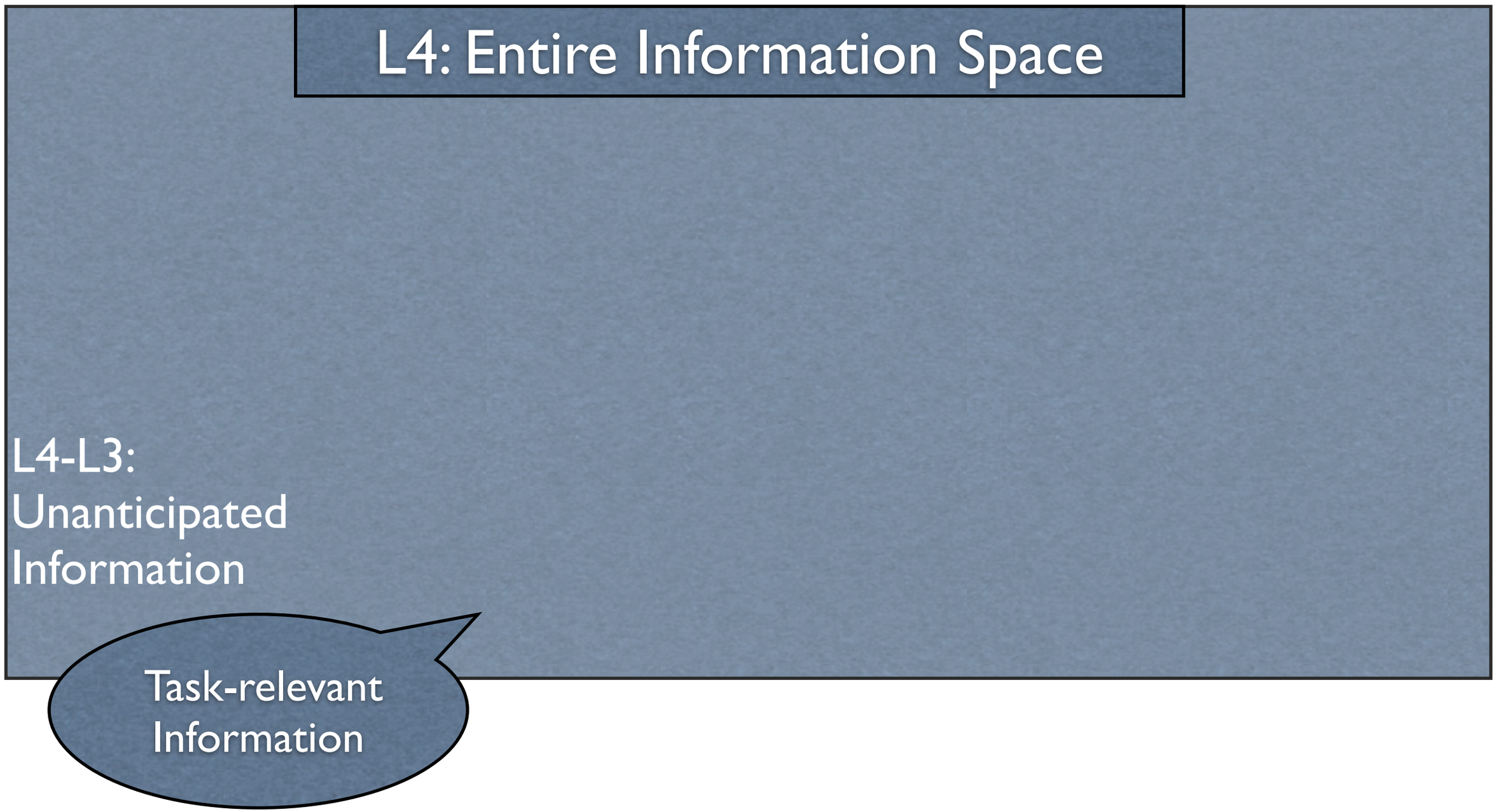
Reuse Requirements

- Creating and maintaining a reuse repository - takes commitment, effort and investment.
- Enabling software developers to build new software systems with components from the reuse repository.

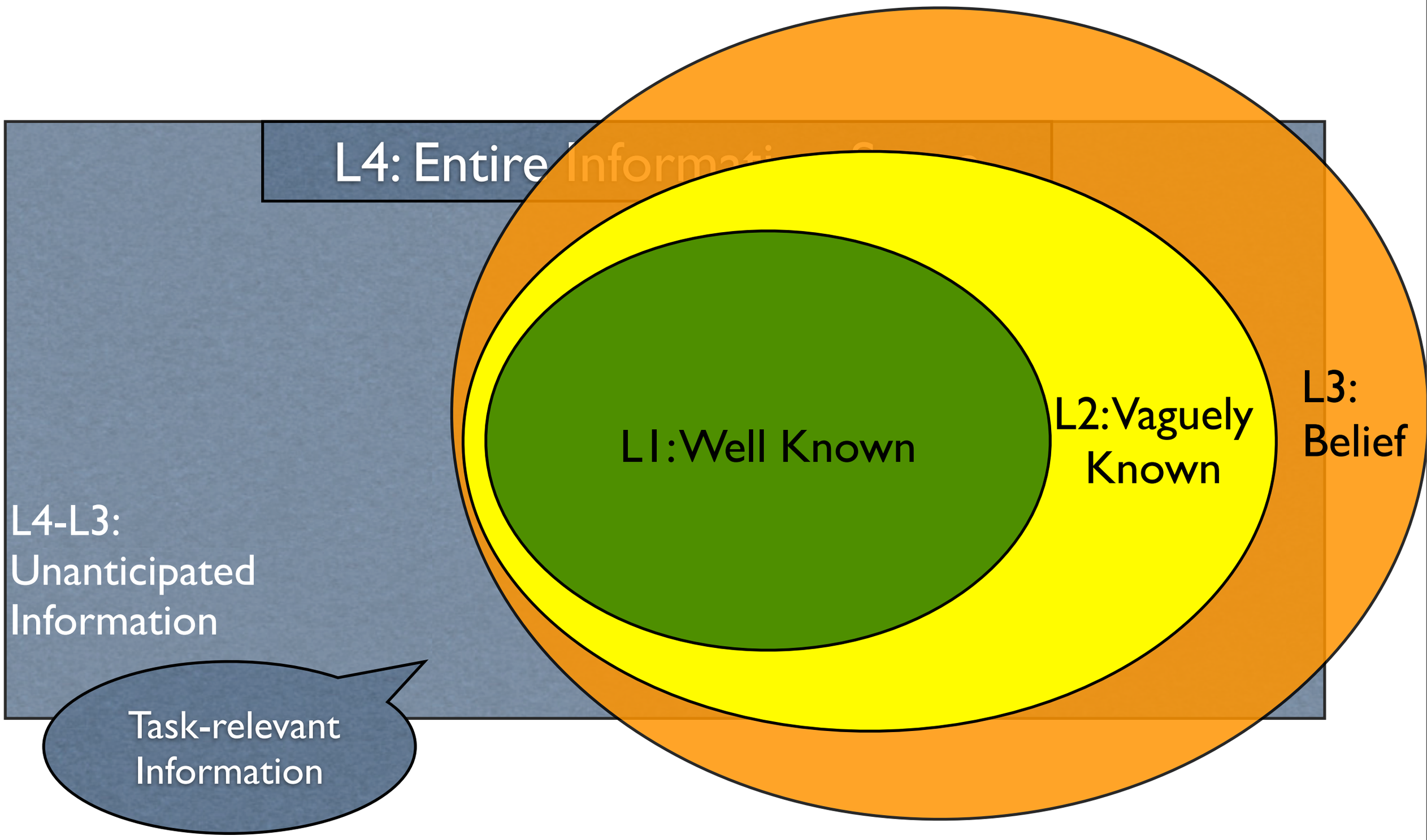
Cognition Issues in Reuse



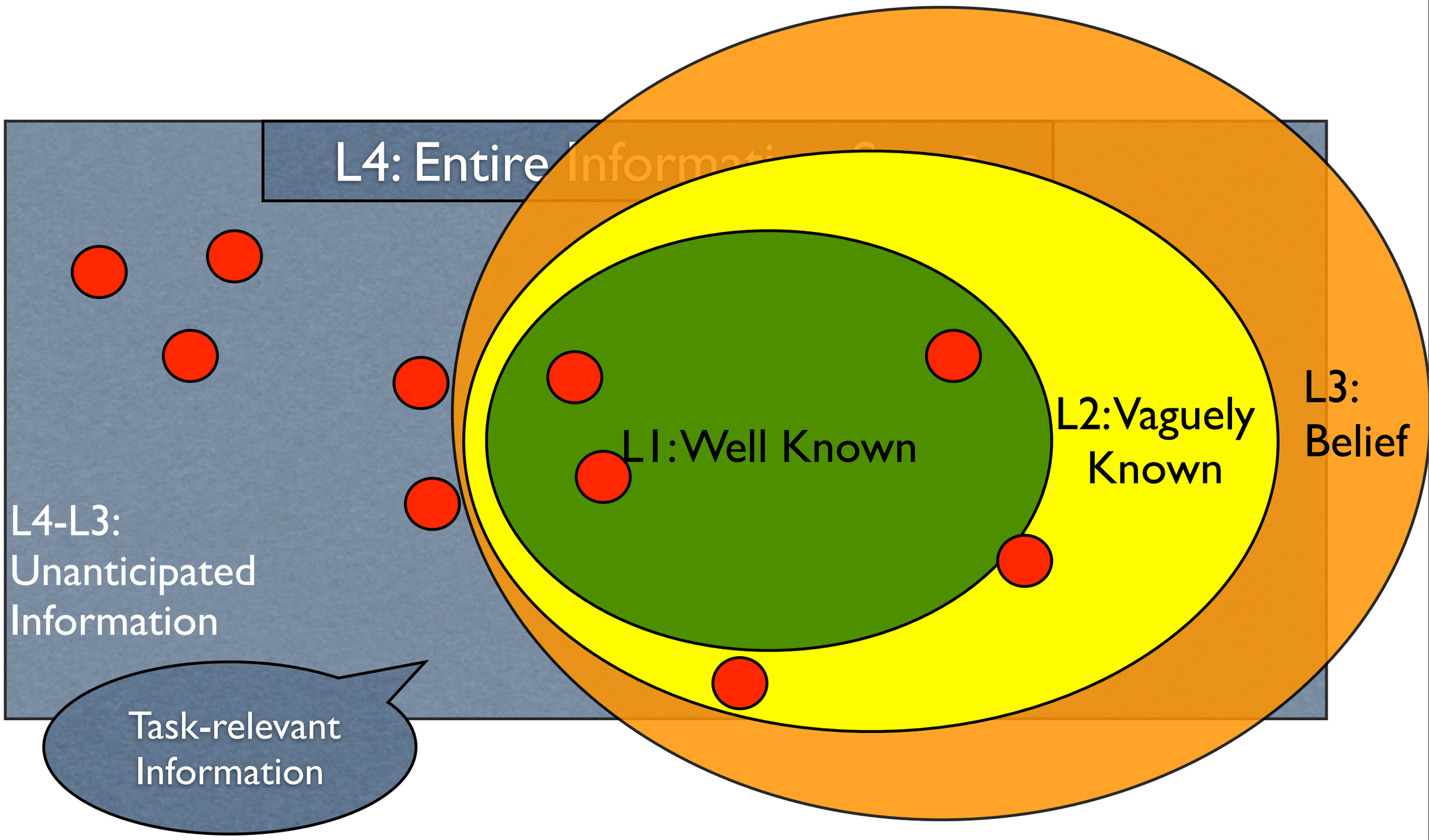
Cognition Issues in Reuse



Cognition Issues in Reuse



Cognition Issues in Reuse

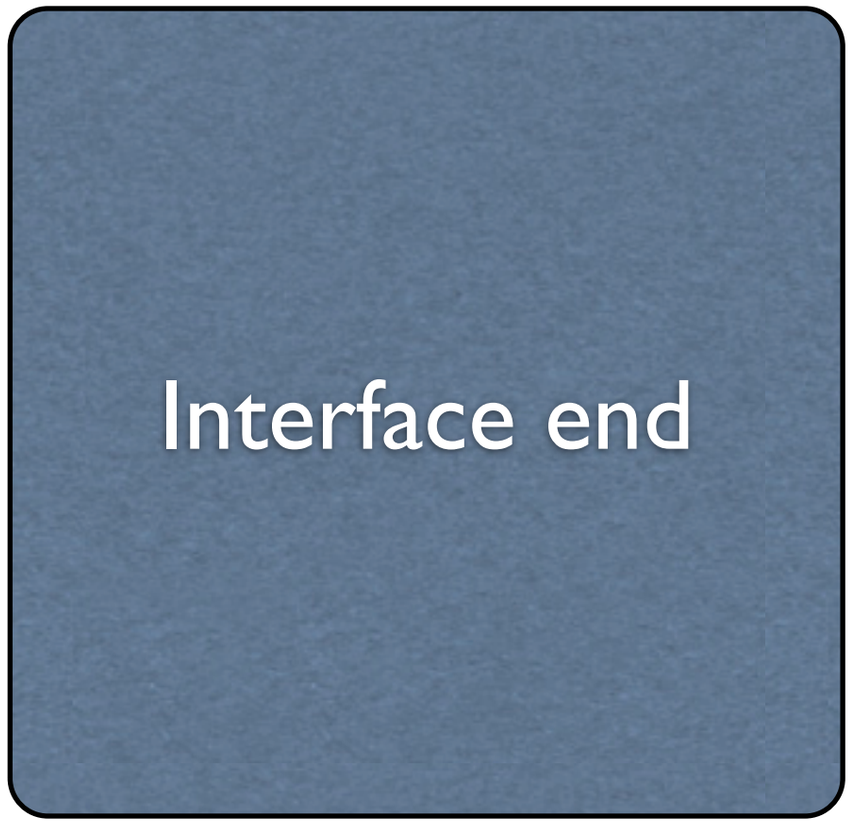


CodeBroker

delivers task relevant and personalised components...

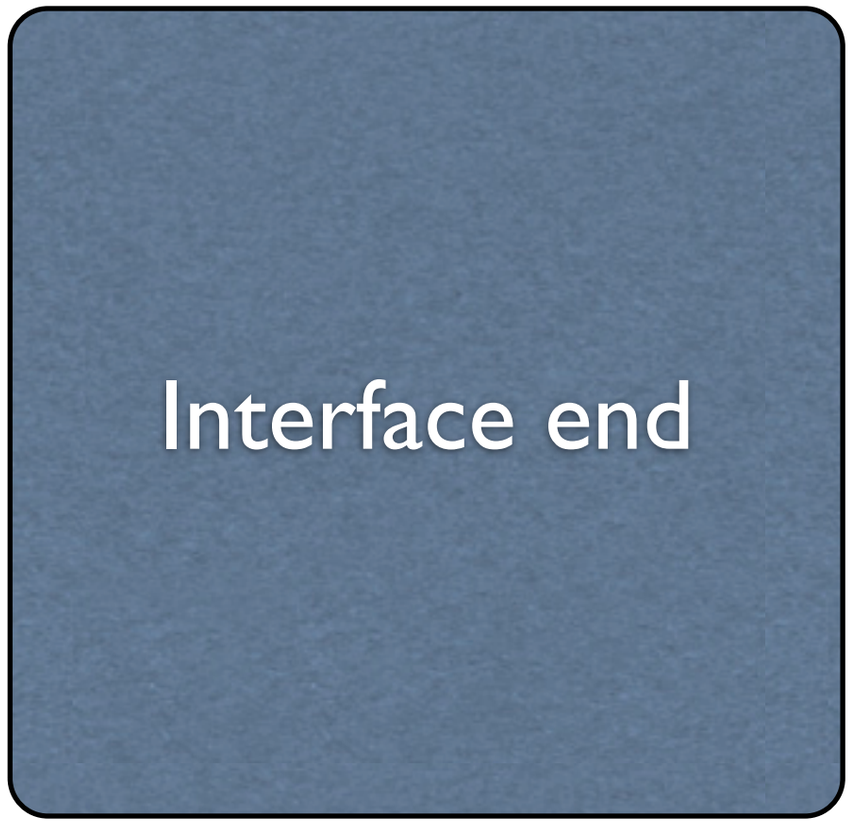
CodeBroker

delivers task relevant and personalised components...



CodeBroker

delivers task relevant and personalised components...



CodeBroker

delivers task relevant and personalised components...

Interface end

Back-end search engine

responds to comments

e.g. // generate random numbers

CodeBroker

delivers task relevant and personalised components...

Interface end

responds to comments

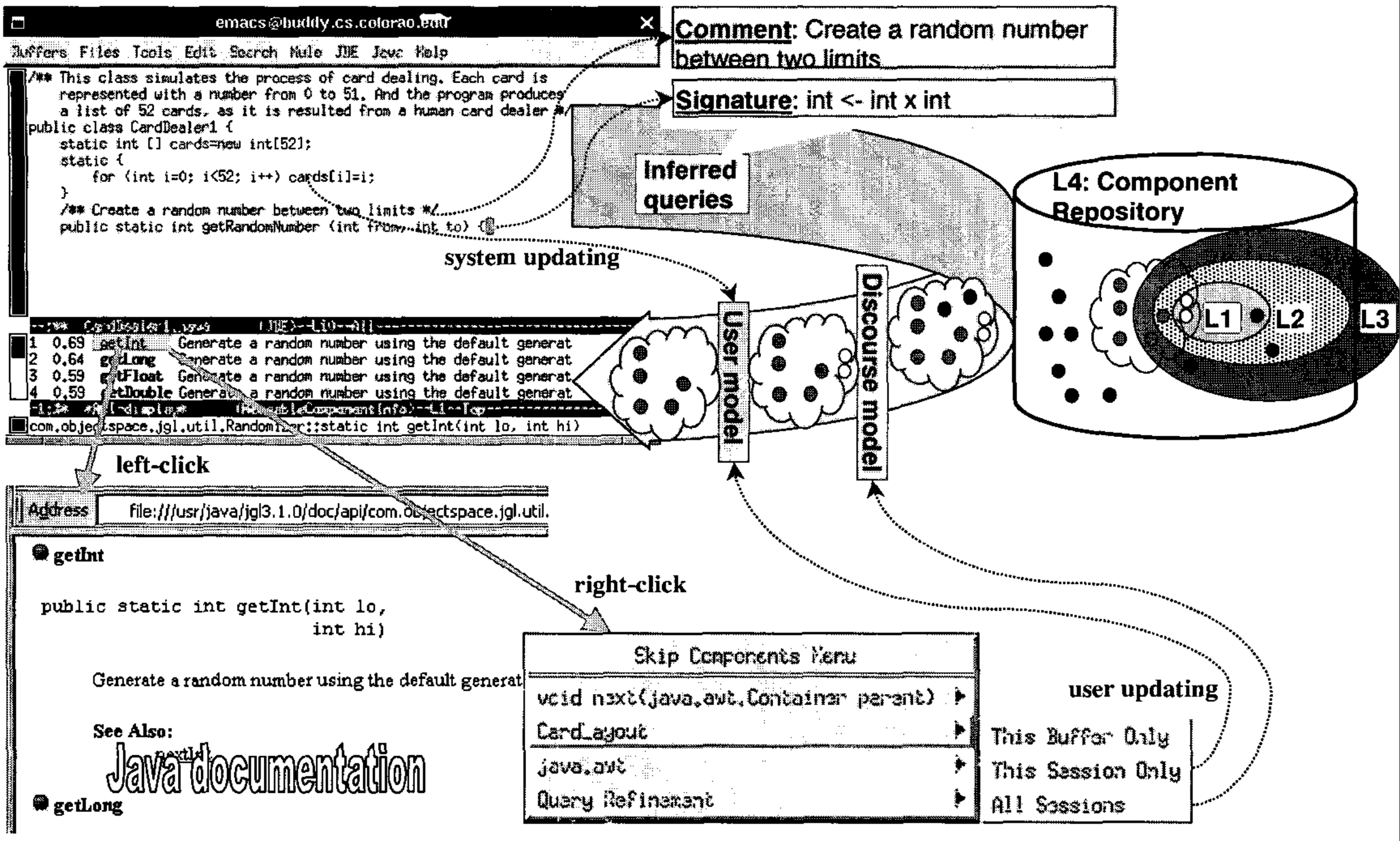
e.g. // generate random numbers

Back-end search engine

clues in method signatures

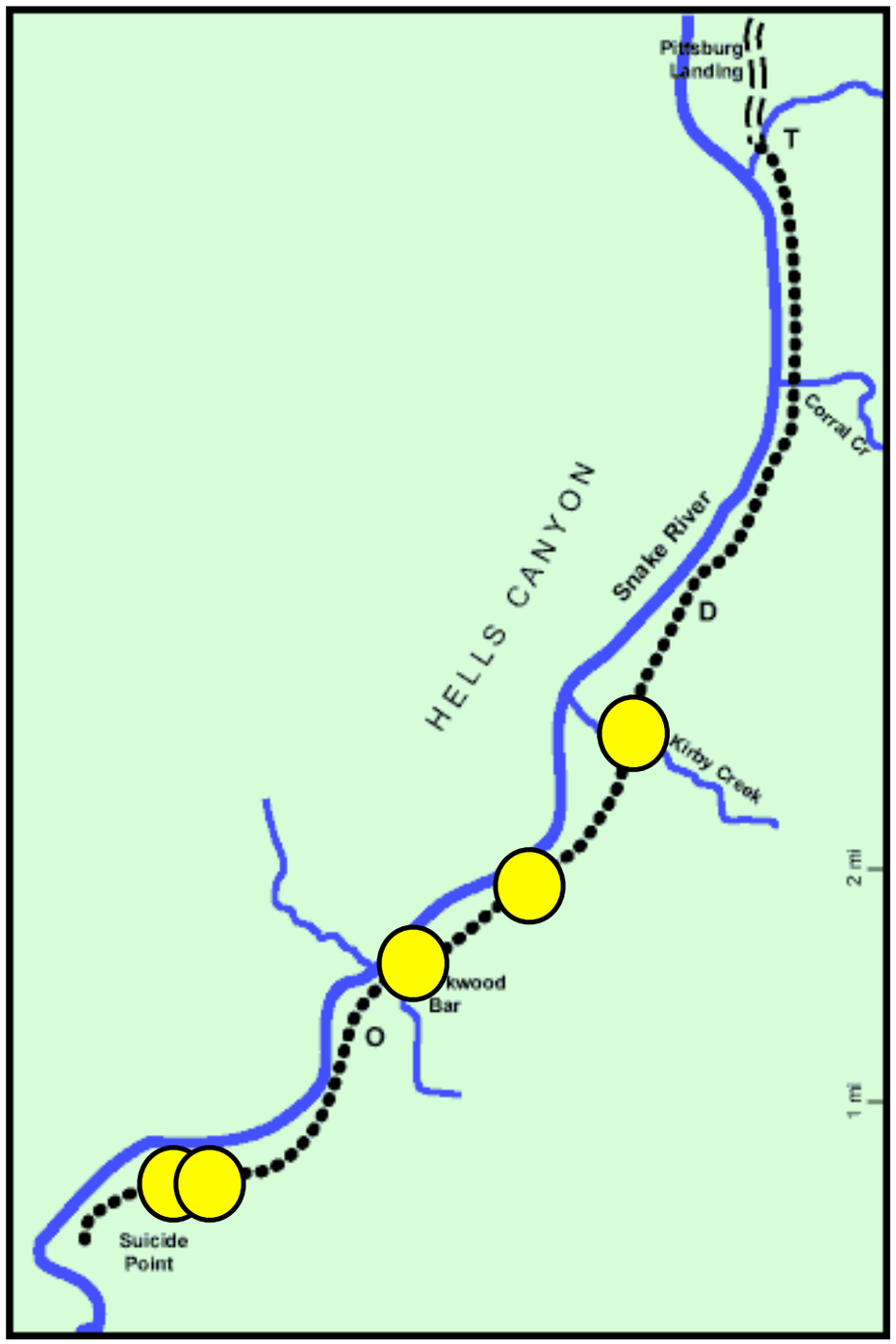
e.g. void randomNumberGenerator (int x)

