Strategies for Confirming we are in the "right place" in the Source Code

```
☑ BufferListener.java ☑ JEditBuffer.java ☒
2776
2777
          //{{{ getListener() method
          private BufferListener getListener(int index)
2778⊜
2779
2780
              return bufferListeners.get(index).listener;
2781
          } //}}}
2782
2783
          //{{{ contentInserted() method
2784
          private void contentInserted(int offset, int length,
2785
              IntegerArray endOffsets)
2786
 2787
2788
2789
                  transaction = true;
2790
2791
                  int startLine = lineMgr.getLineOfOffset(offset);
2792
                  int numLines = endOffsets.getSize();
2793
2794
                  if (!loading)
2795
 2796
                      firePreContentInserted(startLine, offset, numLines, length);
 2797
 2798
 2799
                  lineMgr.contentInserted(startLine,offset,numLines,length,
 2800
                      endOffsets);
2801
                   positionMgr.contentInserted(offset,length);
2802
2803
                   setDirty(true);
2804
2805
                  if(!loading)
2806
2807
                      fireContentInserted(startLine,offset,numLines,length);
2808
2809
                      if(!undoInProgress && !insideCompoundEdit())
2810
                           fireTransactionComplete();
2811
2812
2813
              finally
2814
 2815
2816
                  transaction = false:
2817
2818
          } //}}}
2819
2820
          //{{{ parseBufferLocalProperties() method
          nrivate unid narcaRuffarlocalDronartias(CharSamuence nron)
```

Strategies for Confirming we are in the "right place" in the Source Code

```
↓ JEditBuffer.java 

□
2776
2777
         //{{{ getListener() method
         private BufferListener getListener(int index)
27789
2779
2780
              return bufferListeners.get(index).listener;
2781
         } //}}}
2782
2783
         //{{{ contentInserted() method
2784
         private void contentInserted(int offset, int length,
2785
             IntegerArray endOffsets)
2786
2787
2788
2789
                  transaction = true;
2790
2791
                  int startLine = lineMgr.getLineOfOffset(offset);
2792
                  int numLines = endOffsets.getSize();
2793
2794
                  if (!loading)
2795
2796
                      firePreContentInserted(startLine, offset, numLines, length);
2798
2799
                  lineMgr.contentInserted(startLine,offset,numLines,length,
2800
                      endOffsets);
2801
                  positionMgr.contentInserted(offset,length);
2802
2803
                  setDirty(true);
2804
2805
                  if(!loading)
2806
2807
                      fireContentInserted(startLine,offset,numLines,length);
2809
                      if(!undoInProgress && !insideCompoundEdit())
2810
                          fireTransactionComplete();
2811
2812
2813
              finally
2814
2815
2816
                  transaction = false:
2817
2818
         } //}}}
2819
         //{{{ parseBufferLocalProperties() method
         nrivate unid narcaRuffarlocalDronartias(CharSaguanca nron)
```

- Breakpoint
- System.out.println("message");
- Change observable behavior

Example Concept Location Explanation for Assignment

■ 'I started by locating the loading screen image in the resources folder. Once I identified the loading screen image as "splash.png", I searched exactly for that string in the entire workspace and found the file "SplashScreen.java" [...]'

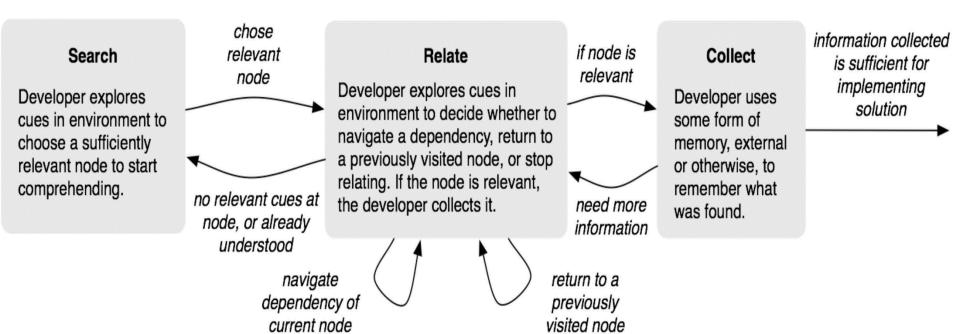
Program Understanding Model

"Developers work on software maintenance tasks by

- searching for relevant code,
- navigating dependencies, and
- constructing a mental model of the dependencies of a feature, suggesting the need for environments that explicitly support feature location through dependency navigation."

[Ko'06] Ko, A.J., Myers, B.A., Coblenz, M.J., and Aung, H.H. "An Exploratory Study of How Developers Seek, Relate, and Collect Relevant Information during Software Maintenance Tasks", 2006, IEEE Transactions on Software Engineering, 32(12), 971-987.

Program Understanding Model



"Developers work on software maintenance tasks by

- searching for relevant code,
- navigating dependencies, and
- constructing a mental model of the dependencies of a feature, suggesting the need for environments that explicitly support feature location through dependency navigation."

[Ko'06] Ko, A.J., Myers, B.A., Coblenz, M.J., and Aung, H.H. "An Exploratory Study of How Developers Seek, Relate, and Collect Relevant Information during Software Maintenance Tasks", 2006, IEEE Transactions on Software Engineering, 32(12), 971-987.