CodeBroker


emacs@buddy.cs.colorao.edu
Buffers Files Tools Edit Search Mule JDE Java Help
/** This class simulates the process of card dealing. Each card is
represented with a number from 0 to 51. And the program produces
a list of 52 cards, as it is resulted from a human card dealer */
public class CardDealer1 {
static int [] cards=new int[52];
static {
for (int i=0; i<52; i++) cards[i]=i;
}
/** Create a random number between two limits */
public static int getRandomNumber (int from, int to) {
Comment: Create a random number between two limits
Signature: int <- int x int
Inferred queries
system updating
L4: Component Repository
L1
L2
L3
User model
Discourse model
1 0.69 getInt Generate a random number using the default generat
2 0.64 getLong Generate a random number using the default generat
3 0.59 getFloat Generate a random number using the default generat
4 0.59 getDouble Generate a random number using the default generat
com.objectspace.jgl.util.RandomNumber::static int getInt(int lo, int hi)
left-click
Address
file:///usr/java/jgl3.1.0/doc/api/com.objectspace.jgl.util.
getInt
public static int getInt(int lo,
int hi)
Generate a random number using the default generat
See Also:
Java documentation
getLong
right-click
Skip Components Menu
void next(java.awt.Container parent)
CardLayout
java.awt
Query Refinement
user updating
This Buffer Only
This Session Only
All Sessions

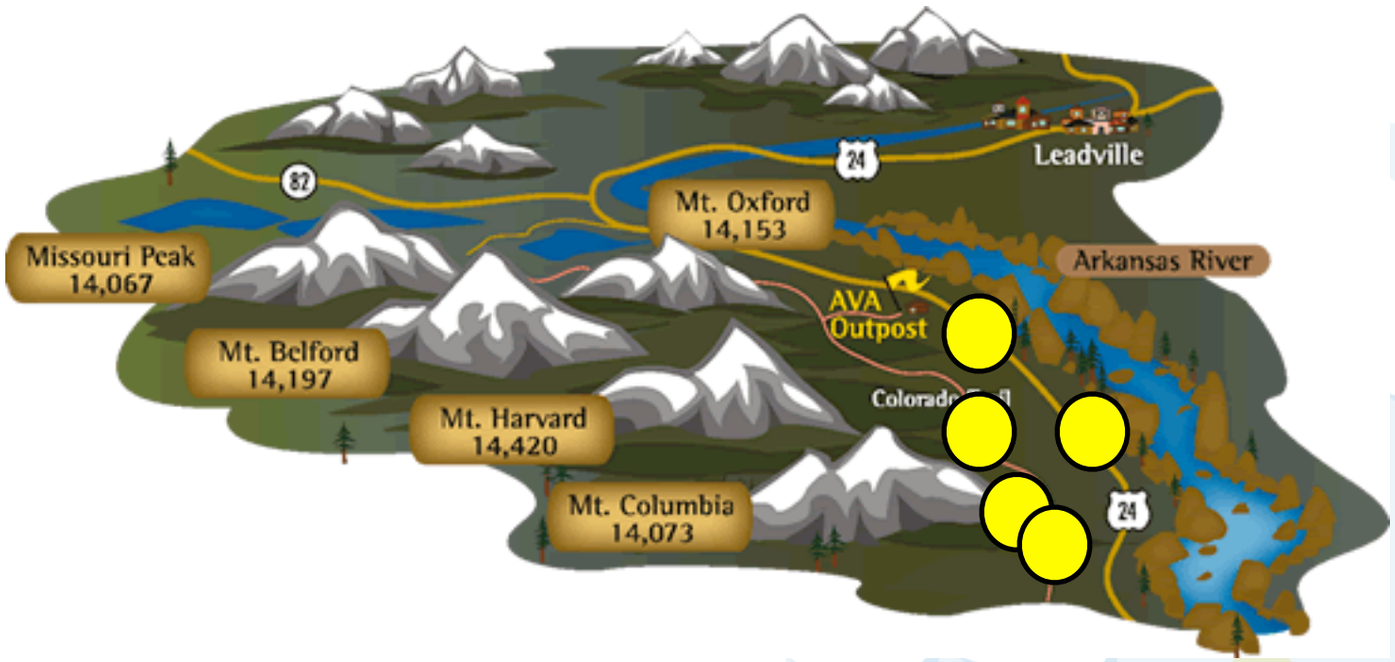
UltiGPX

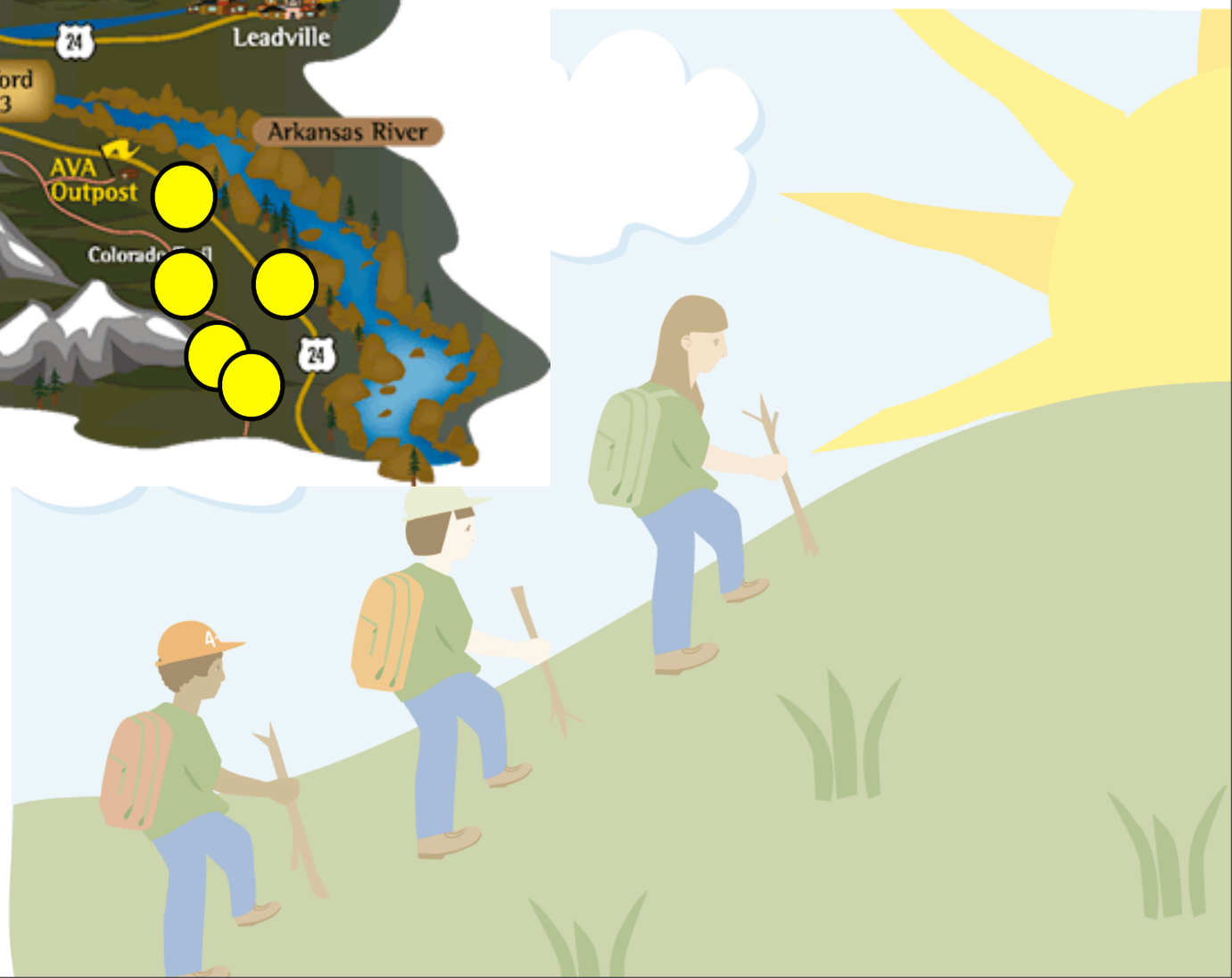
A tool for visualising GPS data collected by hikers



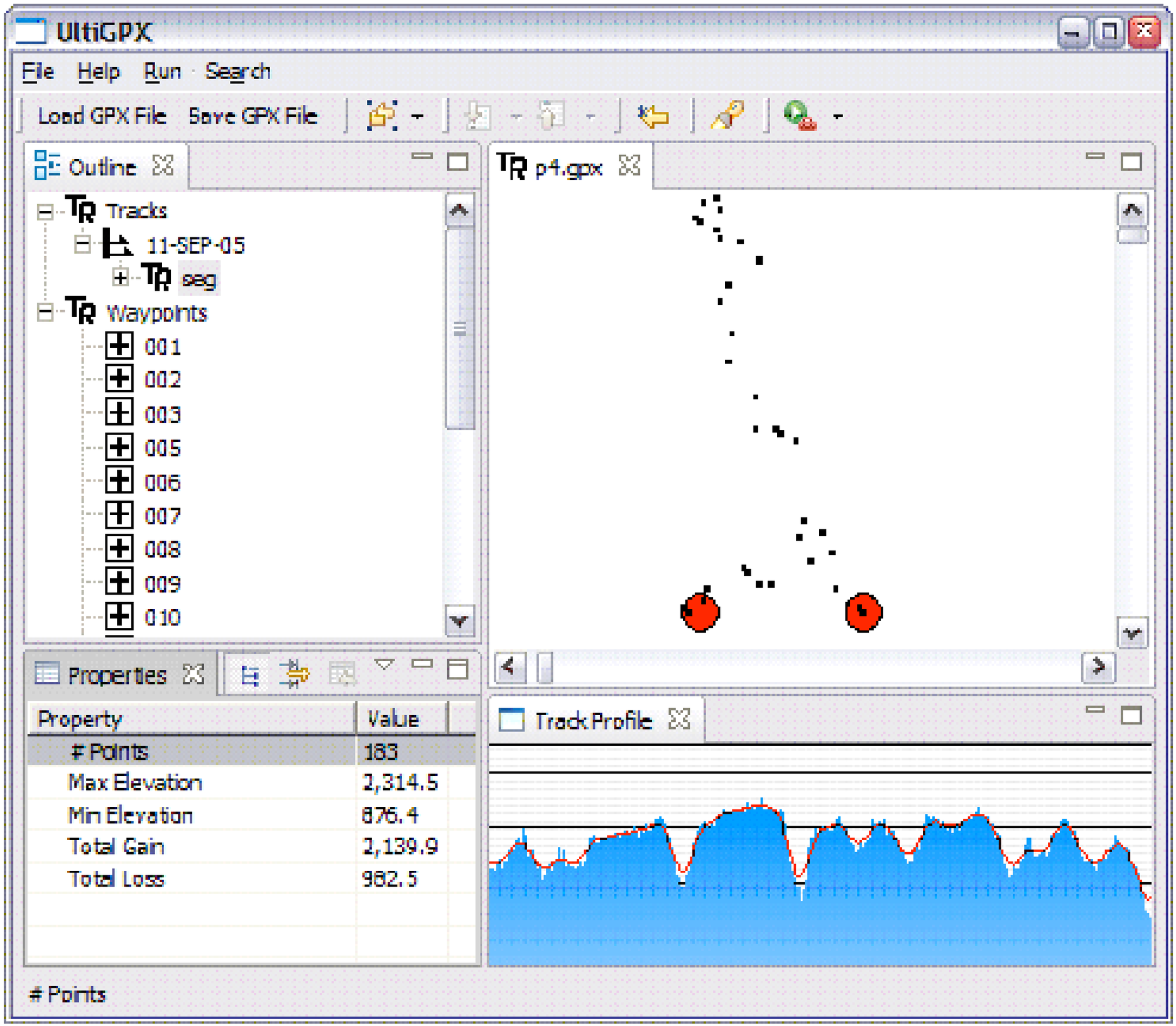
UltiGPX

A tool for visualising GPS data collected by hikers



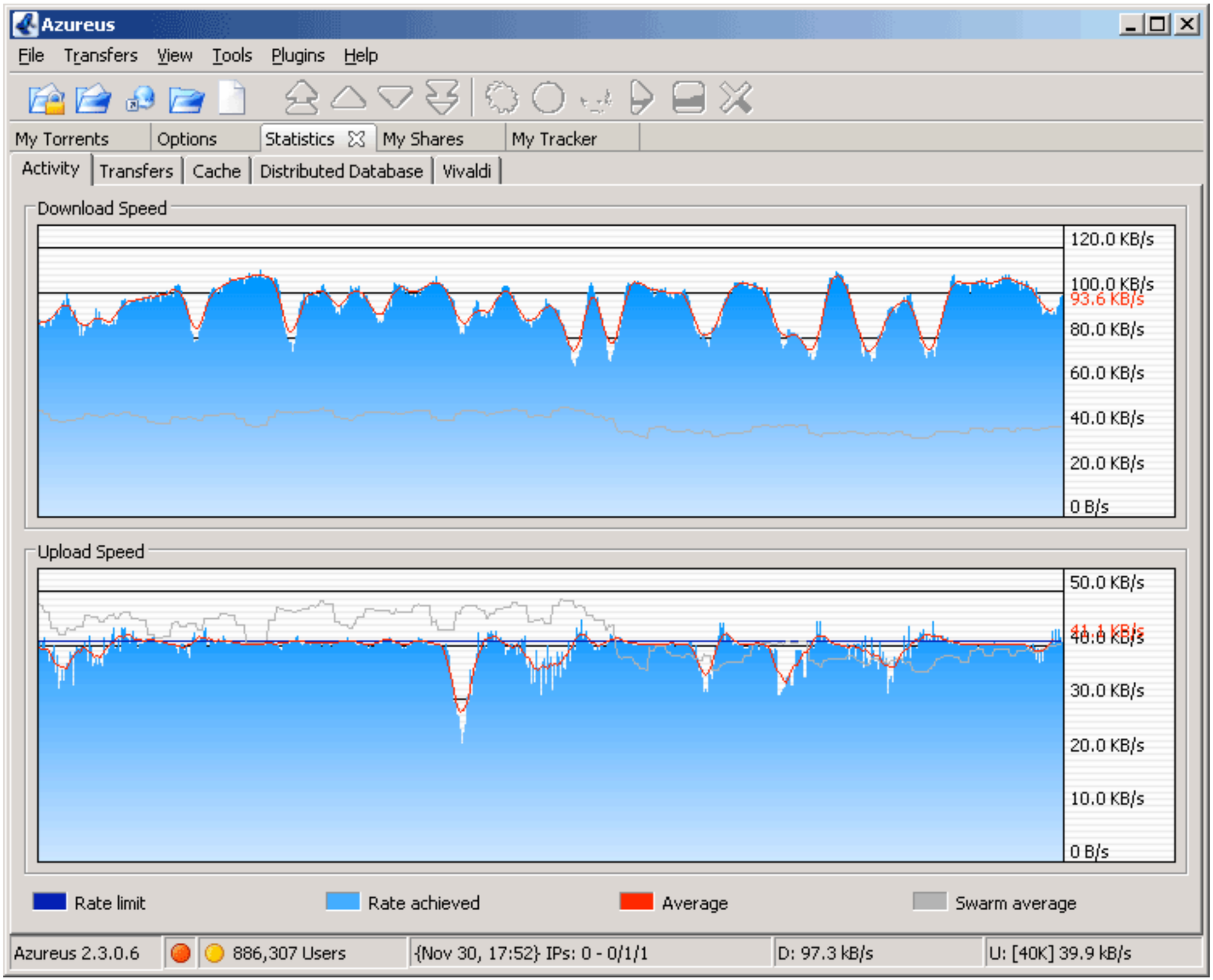


UltiGPX

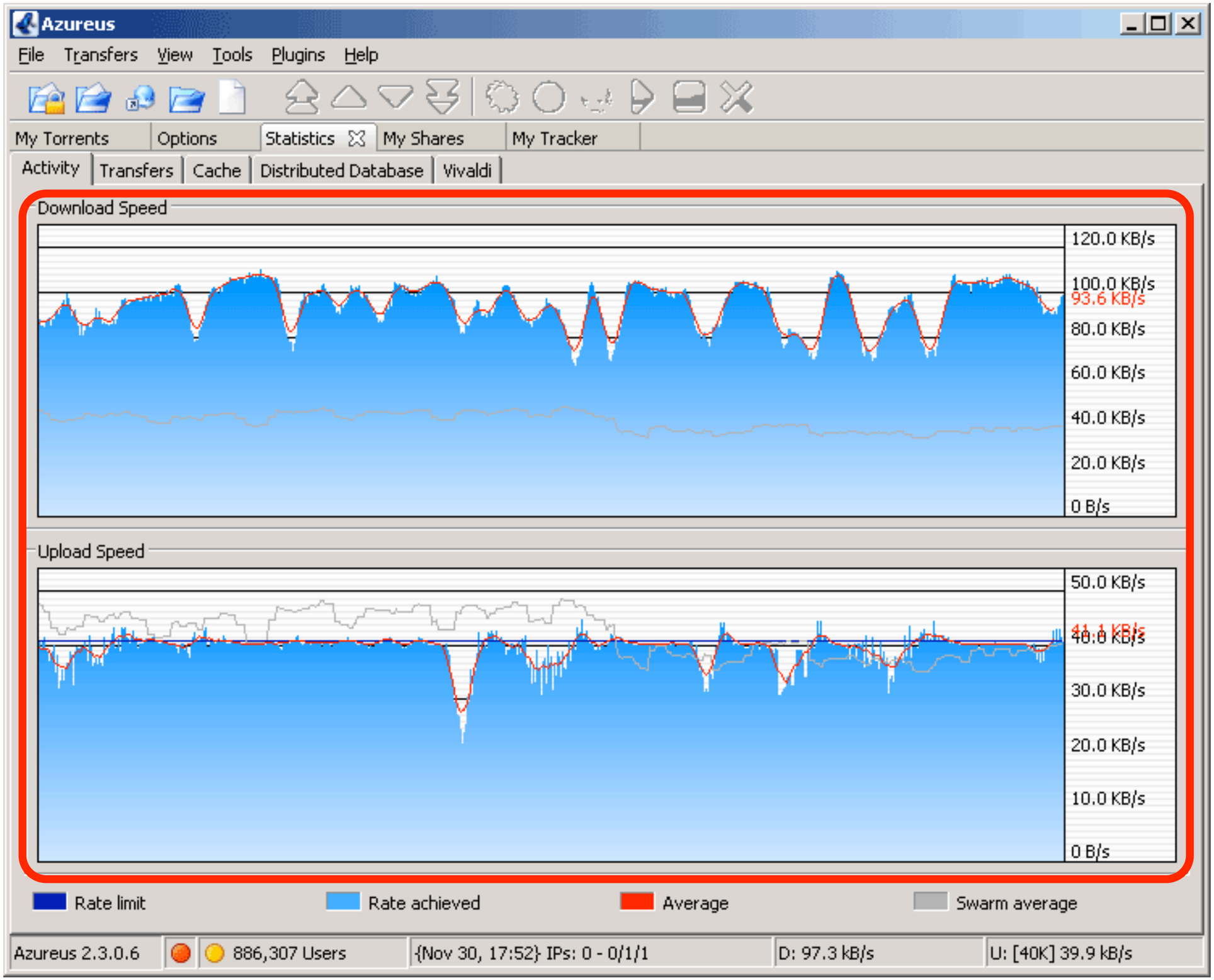

UltiGPX
File Help Run Search
Load GPX File Save GPX File
Outline
Tracks
11-SEP-05
seg
Waypoints
001
002
003
005
006
007
008
009
010
p4.gpx
Properties
Property
Value
Points
183
Max Elevation
2,314.5
Min Elevation
876.4
Total Gain
2,139.9
Total Loss
982.5
Track Profile
Points

Azureus

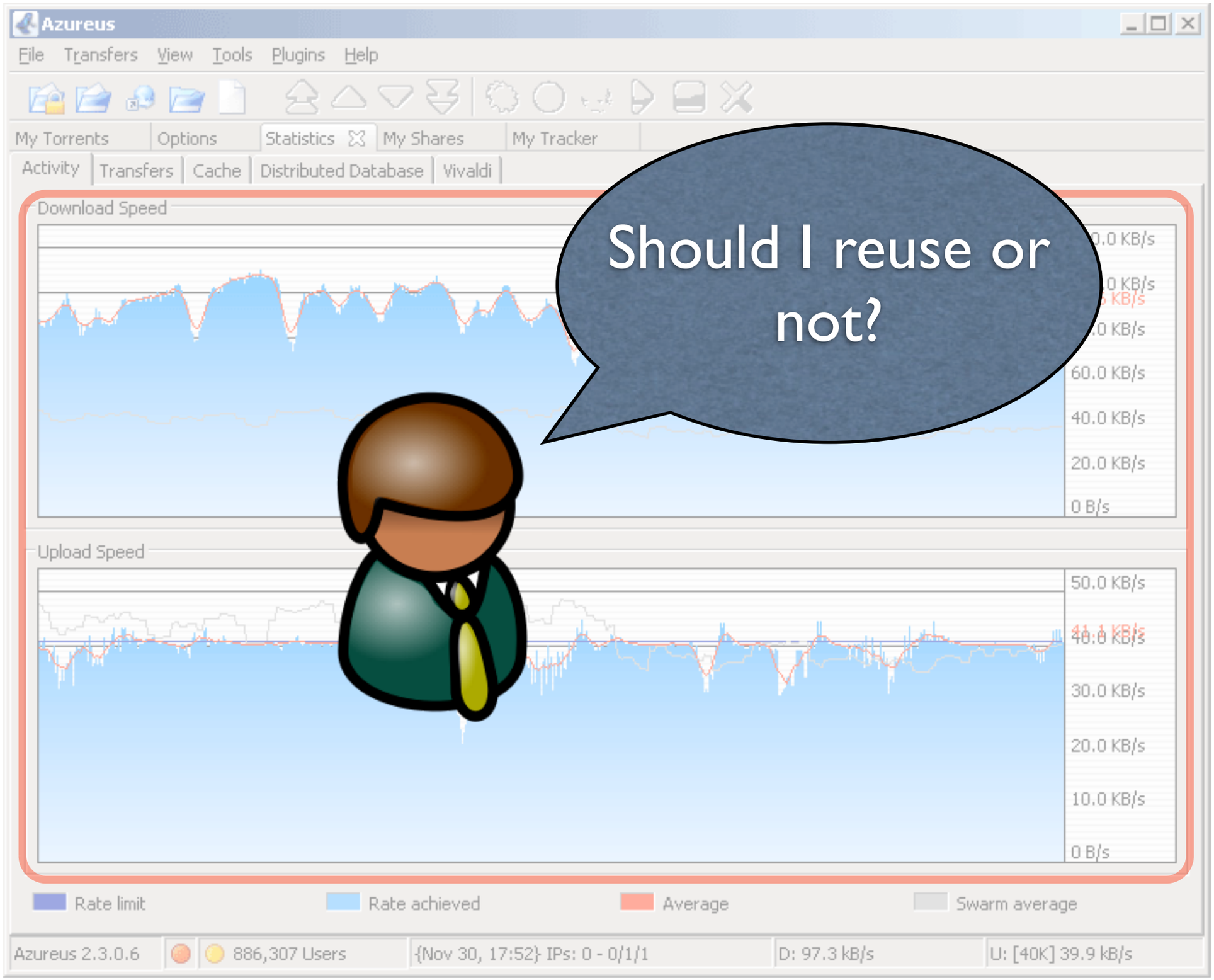
Azureus


Azureus
File Transfers View Tools Plugins Help
My Torrents Options Statistics My Shares My Tracker
Activity Transfers Cache Distributed Database Vivaldi
Download Speed
120.0 KB/s
100.0 KB/s
93.6 KB/s
80.0 KB/s
60.0 KB/s
40.0 KB/s
20.0 KB/s
0 B/s
Upload Speed
50.0 KB/s
41.1 KB/s
40.0 KB/s
30.0 KB/s
20.0 KB/s
10.0 KB/s
0 B/s
Rate limit
Rate achieved
Average
Swarm average
Azureus 2.3.0.6
886,307 Users
{Nov 30, 17:52} IPs: 0 - 0/1/1
D: 97.3 kB/s
U: [40K] 39.9 kB/s

Azureus

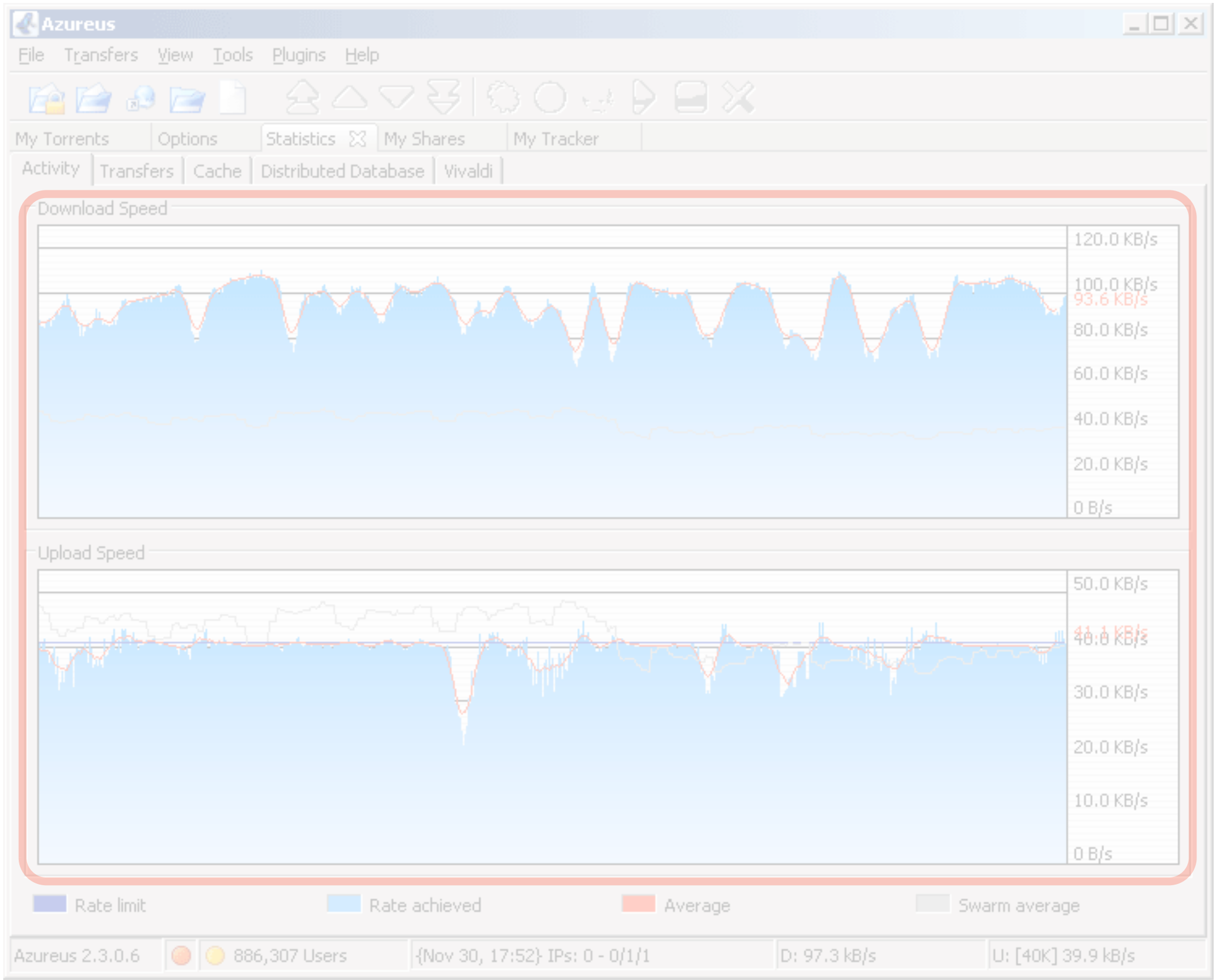

Azureus
File Transfers View Tools Plugins Help
My Torrents Options Statistics My Shares My Tracker
Activity Transfers Cache Distributed Database Vivaldi
Download Speed
120.0 KB/s
100.0 KB/s
93.6 KB/s
80.0 KB/s
60.0 KB/s
40.0 KB/s
20.0 KB/s
0 B/s
Upload Speed
50.0 KB/s
41.1 KB/s
40.0 KB/s
30.0 KB/s
20.0 KB/s
10.0 KB/s
0 B/s
Rate limit
Rate achieved
Average
Swarm average
Azureus 2.3.0.6
886,307 Users
{Nov 30, 17:52} IPs: 0 - 0/1/1
D: 97.3 kB/s
U: [40K] 39.9 kB/s

Azureus

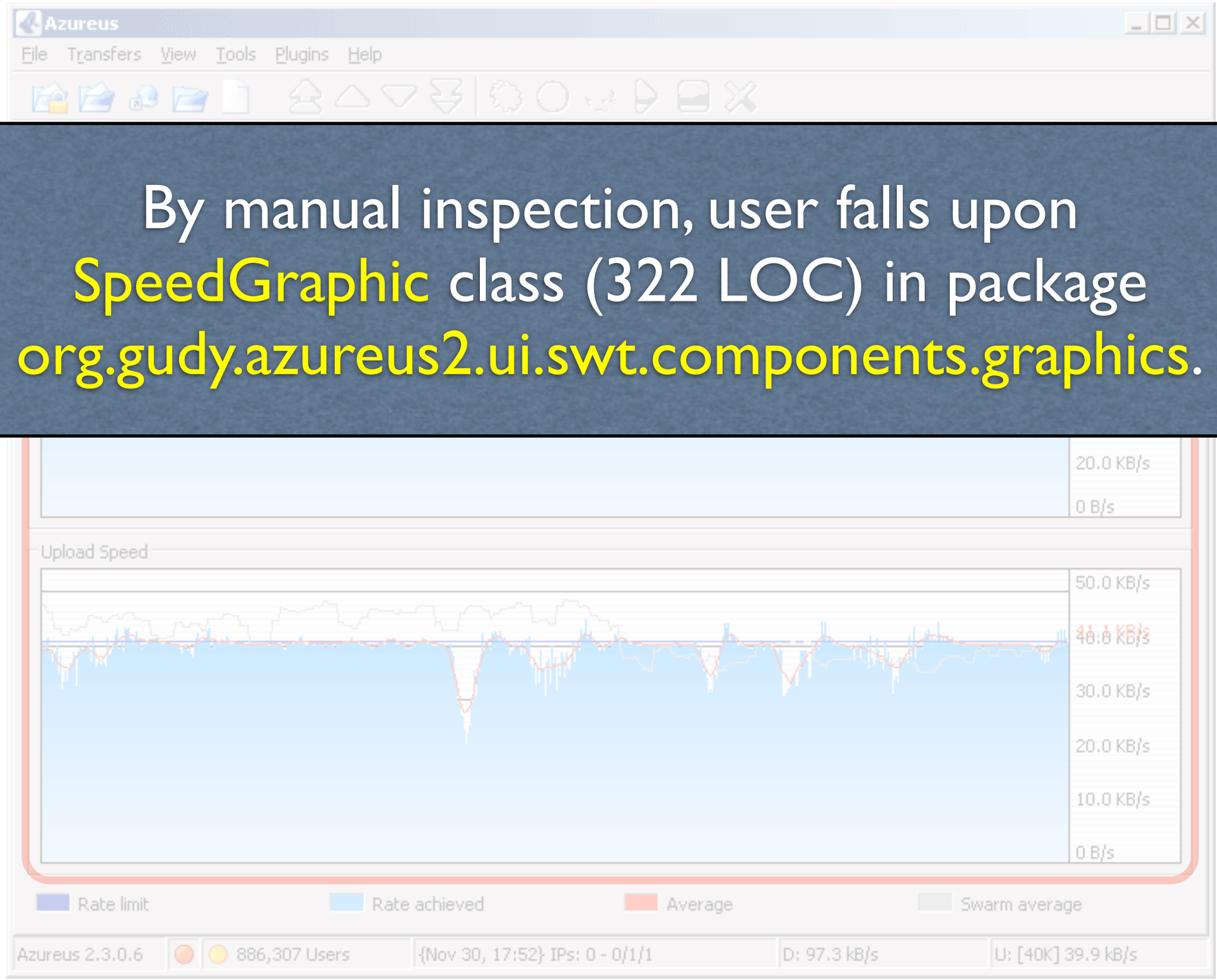

Azureus
File Transfers View Tools Plugins Help
My Torrents Options Statistics My Shares My Tracker
Activity Transfers Cache Distributed Database Vivaldi
Download Speed
60.0 KB/s
40.0 KB/s
20.0 KB/s
0 B/s
Should I reuse or not?
Upload Speed
50.0 KB/s
30.0 KB/s
20.0 KB/s
10.0 KB/s
0 B/s
Rate limit
Rate achieved
Average
Swarm average
Azureus 2.3.0.6
886,307 Users
{Nov 30, 17:52} IPs: 0 - 0/1/1
D: 97.3 kB/s
U: [40K] 39.9 kB/s

Classic Code Searching

Azureus

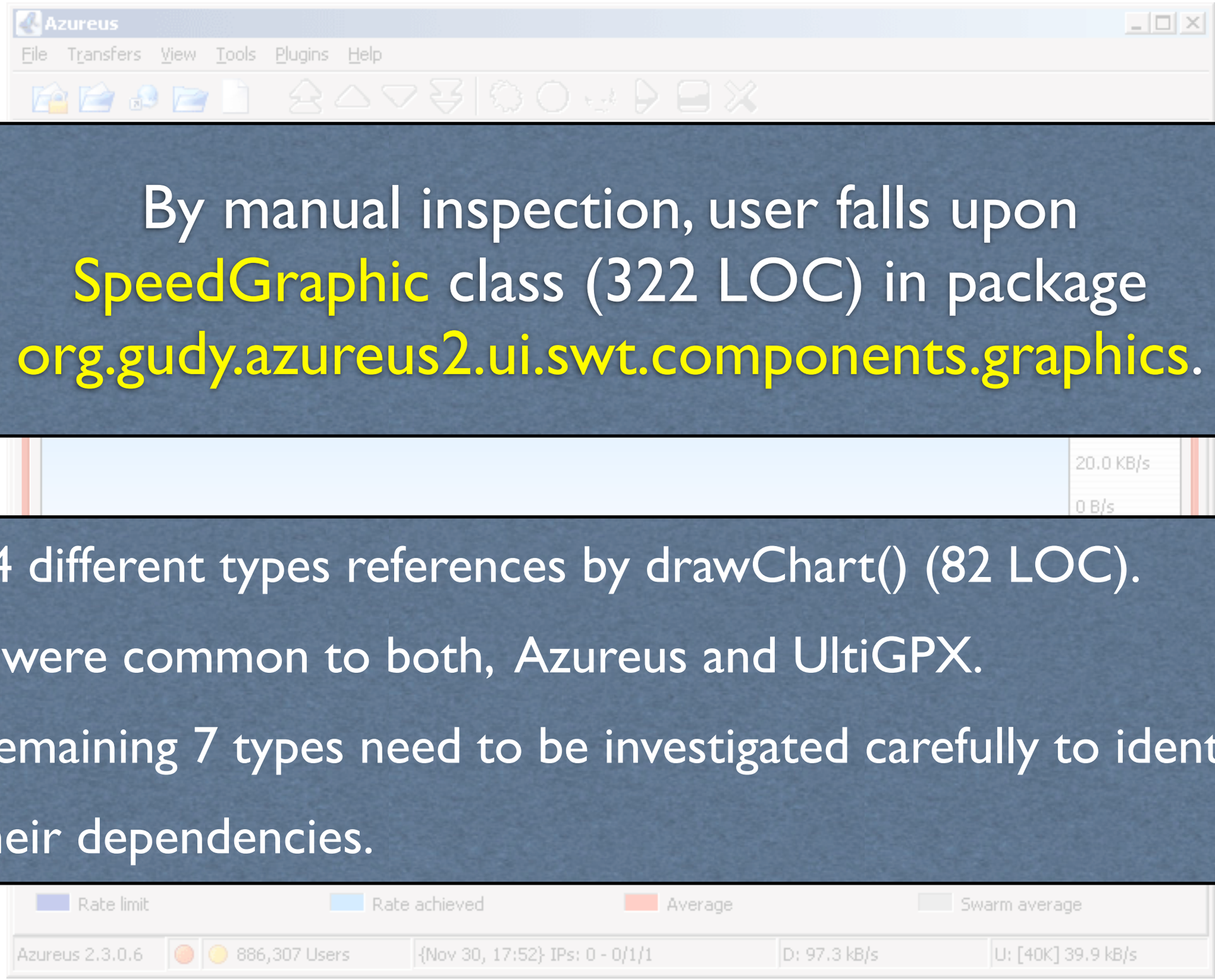

Azureus
File Transfers View Tools Plugins Help
My Torrents Options Statistics My Shares My Tracker
Activity Transfers Cache Distributed Database Vivaldi
Download Speed
120.0 KB/s
100.0 KB/s
93.6 KB/s
80.0 KB/s
60.0 KB/s
40.0 KB/s
20.0 KB/s
0 B/s
Upload Speed
50.0 KB/s
41.1 KB/s
40.0 KB/s
30.0 KB/s
20.0 KB/s
10.0 KB/s
0 B/s
Rate limit
Rate achieved
Average
Swarm average
Azureus 2.3.0.6
886,307 Users
{Nov 30, 17:52} IPs: 0 - 0/1/1
D: 97.3 kB/s
U: [40K] 39.9 kB/s

Azureus



By manual inspection, user falls upon SpeedGraphic class (322 LOC) in package org.gudy.azureus2.ui.swt.components.graphics.

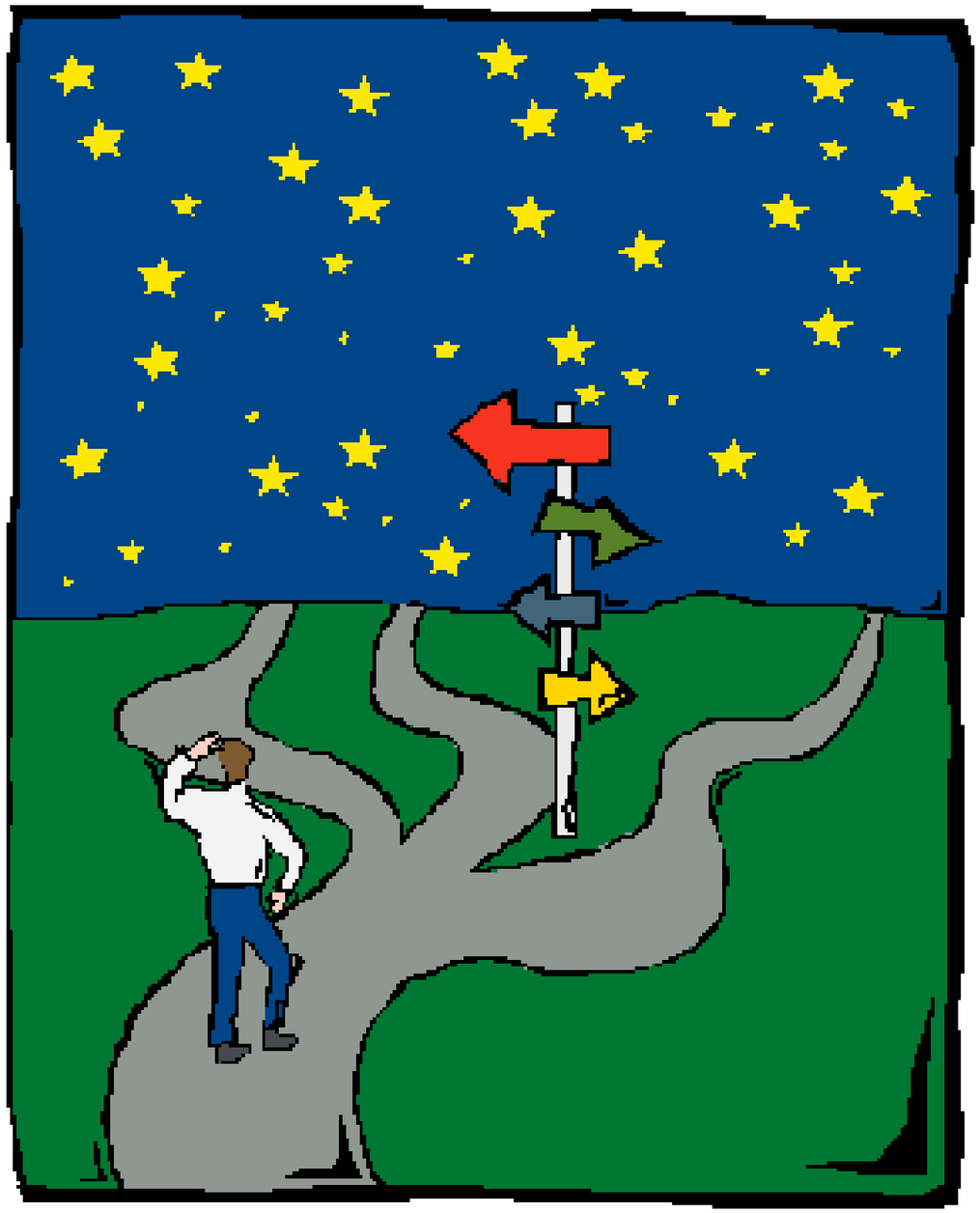
Azureus



By manual inspection, user falls upon SpeedGraphic class (322 LOC) in package org.gudy.azureus2.ui.swt.components.graphics.

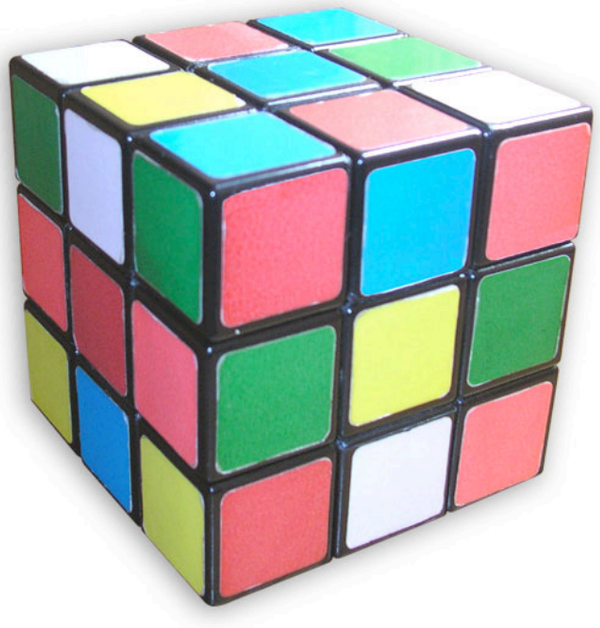
- 14 different types references by drawChart() (82 LOC).
- 7 were common to both, Azureus and UltiGPX.
- Remaining 7 types need to be investigated carefully to identify their dependencies.

What decisions were made?



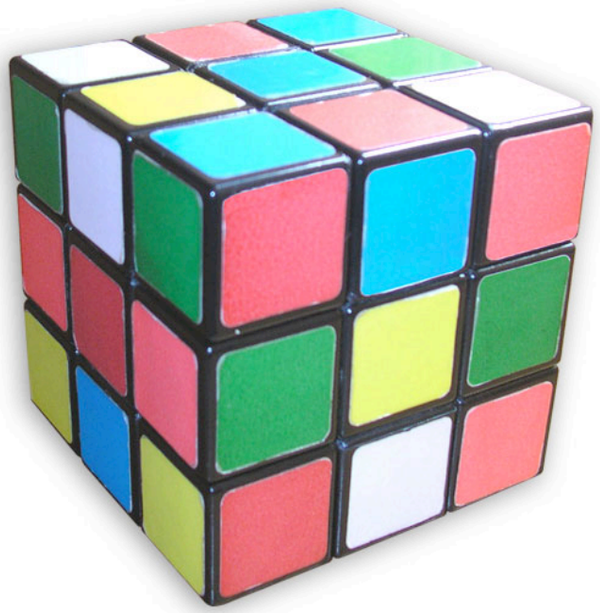
Reuse Process

Reuse Process

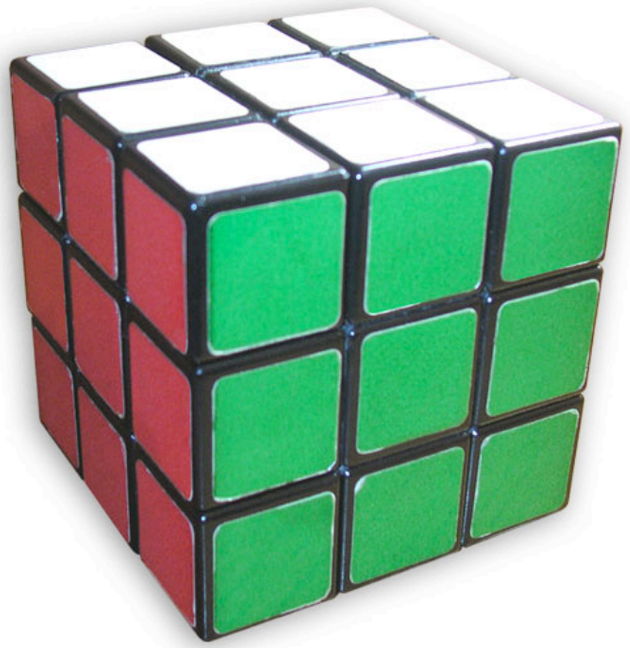


Identification

Reuse Process

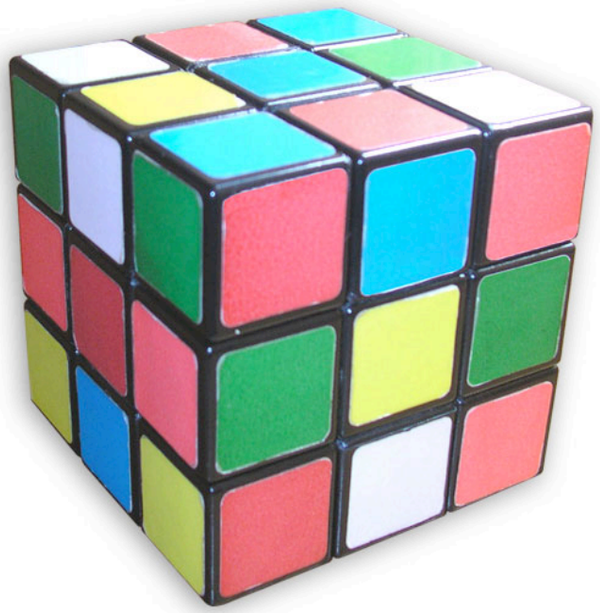


Identification

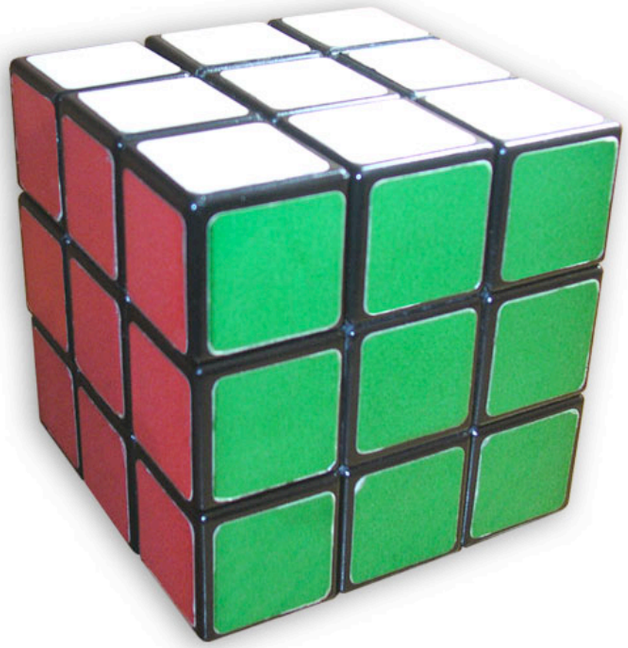


Delineation

Reuse Process



Identification

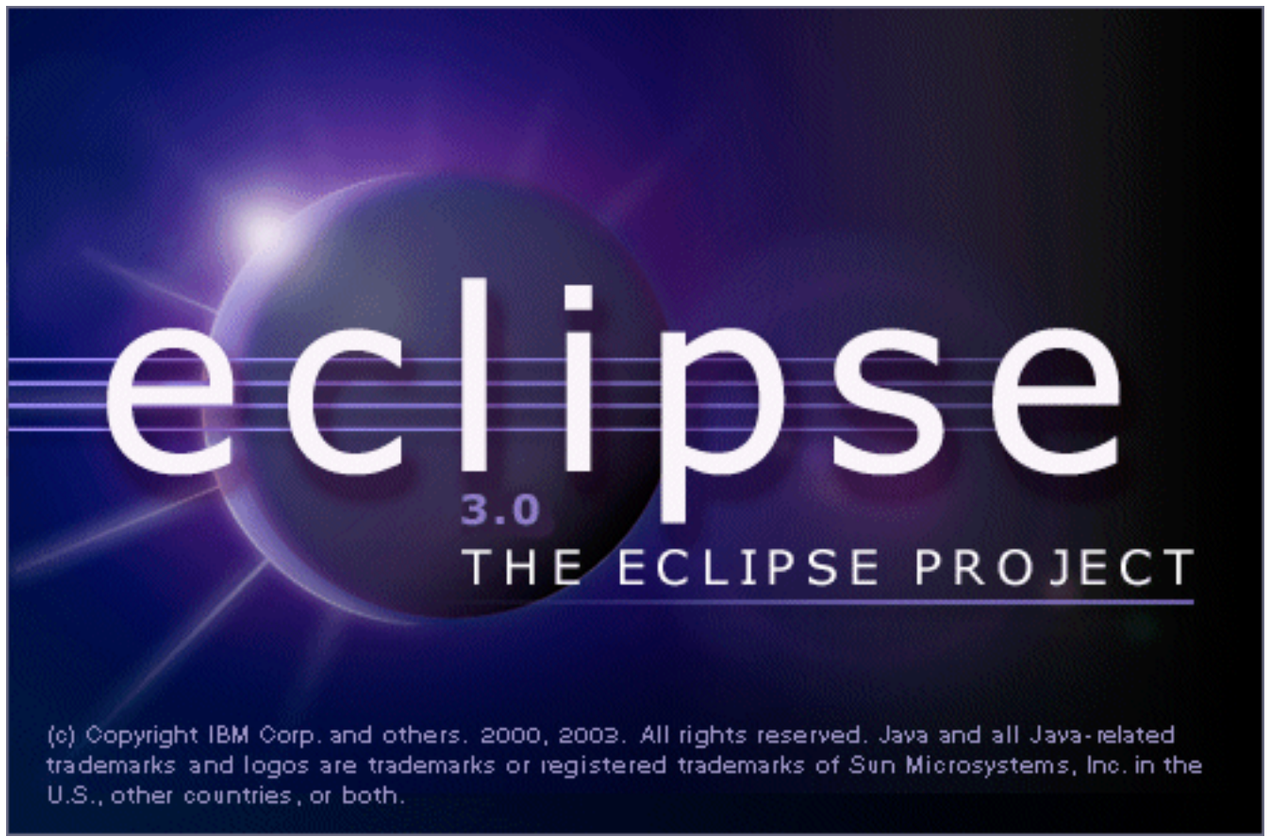


Delineation

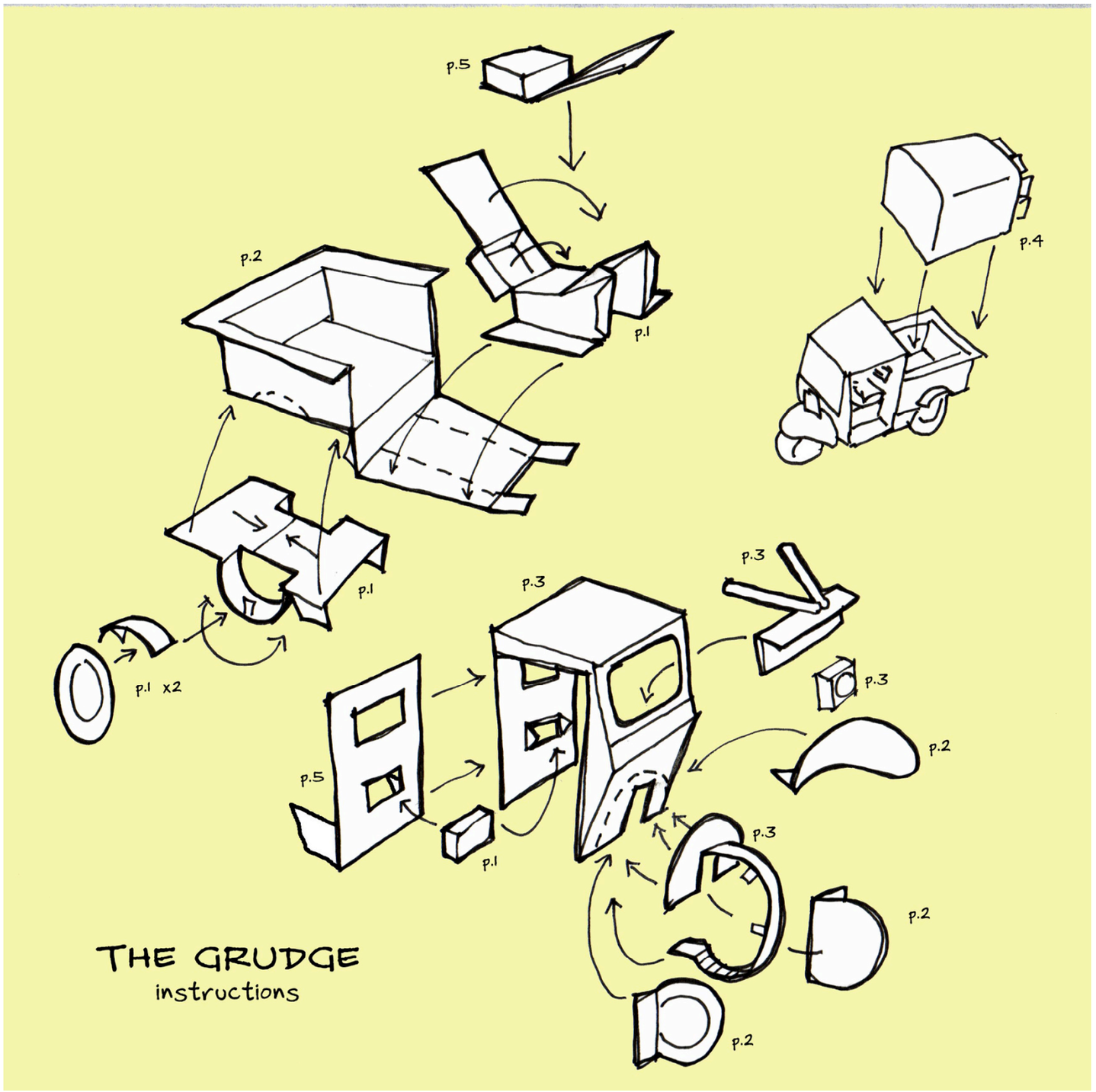
Extraction

Gilligan

a tool to support software reuse...

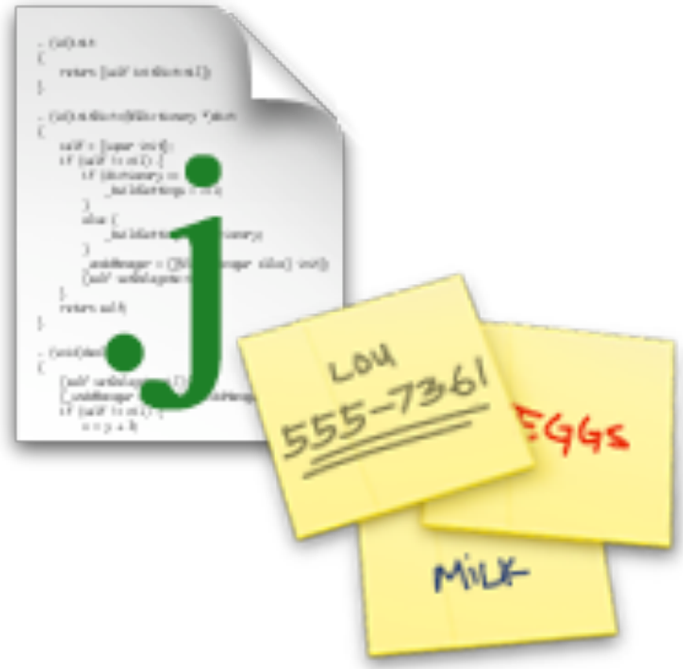


What does Gilligan do?



What does Gilligan do?

Supports navigation



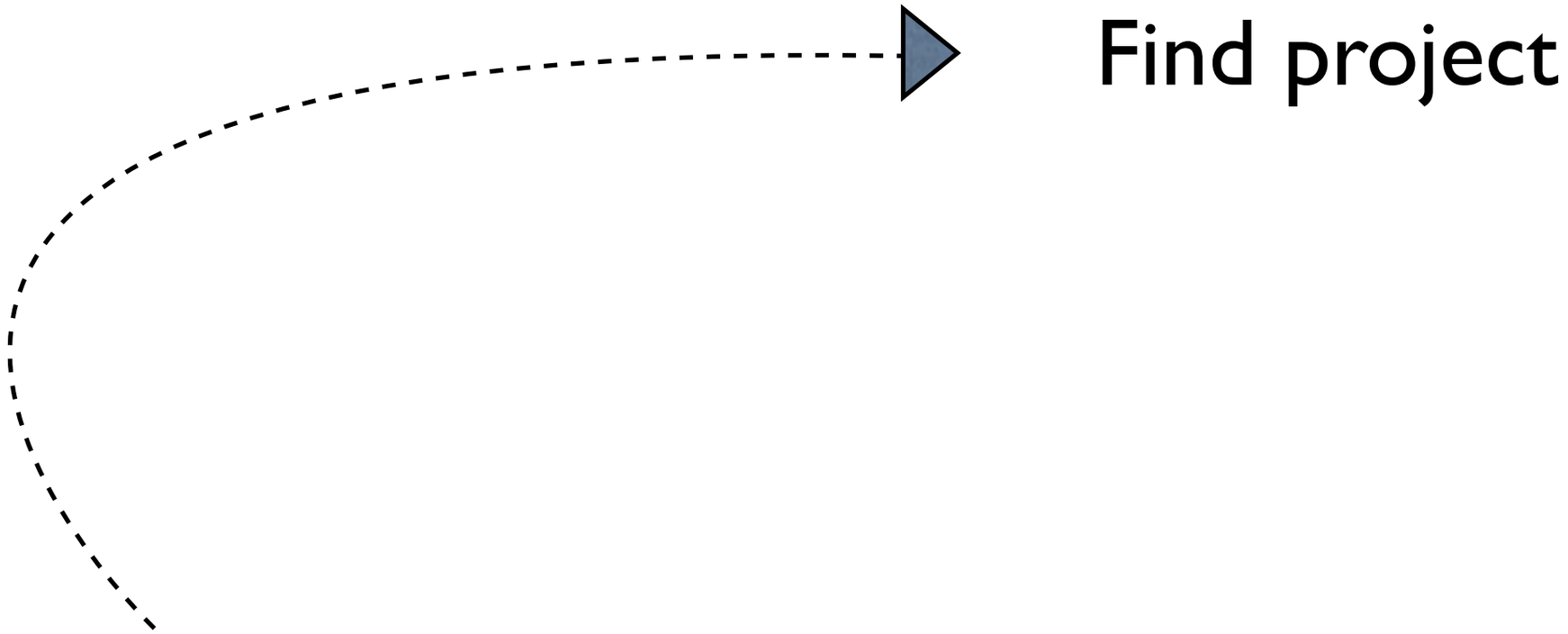
Annotate structurally dependent code fragments

What does Gilligan do?

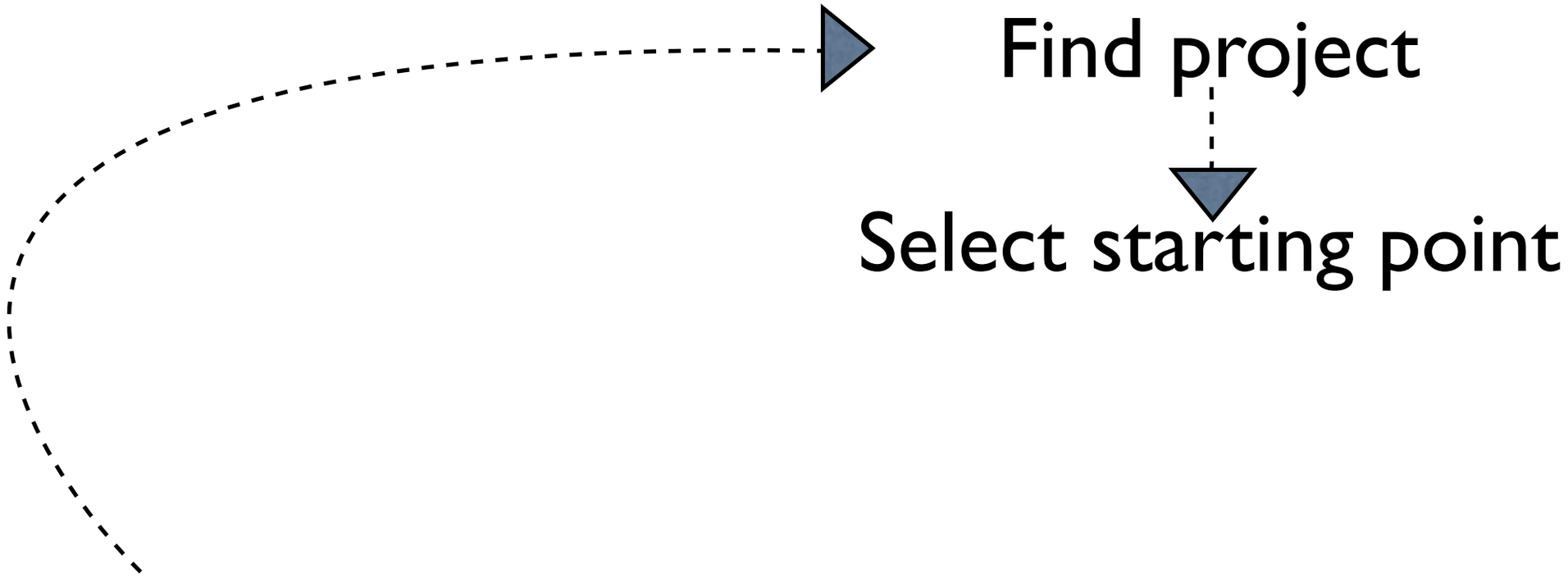
What does Gilligan do?

How does Gilligan work?

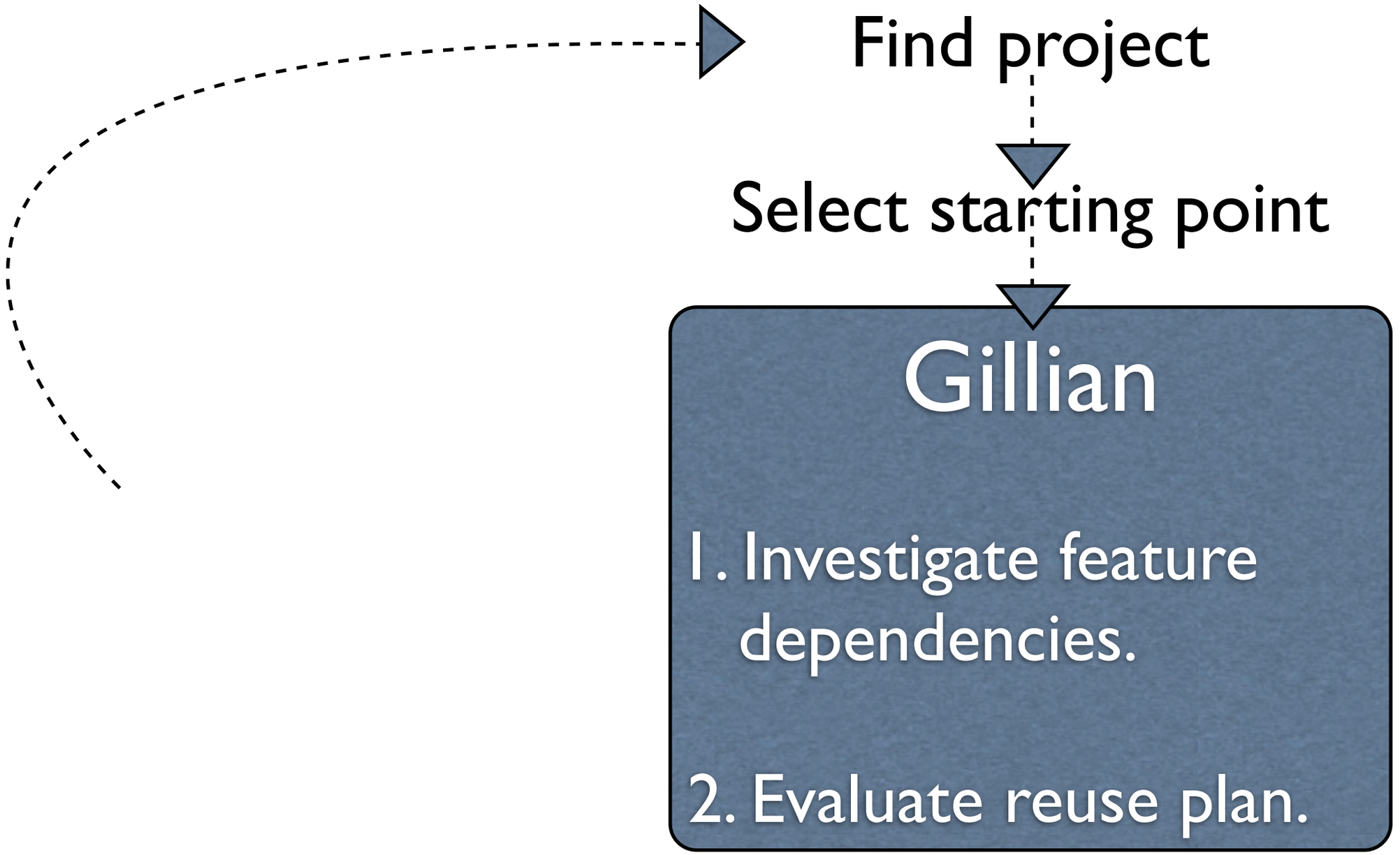
How does Gilligan work?



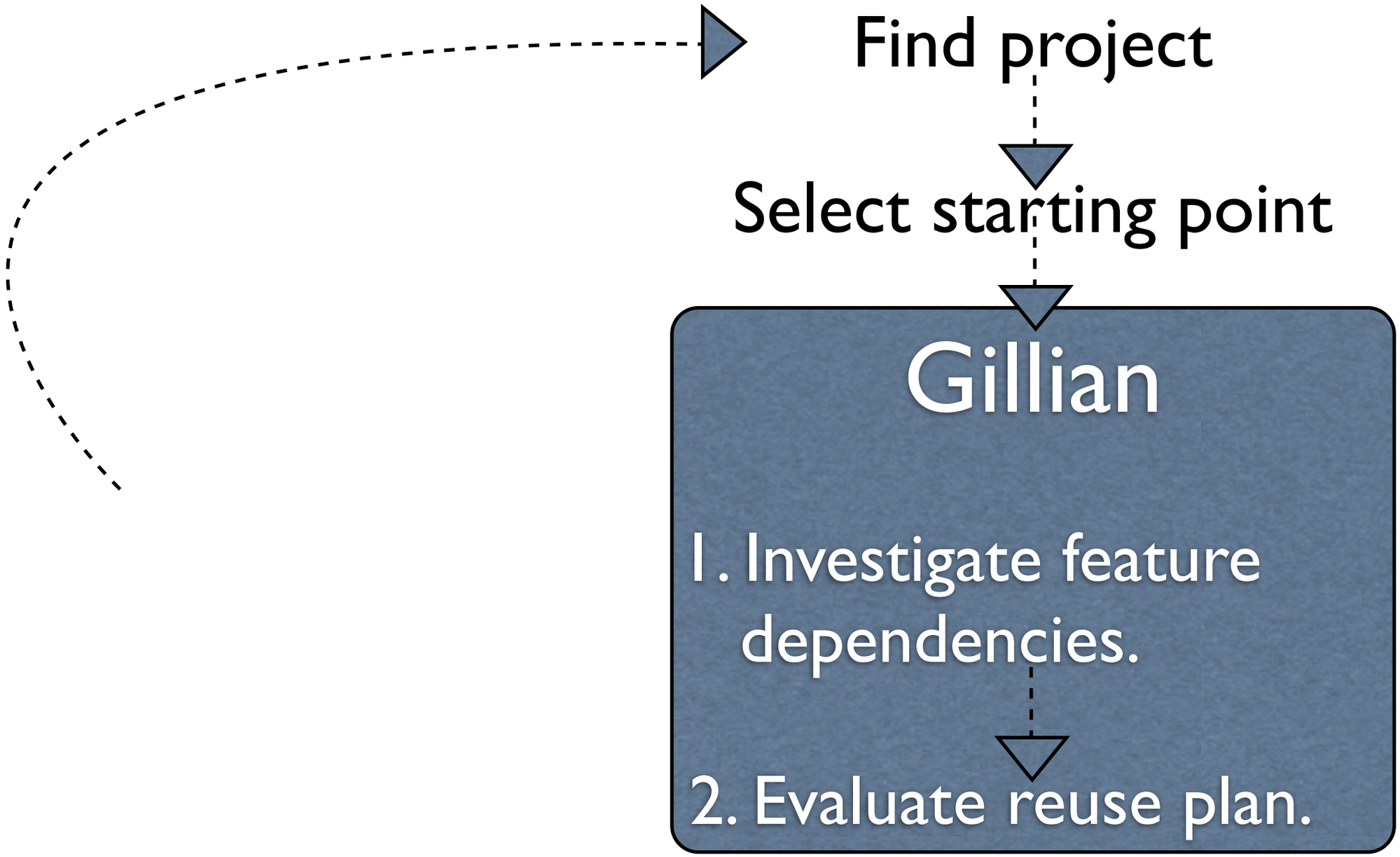
How does Gilligan work?



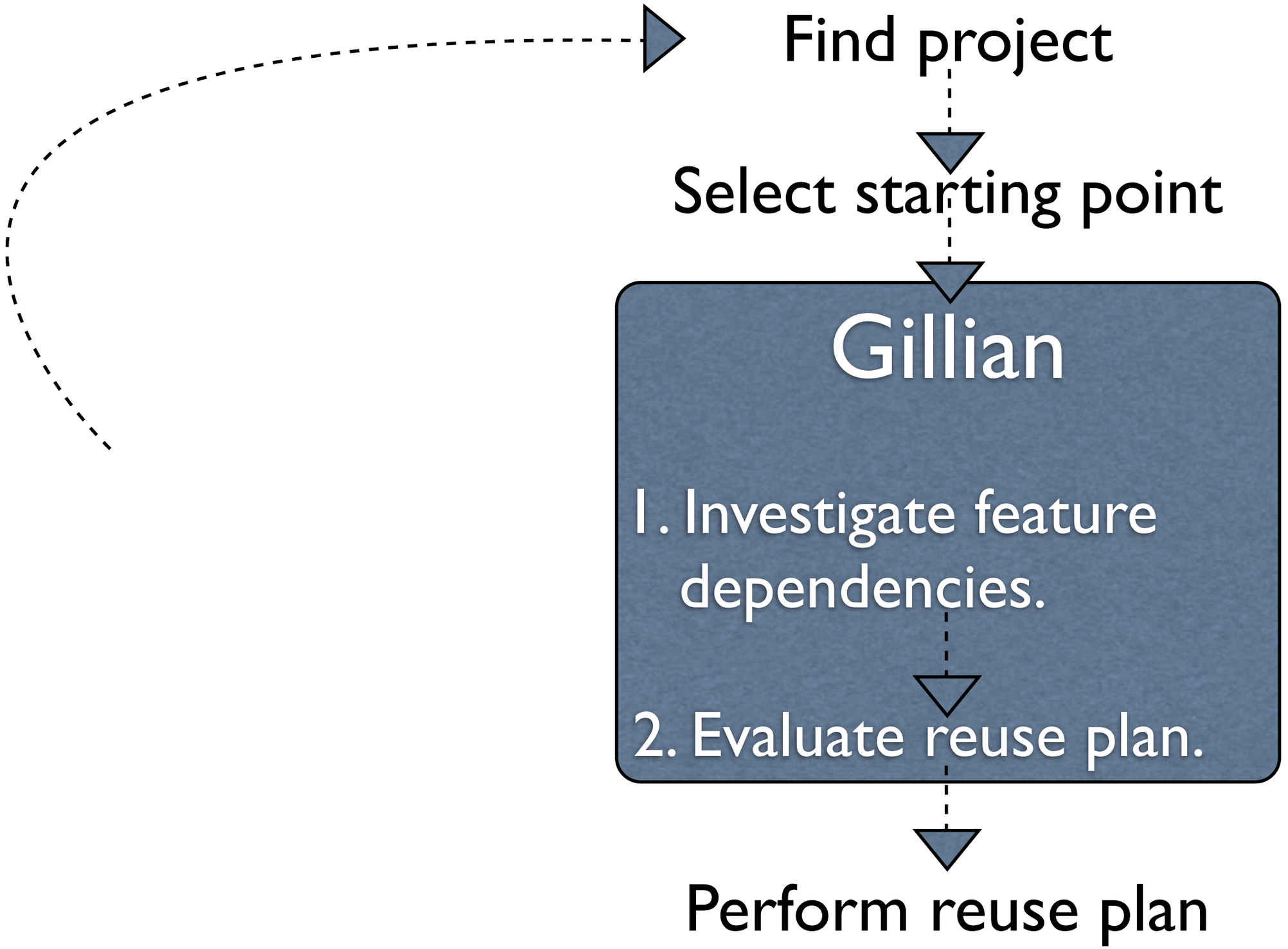
How does Gilligan work?



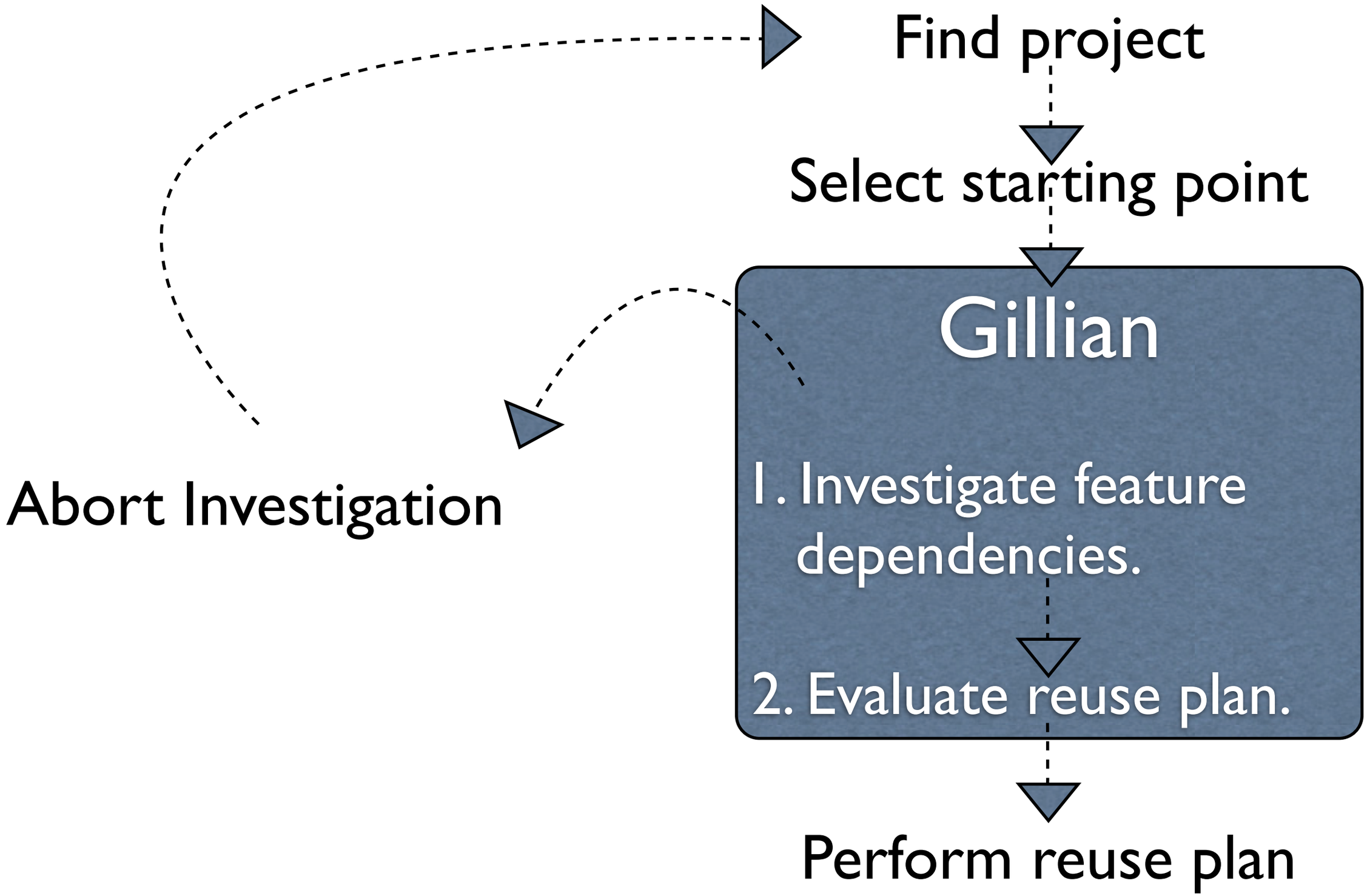
How does Gilligan work?



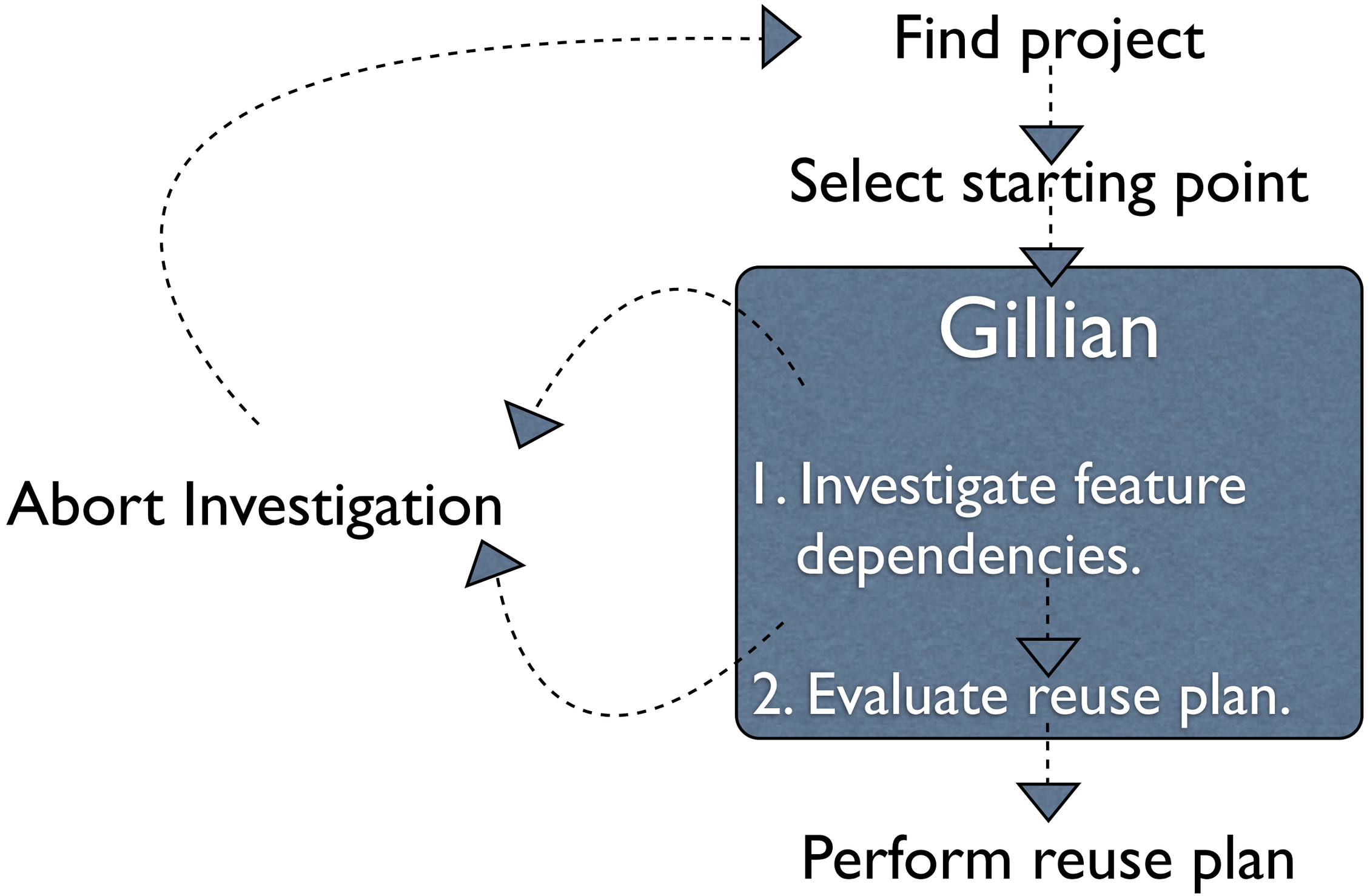
How does Gilligan work?



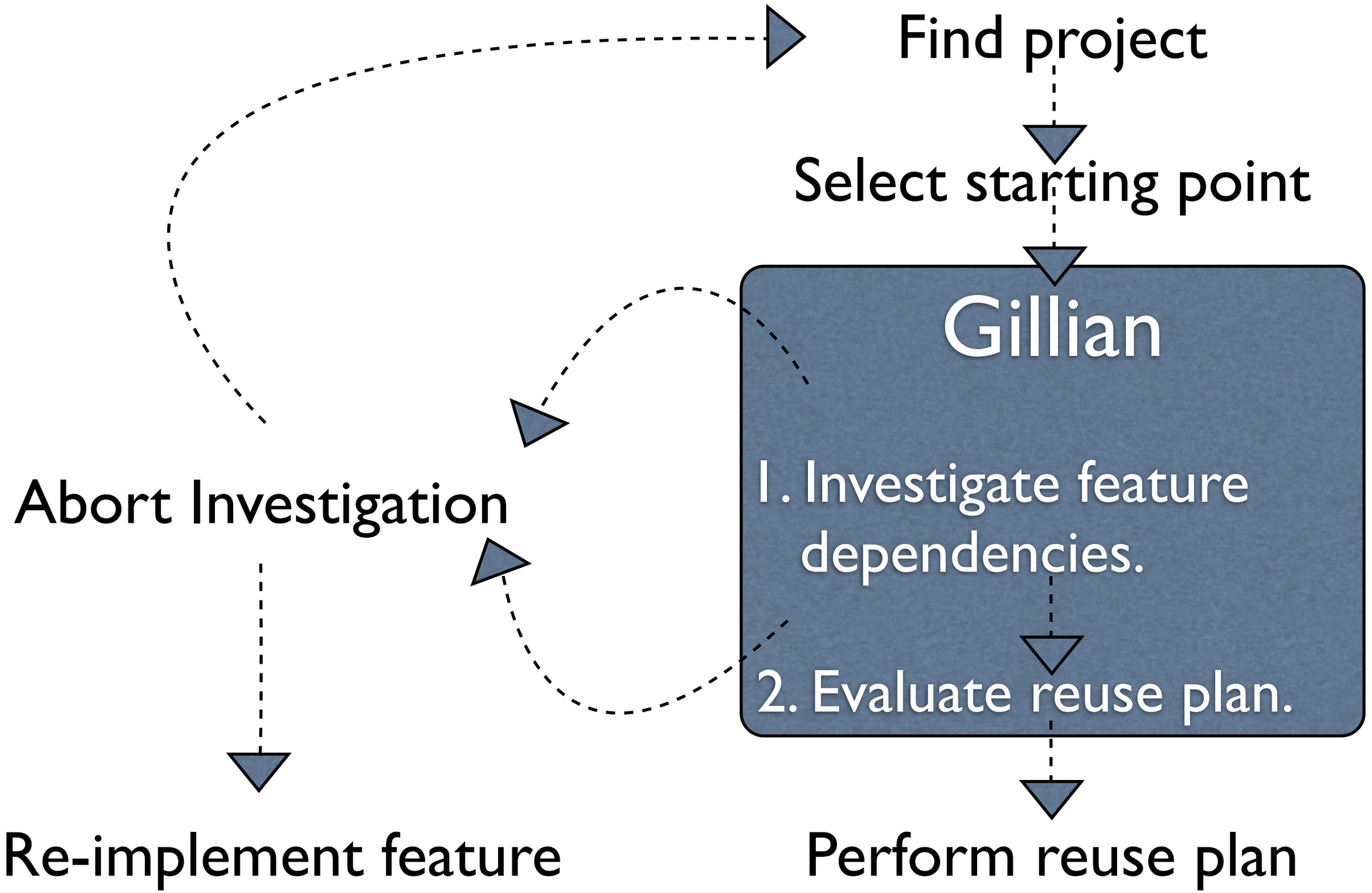
How does Gilligan work?



How does Gilligan work?



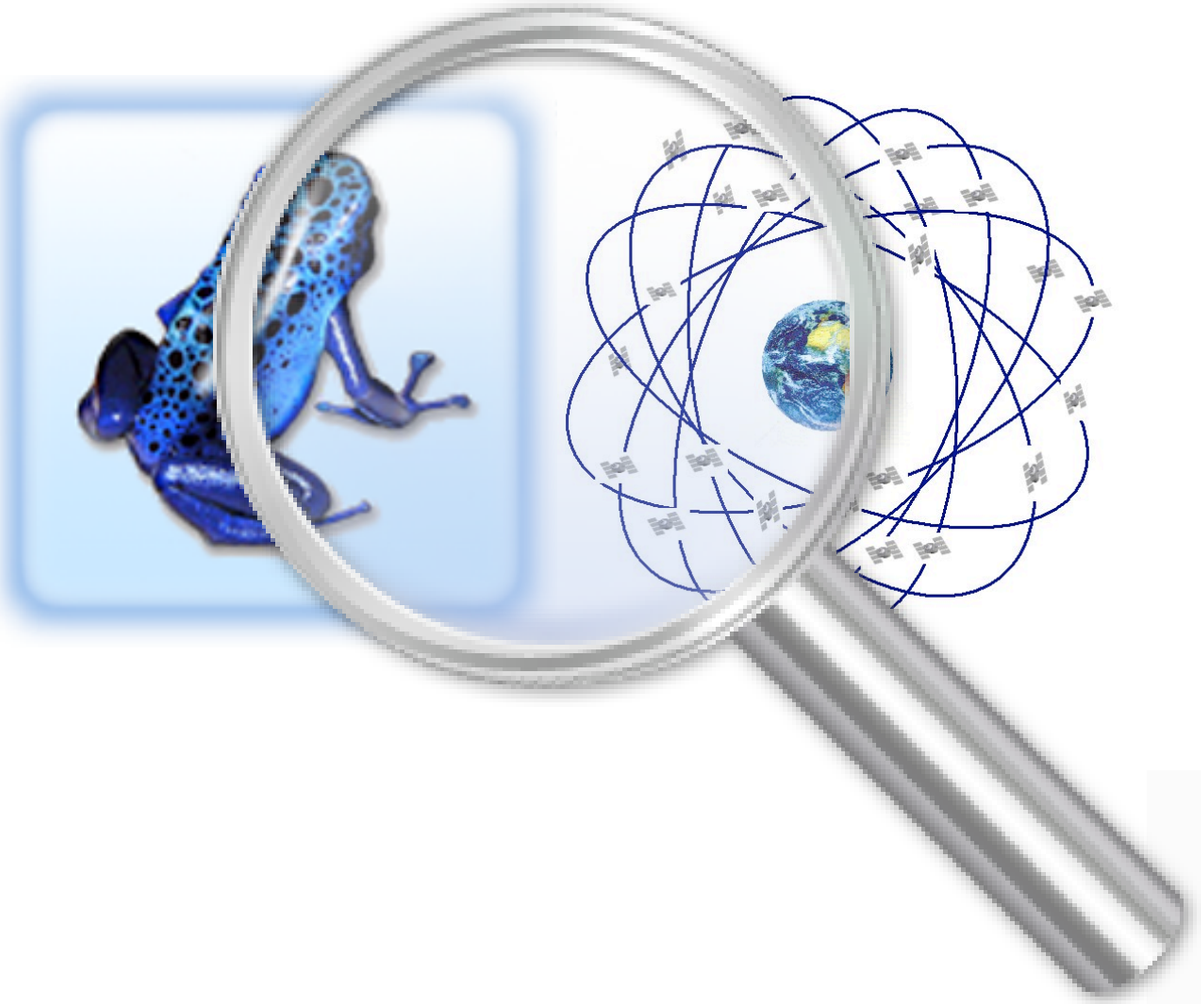
How does Gilligan work?



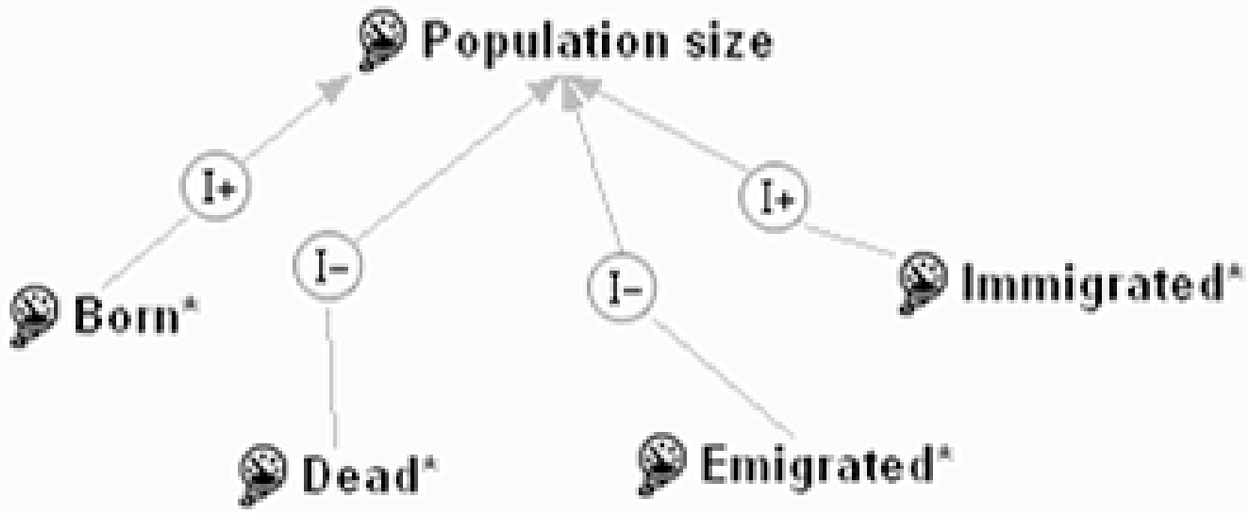
How does Gilligan work?

How does Gilligan work?

How does Gilligan work?



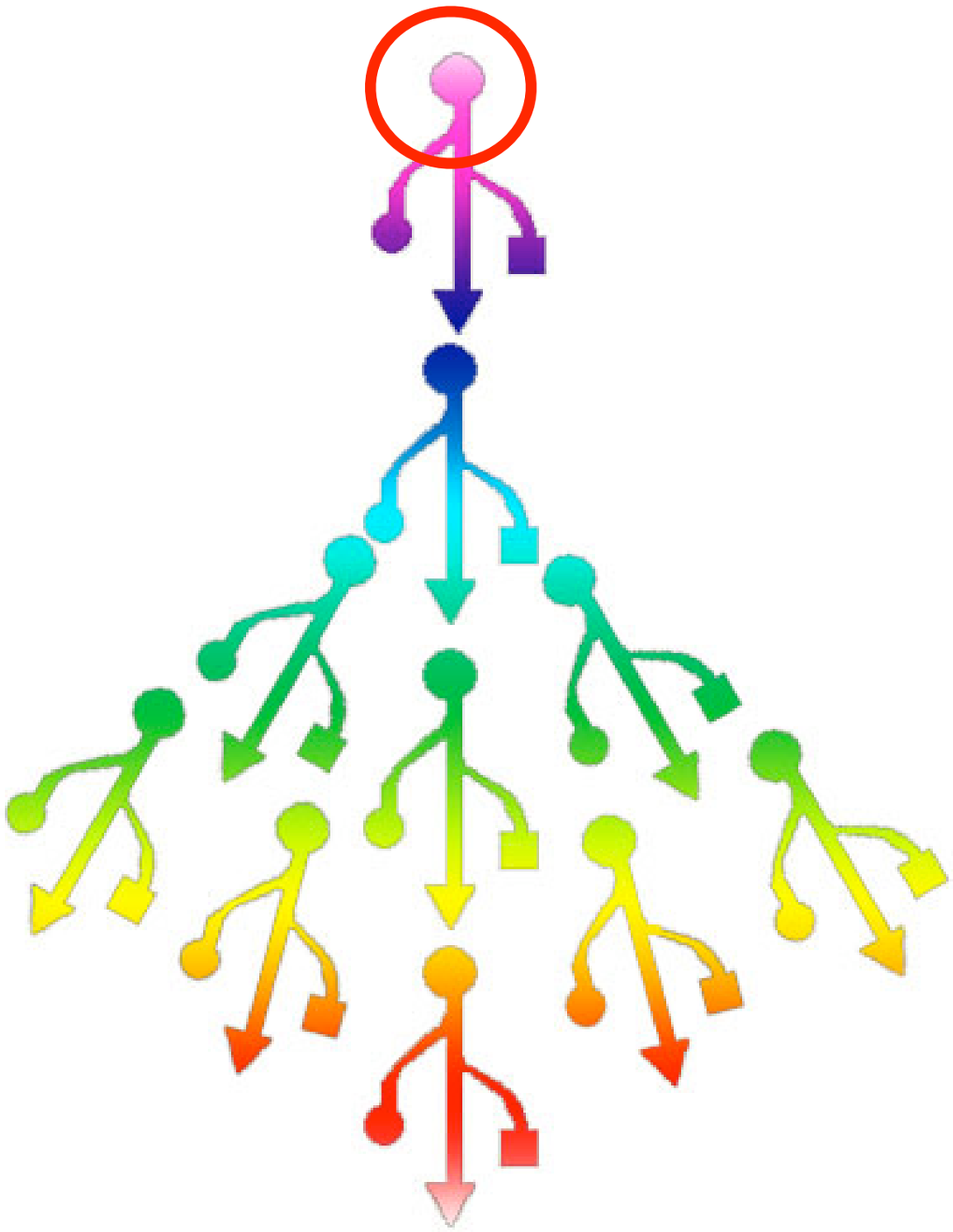
Search for dependencies



How does Gilligan work?

How does Gilligan work?

How does Gilligan work?

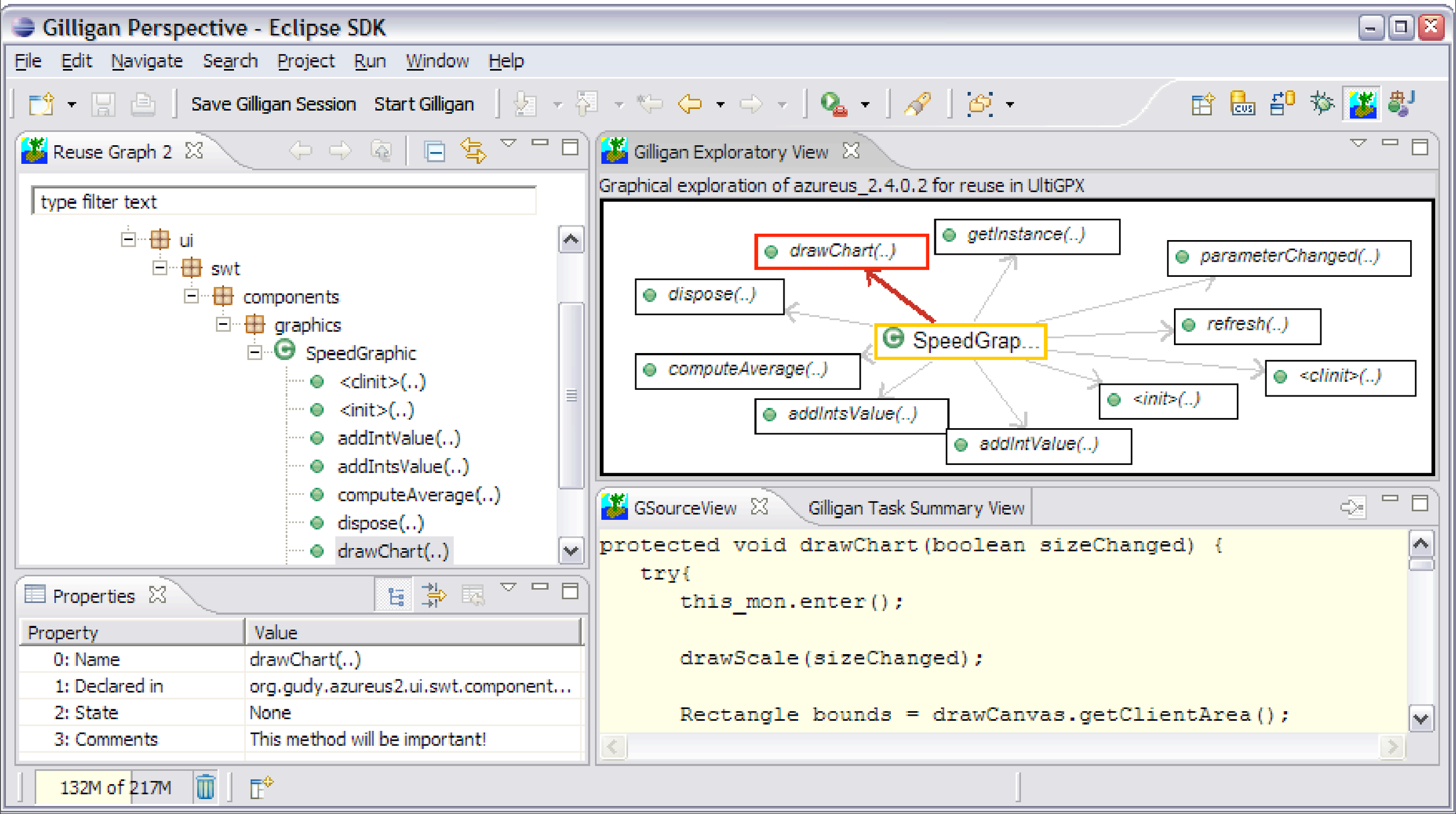


Starting point
(package, class or method)

Relevant dependencies

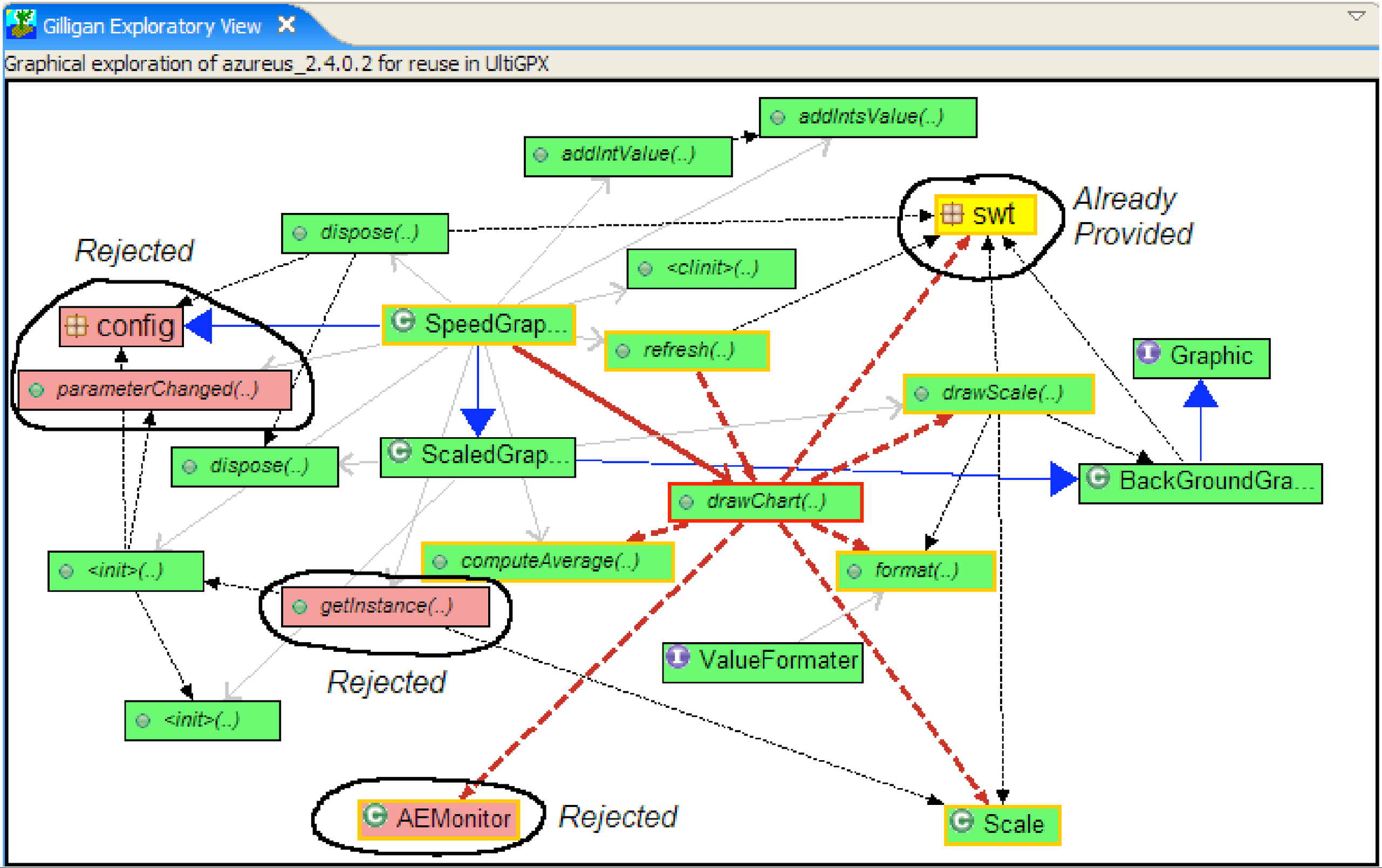
Screenshot of Gilligan

Screenshot of Gilligan

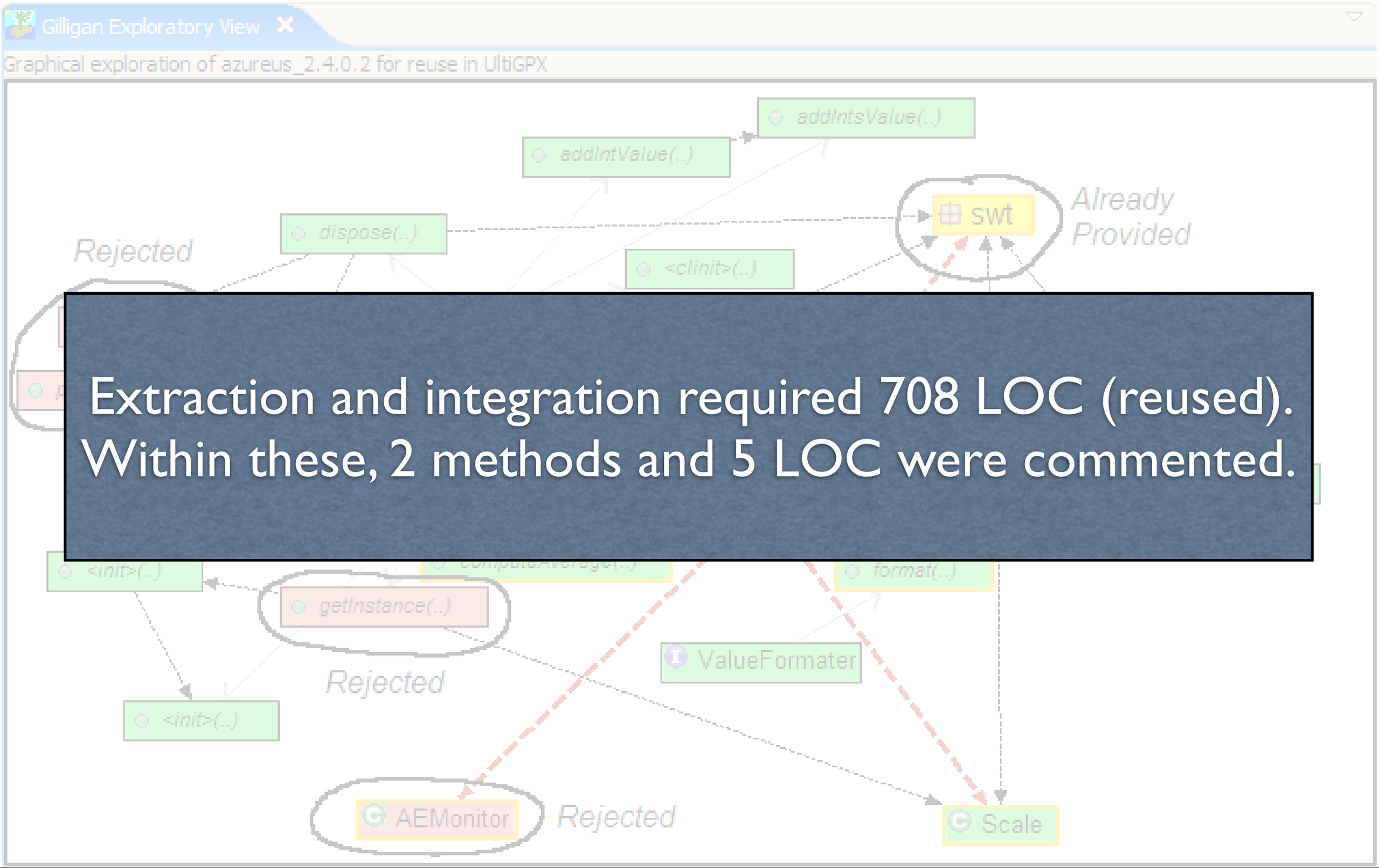


Azureus Reuse Plan

Azureus Reuse Plan



Azureus Reuse Plan



Extraction and integration required 708 LOC (reused).
Within these, 2 methods and 5 LOC were commented.

Reuse in Industry - Survey

#	Question	Strongly Agree	Agree	Somewhat Agree	Neutral	Somewhat Disagree	Disagree	Strongly Disagree
1	I have reused source code	10	2	0	0	0	0	0
2	I have reused whole classes	7	3	2	0	0	0	0
3	I have reused whole features	4	0	1	1	4	2	0
4	I would rather reimplement a feature than reuse an existing one	0	0	2	1	3	4	2
5	I reuse code to save time	6	4	2	0	0	0	0
6	I reuse code to increase reliability	4	7	0	1	0	0	0
7	Keeping track of the relevant details of a piece of source code while navigating its text can be difficult	2	6	2	1	0	1	0
8	Understanding what dependencies a feature has on its context is important for me to determine whether I should reuse it	7	5	0	0	0	0	0
9	My organization has a large amount of code available to be reused	5	1	4	1	1	0	0
10	Portions of features I am developing already exist	1	4	5	0	1	1	0

Evaluation - Case Study

From SWT framework (65 kloc in 458 classes), extract...

BMP implementation

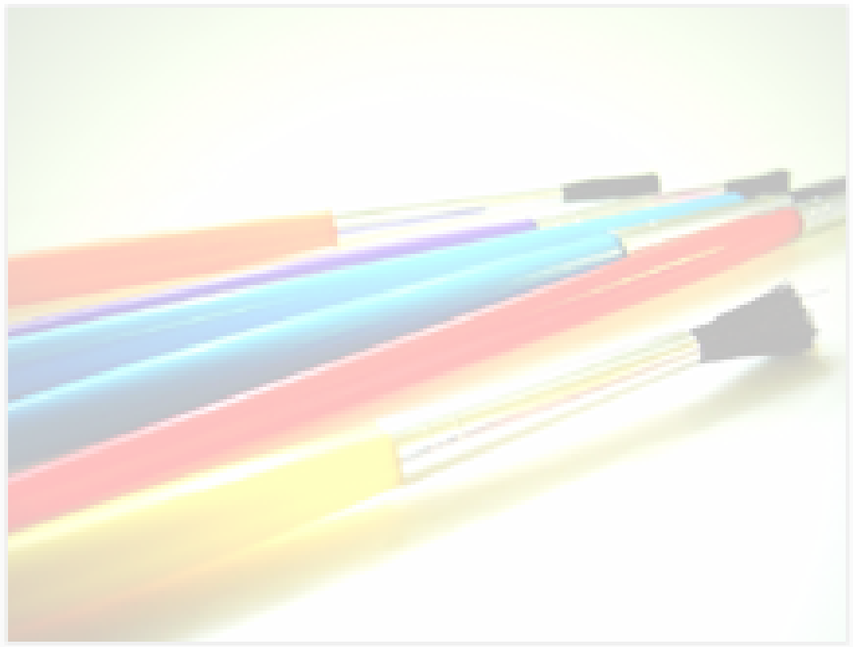
PNG implementation

Evaluation - Case Study

BMP extraction

PNG extraction

Evaluation - Case Study



BMP extraction

PNG extraction

Evaluation - Case Study

- Started with WinBMPFileFormat.
- Led him to LEDataInputStream and ImageData.
- Reused 497 LOC.
- Of 14,081 nodes in SWT, only 60 visited.
- 38 accepted, 16 rejected, remapped 2 and 4 already provided.

BMP extraction

PNG extraction

Evaluation - Case Study

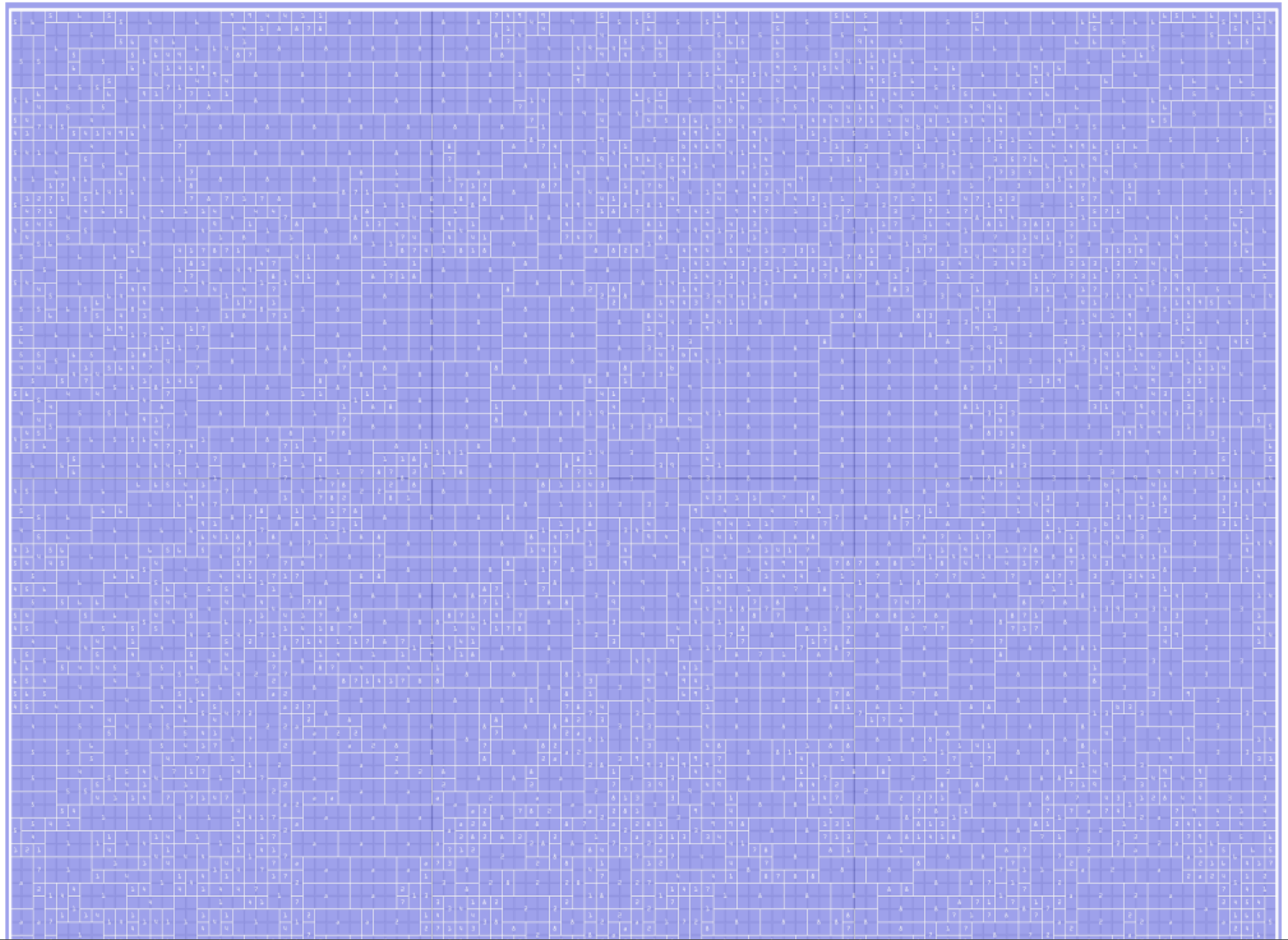
- Started with WinBMPFileFormat.
- Led him to LEDataInputStream and ImageData.
- Reused 497 LOC.
- Of 14,081 nodes in SWT, only 60 visited.
- 38 accepted, 16 rejected, remapped 2 and 4 already provided.

BMP extraction

- Started with a 471 LOC class.
- Of 92 nodes on screen, 20 were relevant to him.
- He marked 20 accepted, 2 rejected and 1 remapped.
- Eventually he reused 23 classes (approx 3 kloc).
- Changed mind often!

PNG extraction

Strathcona



Java 6 API

Overview (Java Platform SE 6)

http://java.sun.com/javase/6/docs/api/

Getting Started Latest Headlines

Java™ Platform Standard Ed. 6

All Classes

Packages
java.applet
java.awt
java.awt.color
java.awt.datatransfer
java.awt.dnd

All Classes
AbstractAction
AbstractAnnotationValueVisitor6
AbstractBorder
AbstractButton
AbstractCellEditor
AbstractCollection
AbstractColorChooserPanel
AbstractDocument
AbstractDocument.AttributeCon
AbstractDocument.Content
AbstractDocument.ElementEdit
AbstractElementVisitor6
AbstractExecutorService
AbstractInterruptibleChannel
AbstractLayoutCache
AbstractLayoutCache.NodeDime
AbstractList
AbstractListModel
AbstractMap
AbstractMap.SimpleEntry
AbstractMap.SimpleImmutableE
AbstractMarshallerImpl
AbstractMethodError
AbstractOwnableSynchronizer
AbstractPreferences
AbstractProcessor
AbstractQueue
AbstractQueuedLongSynchroniz
AbstractQueuedSynchronizer
AbstractScriptEngine

Overview Package Class Use Tree Deprecated Index Help

PREV NEXT FRAMES NO FRAMES

Java™ Platform Standard Ed. 6

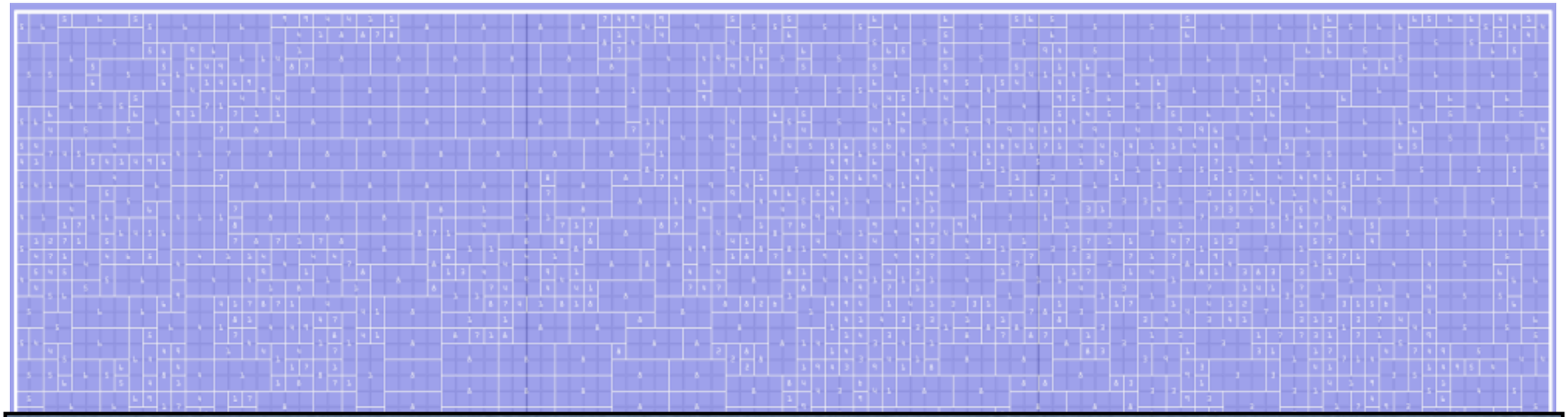
Java™ Platform, Standard Edition 6 API Specification

This document is the API specification for version 6 of the Java™ Platform, Standard Edition.

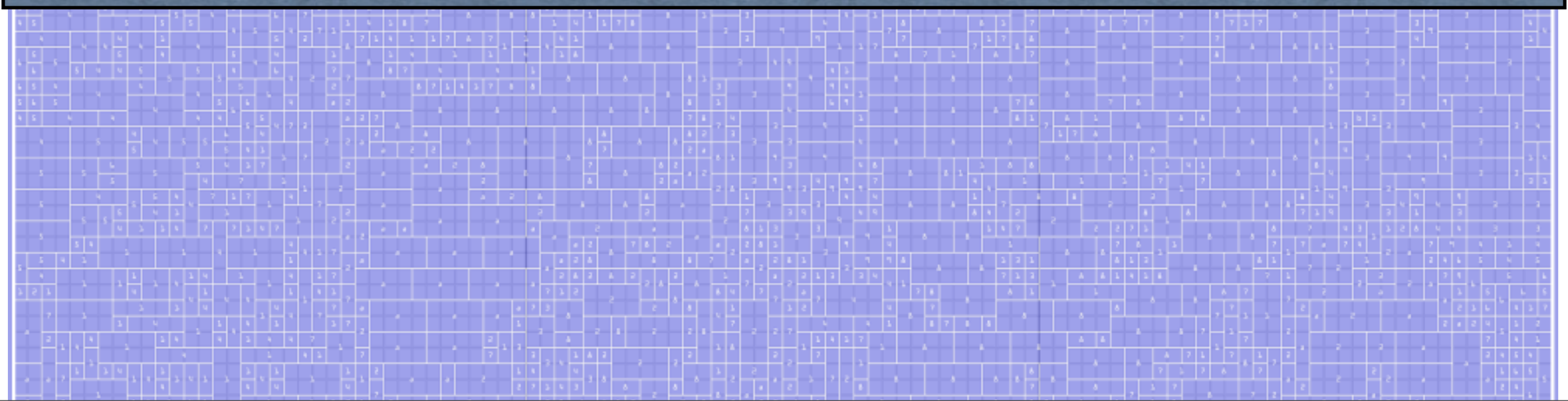
See:
Description

Packages	
java.applet	Provides the classes necessary to create an applet and the classes an applet uses to communicate with its applet context.
java.awt	Contains all of the classes for creating user interfaces and for painting graphics and images.
java.awt.color	Provides classes for color spaces.
java.awt.datatransfer	Provides interfaces and classes for transferring data between and within applications.
java.awt.dnd	Drag and Drop is a direct manipulation gesture found in many Graphical User Interface systems that provides a mechanism to transfer information between two entities logically associated with presentation elements in the GUI.
java.awt.event	Provides interfaces and classes for dealing with different types of events fired by AWT components.
java.awt.font	Provides classes and interface relating to fonts.
java.awt.geom	Provides the Java 2D classes for defining and performing operations on objects related to two-dimensional geometry.
java.awt.im	Provides classes and interfaces for the input method framework.
java.awt.im.spi	Provides interfaces that enable the development of input methods that can be used with any Java runtime environment.
java.awt.image	Provides classes for creating and modifying images.

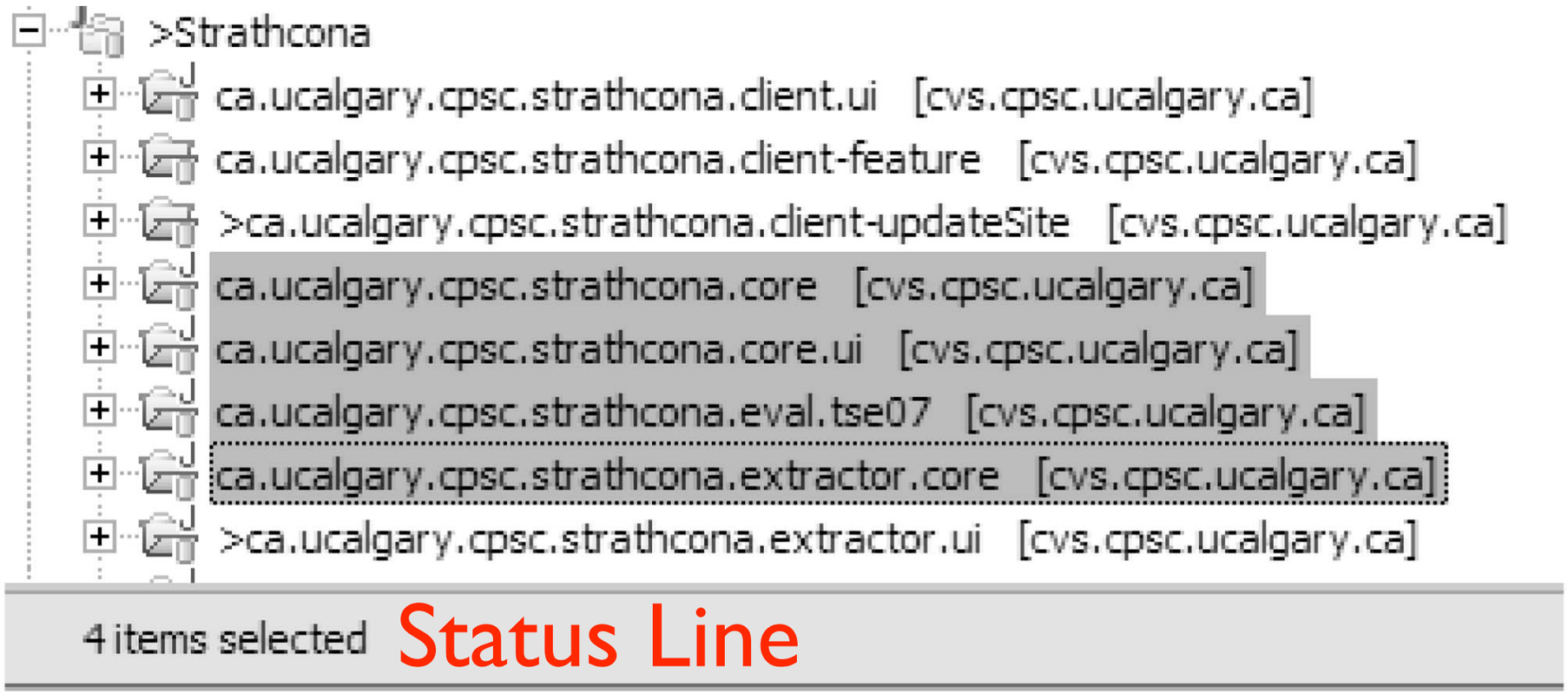
Strathcona



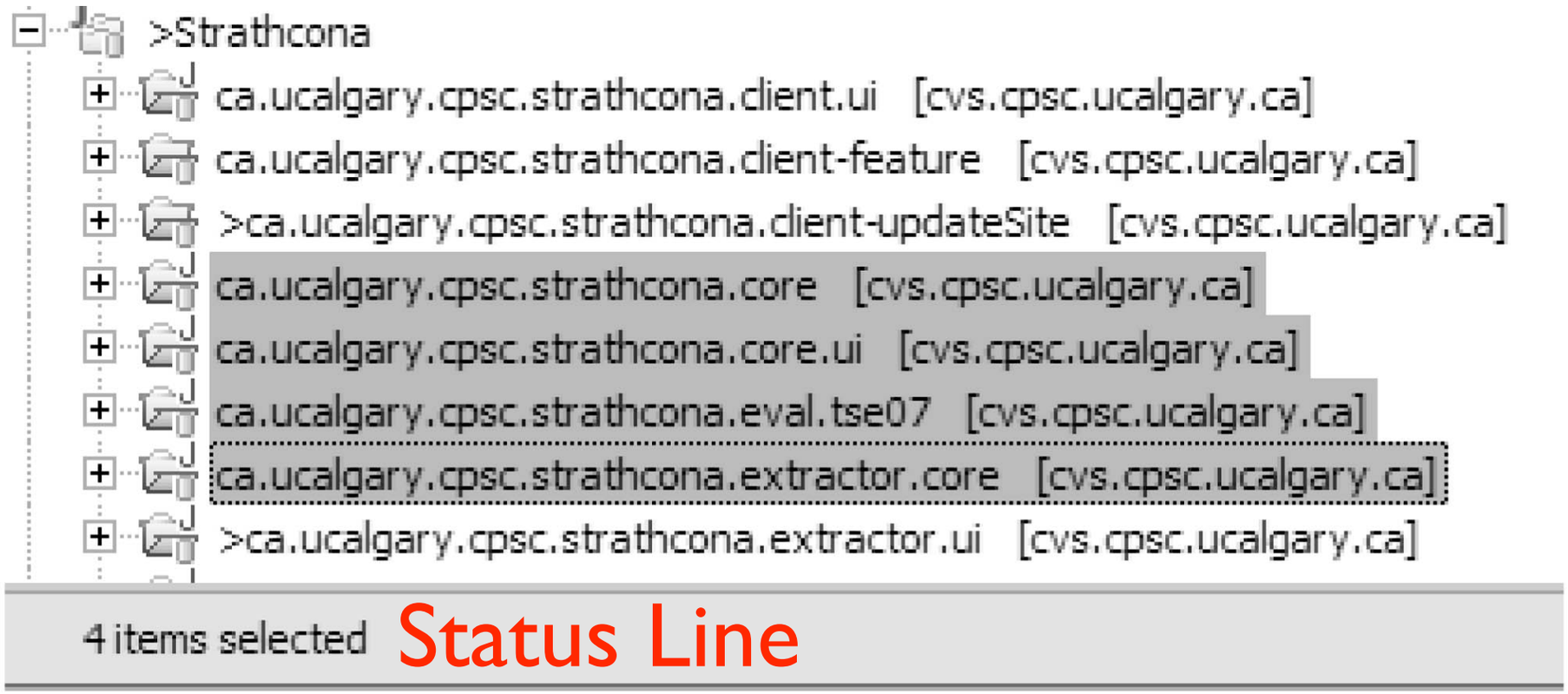
To ease the burden of the developer, Strathcona uses the structure of the source code to find relevant examples in a repository.



Example Task



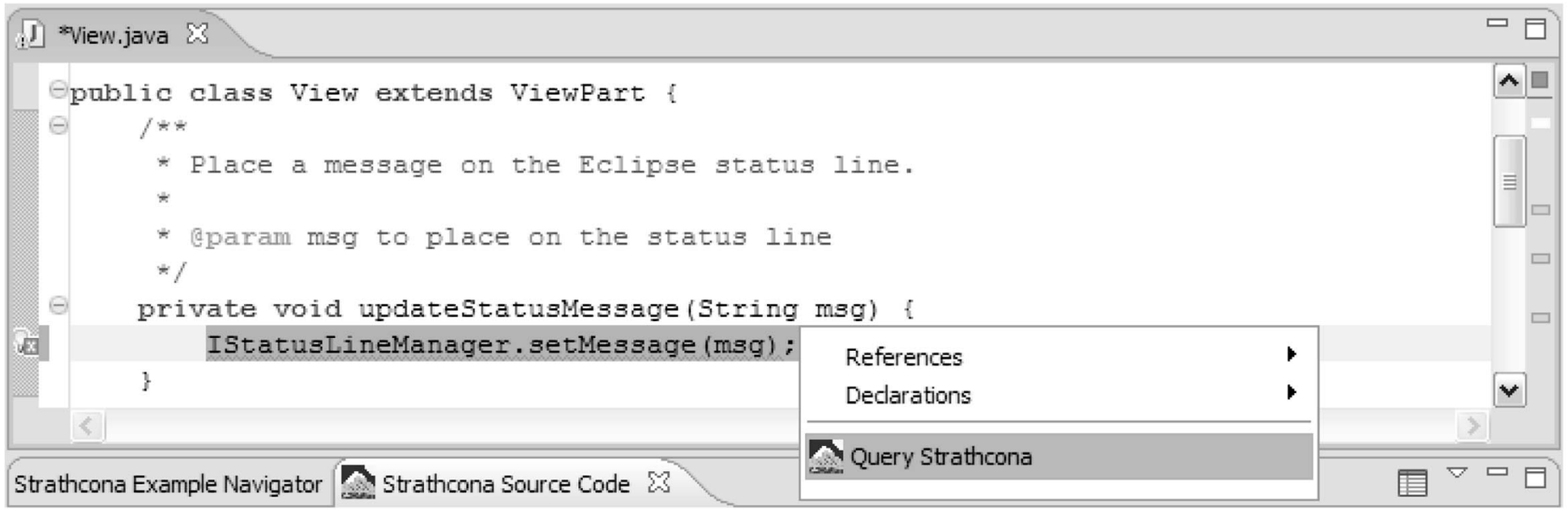
Example Task



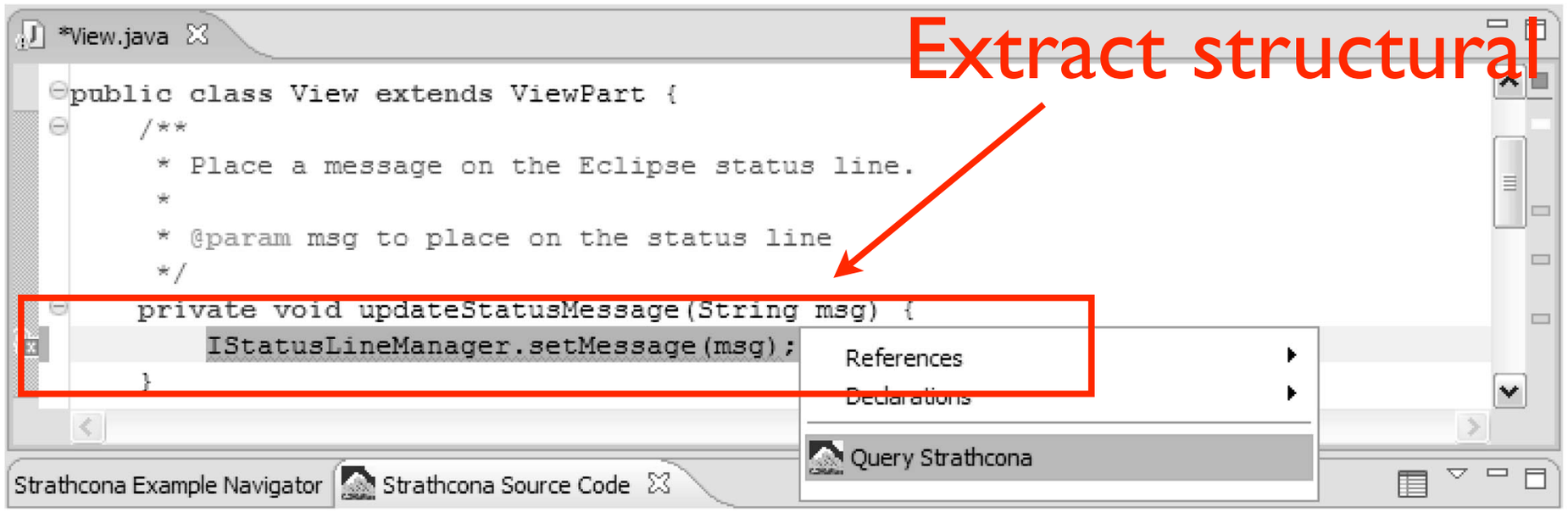
API Documentation for IStatusLineManager

Method Summary	
void	setMessage(String message) Sets the message text to be displayed on the status line.

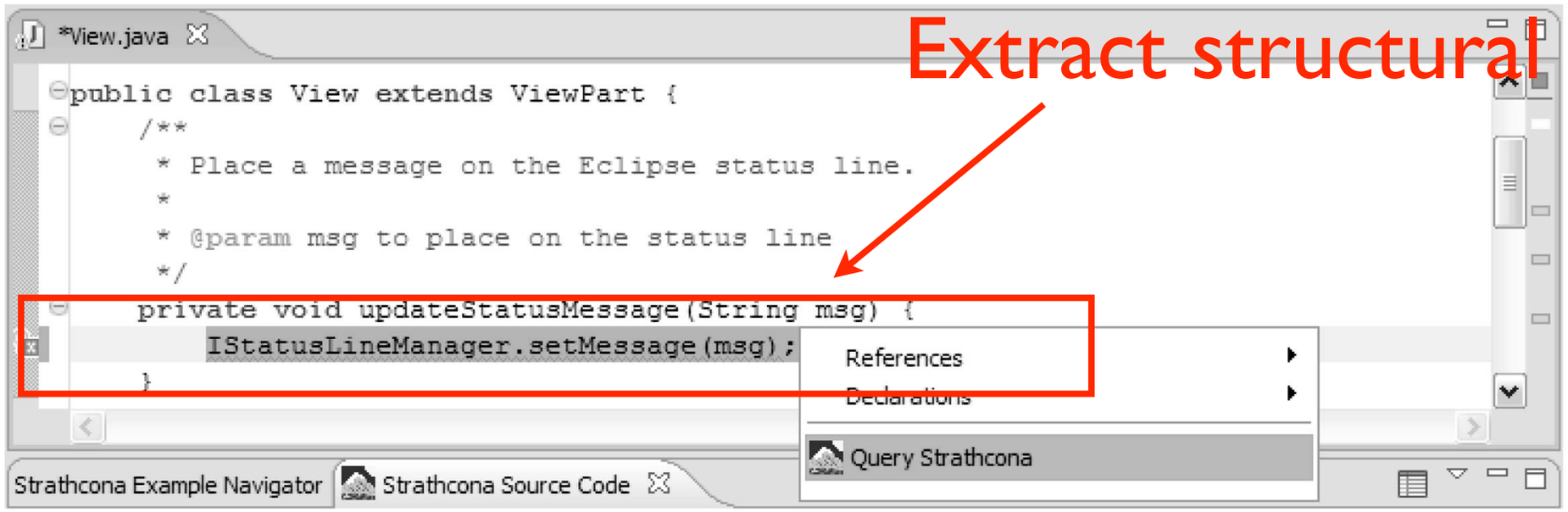
Client: Determining Structural Context



Client: Determining Structural Context



Client: Determining Structural Context



- declaring type is View
- declares method updateStatusMessage.java
- extends org.eclipse.ui.part.ViewPart
- ...

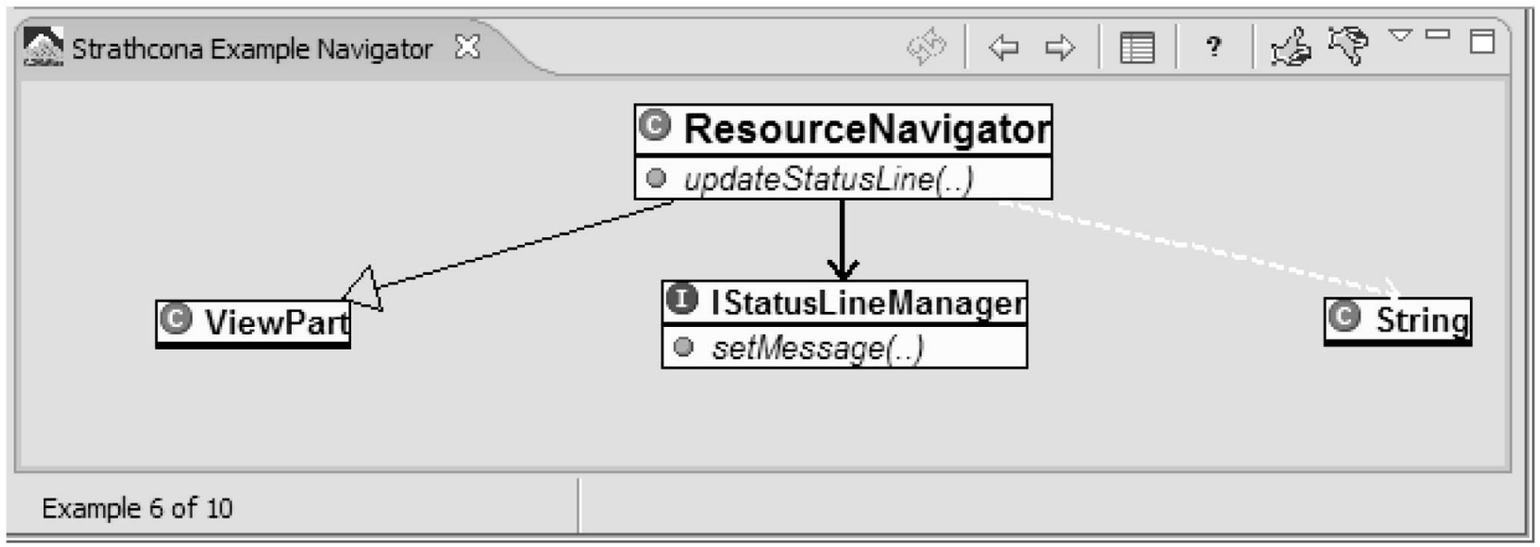
Server Repository

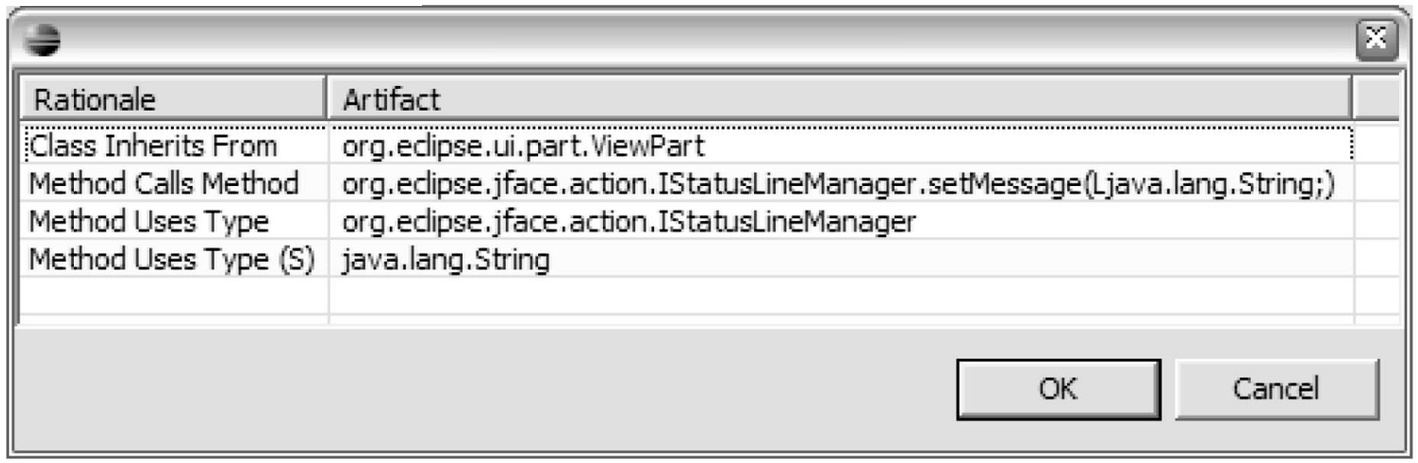
Conditions for addition

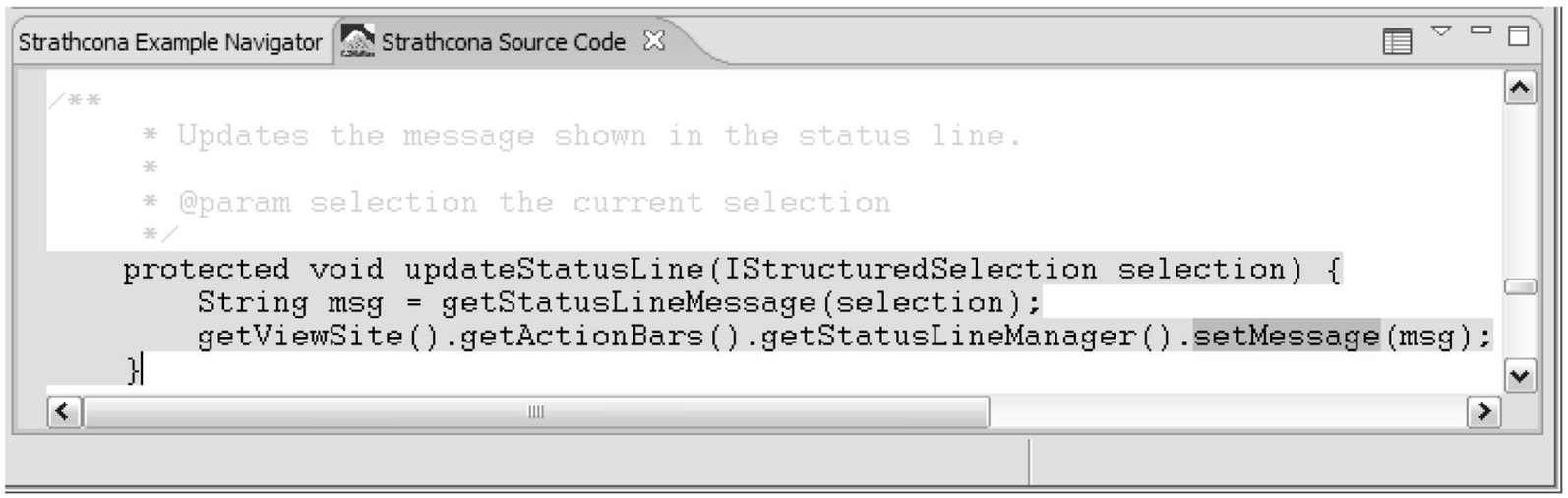
- Code must be parsable by Eclipse compiler.
- Code must represent good usage of API.

Structural contexts are stored in PostGRES tables.

Output Examples







Code Search Engines

SPARSJ

koders

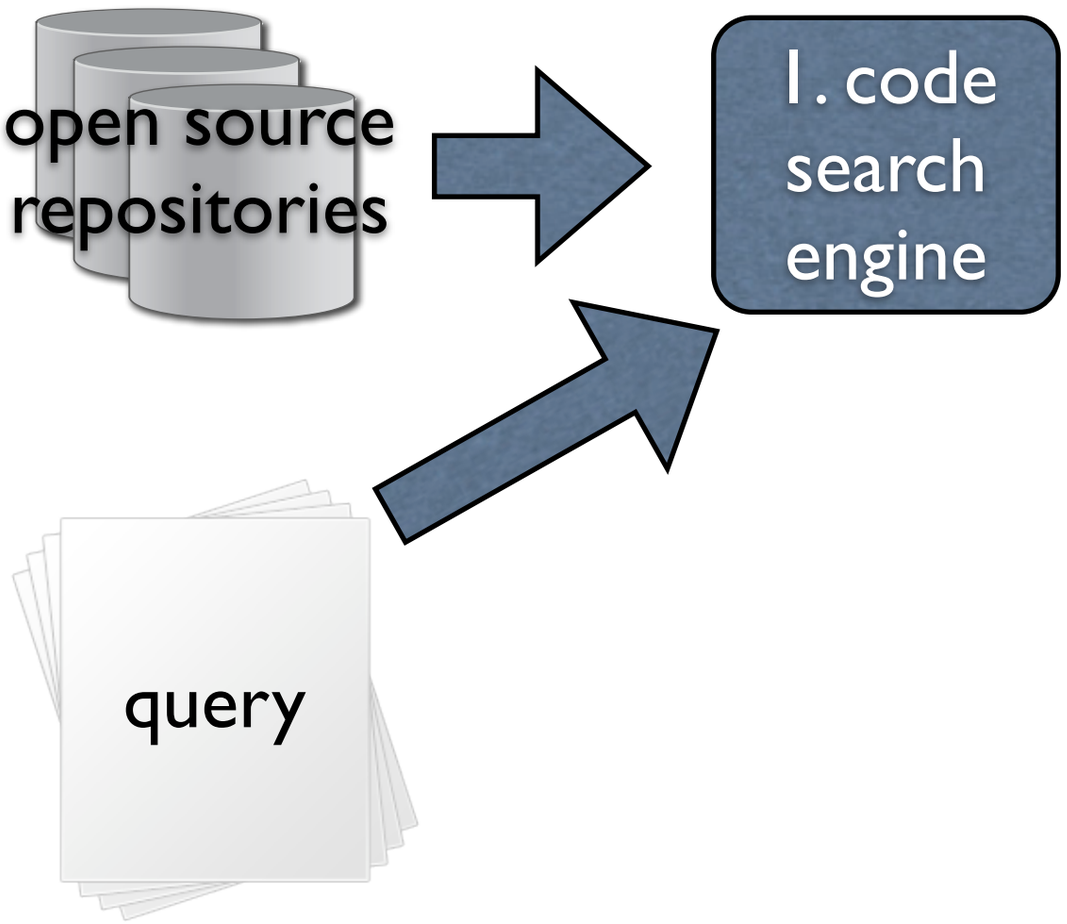
Google™ /*Code Search*/ LABS

Docjar

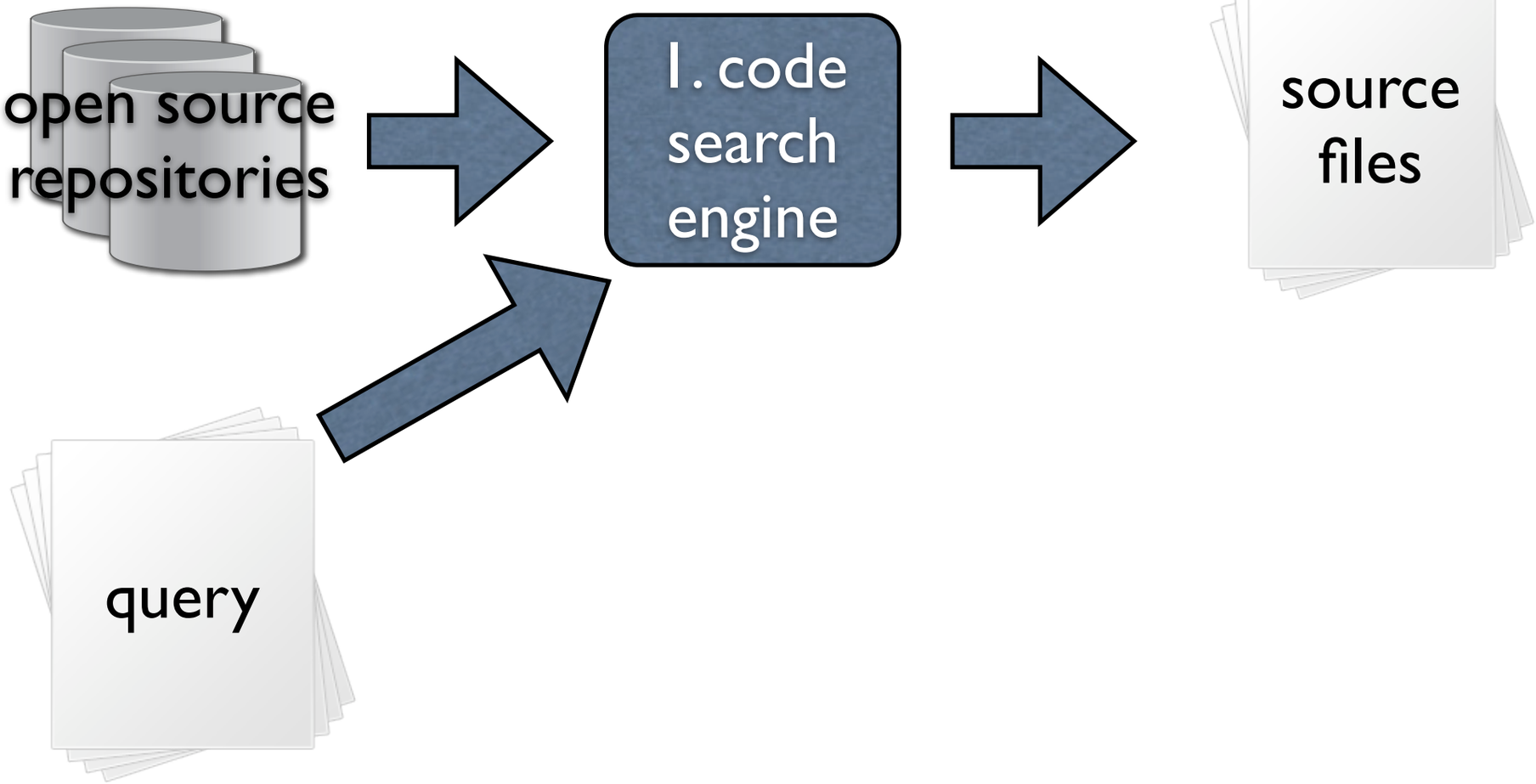
MAPO

MAPO

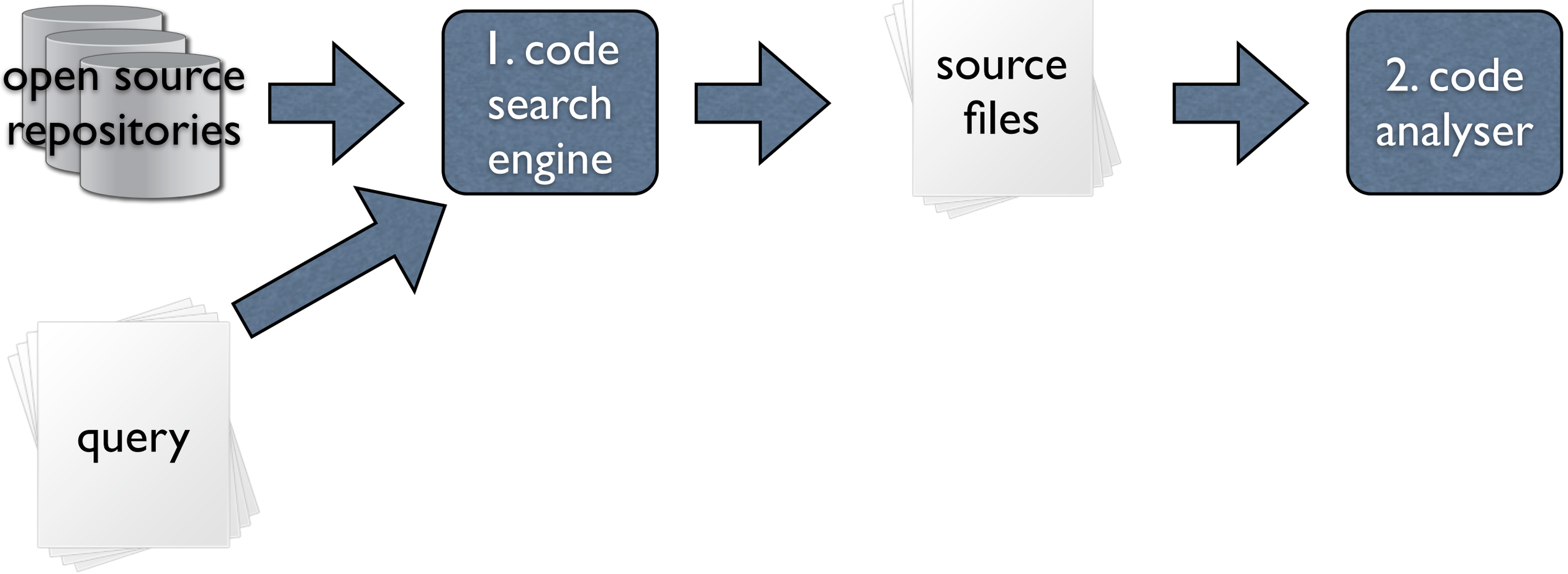
MAPO



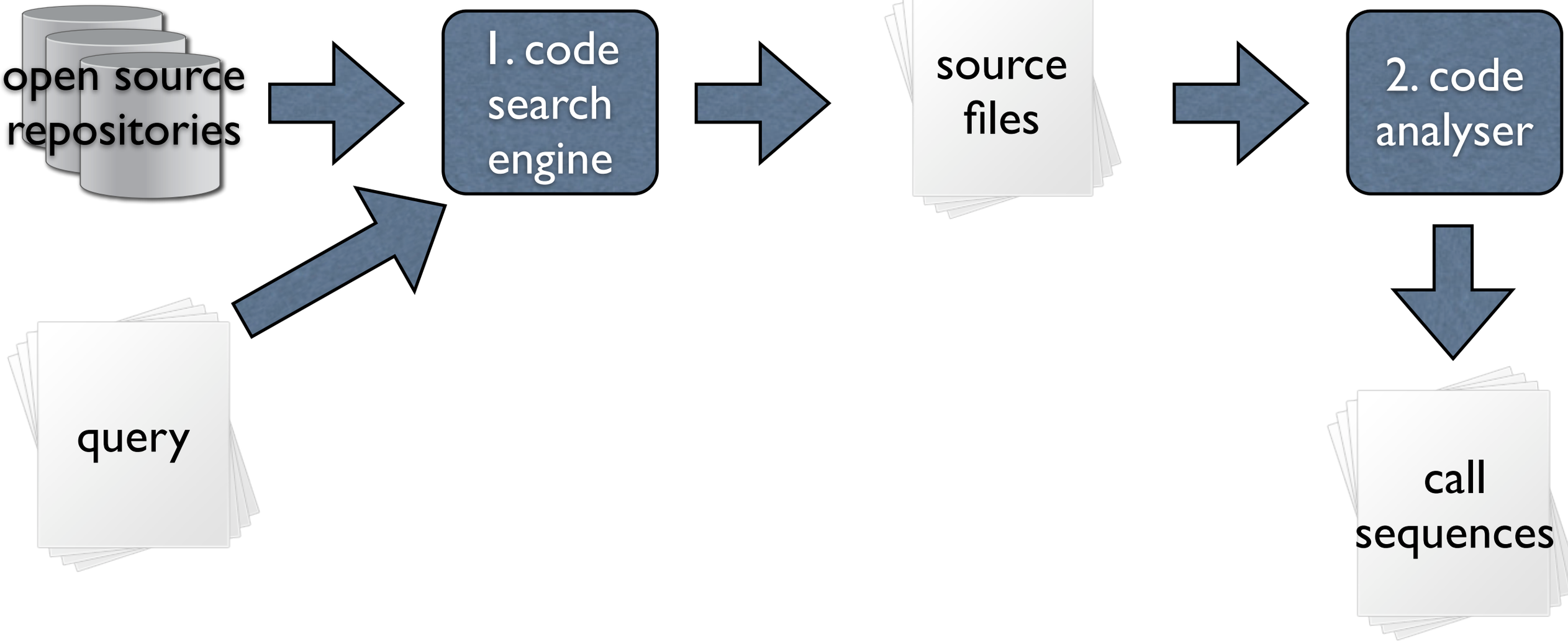
MAPO


open source repositories
1. code search engine
source files
query

MAPO

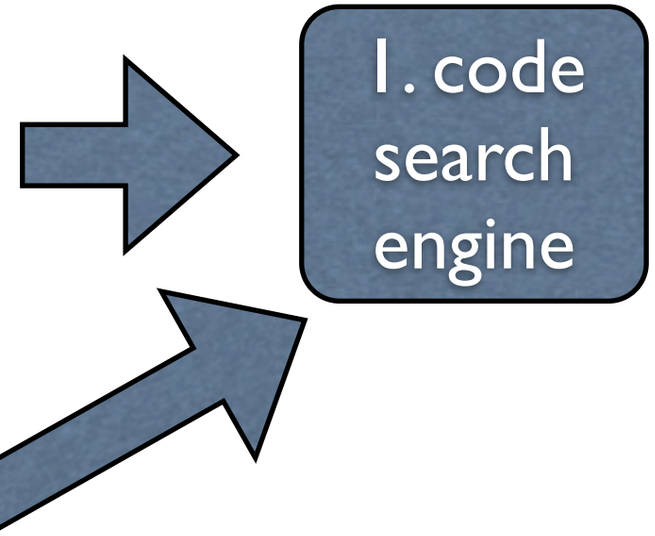

open source repositories
1. code search engine
source files
2. code analyser
query

MAPO

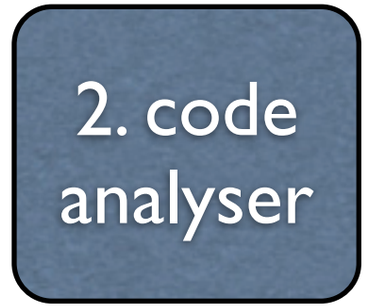

open source repositories
1. code search engine
source files
2. code analyser
query
call sequences

MAPO


open source
repositories

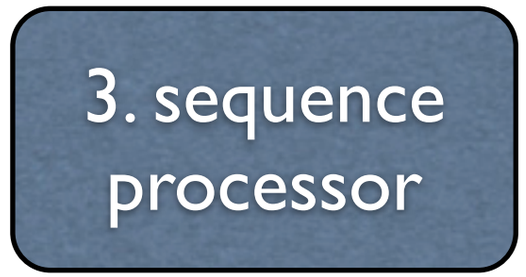

1. code
search
engine


source
files


2. code
analyser


query


call
sequences


3. sequence
processor

MAPO

1. code search engine

source files

2. code analyser

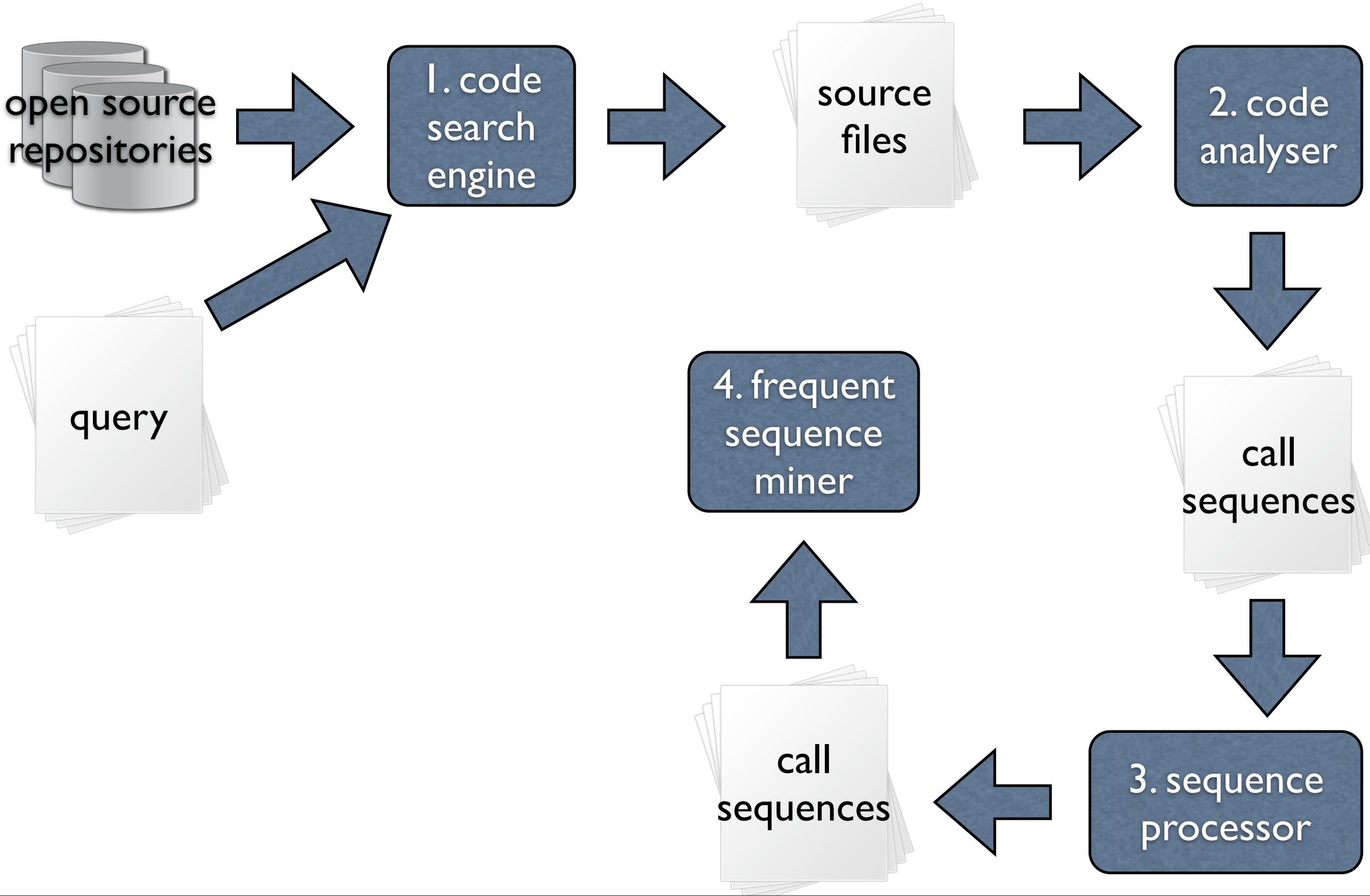
query

call sequences

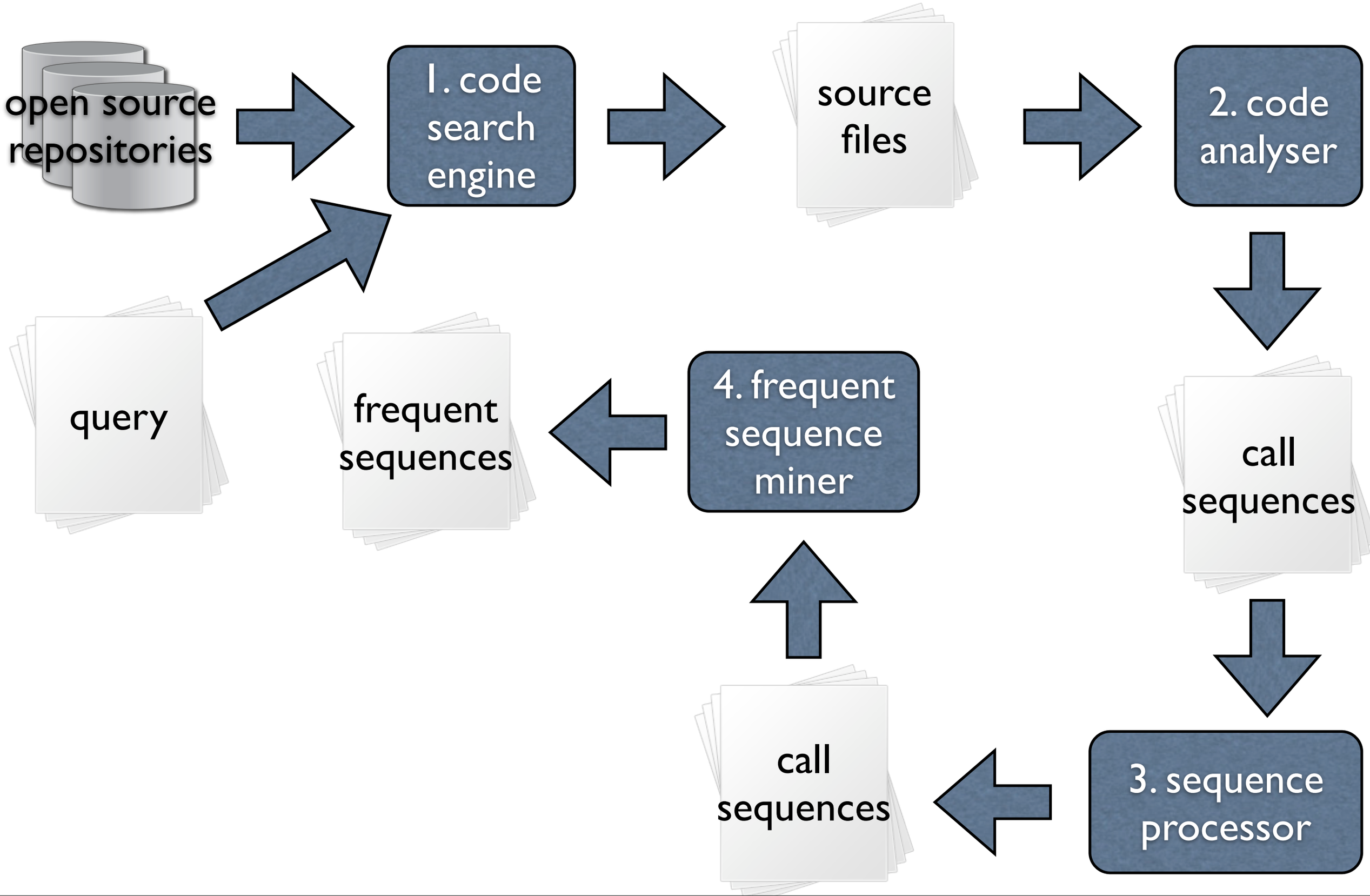
3. sequence processor

call sequences

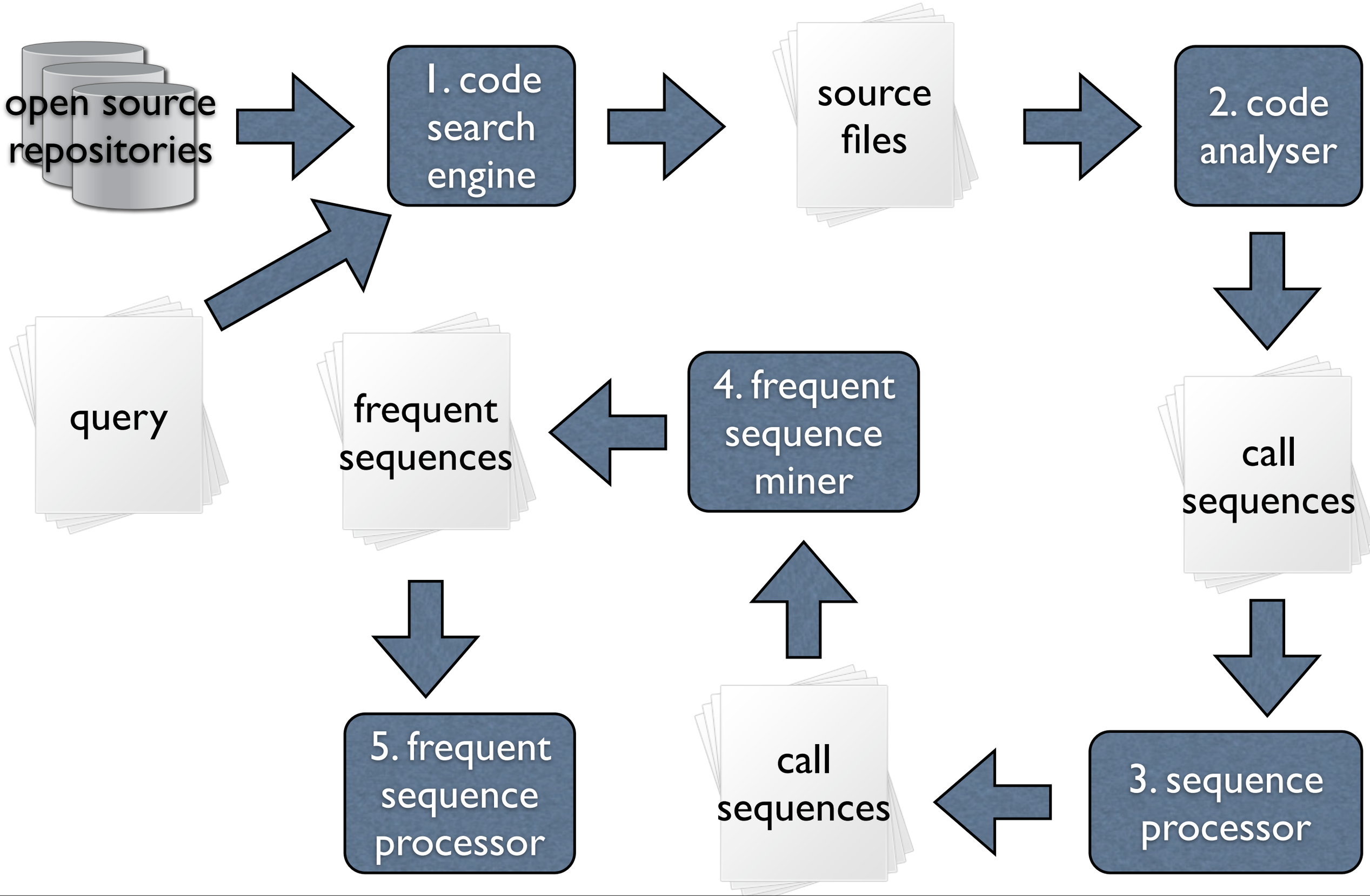
MAPO


open source repositories
1. code search engine
source files
2. code analyser
query
4. frequent sequence miner
call sequences
call sequences
3. sequence processor

MAPO


open source repositories
1. code search engine
source files
2. code analyser
query
frequent sequences
4. frequent sequence miner
call sequences
3. sequence processor
call sequences

MAPO


open source repositories
1. code search engine
source files
2. code analyser
query
frequent sequences
4. frequent sequence miner
call sequences
5. frequent sequence processor
call sequences
3. sequence processor

MAPO

1. code search engine

source files

2. code analyser

call sequences

3. sequence processor

call sequences

4. frequent sequence miner

frequent sequences

5. frequent sequence processor

API Usages

query