Grep-based Concept Location

Grep-based Concept Location

 Source code is not processed (treated as "text") and search mechanism is regular expression matching

- •Queries are regular expressions (i.e., formal language):
 - •[hc]at, .at, *at, [^b]at, ^[hc]at, [hc]at\$, etc.
- Results are unordered lines of text where the query is matched
- Works great for developers who know exactly what to search for

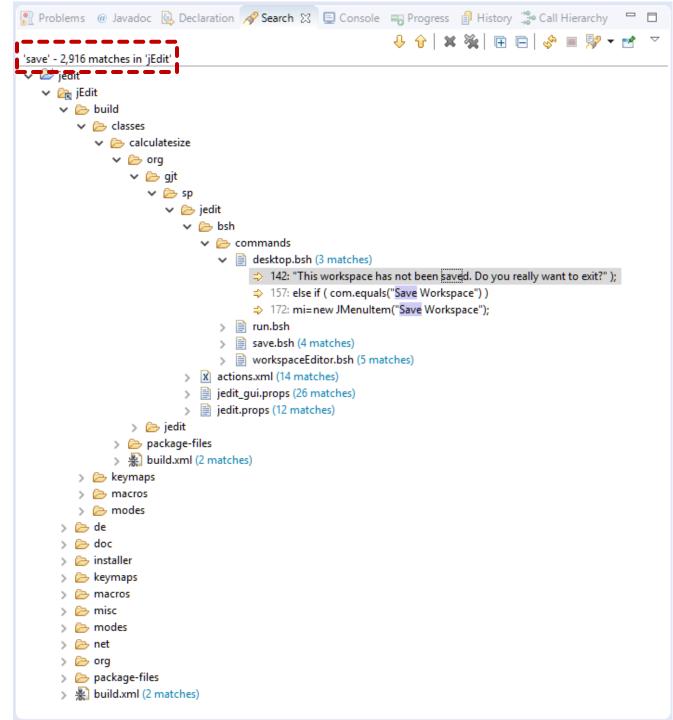
Grep-based Concept Location Example

query = "save"

	_		×
File Search Task Search Git Search Java Search	<i>≫</i> P	lug-in Sea	rch
Containing text:			
save	Case	e sensitive	
(* = any string, ? = any character, \ = escape for literals: * ? \)		ular e <u>x</u> pres o <u>l</u> e word	sion
File name <u>p</u> atterns (separated by comma):			
*	Chg	<u>o</u> ose	
(* = any string, ? = any character, !x = excluding x) Search In Derived resources Binary files			
Scope Oworkspace Selected resources Enclosing projects Working set:		C <u>h</u> oose	
? Customize <u>Replace Search</u>		Canc	el

Grep-based Concept Location Example

~3,000 results



Limitations of Grep-based search

 Query formulation (regular expression) is more difficult than natural language (e.g., google searches)

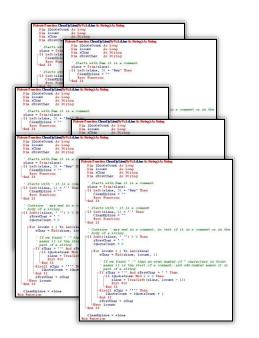
Results unranked

- Exact matches only
 - •i.e., cannot find synonyms, or conceptually similar topics

How Can We Do Better?







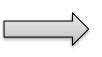


Vector Space Model
Latent Semantic Indexing
Latent Dirichlet Allocation
Relational Topic Model

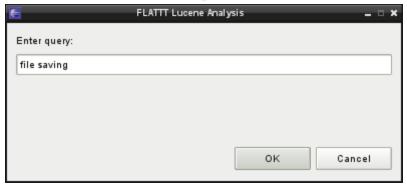
Apache Lucene*



```
The section of the comment of the co
```

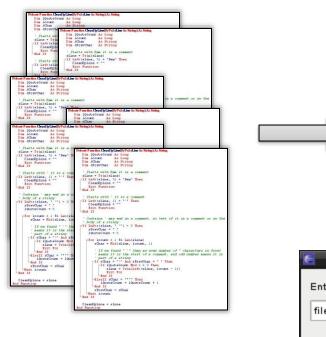






http://www.cs.wm.edu/semeru/flat3/



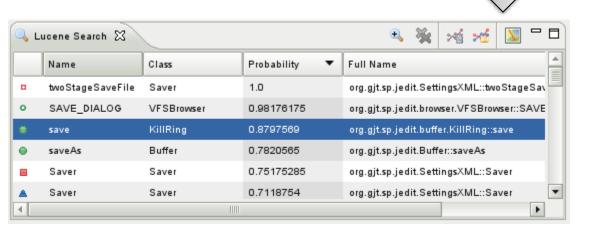


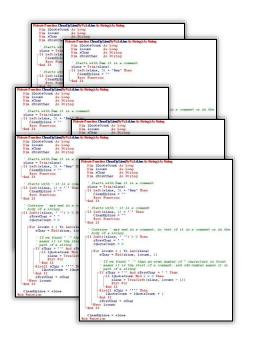




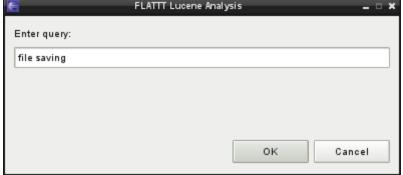
http://www.cs.wm.edu/semeru/flat3/

FLAT³





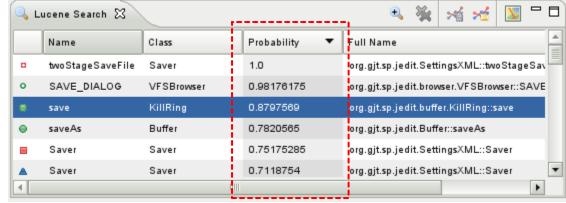




http://www.cs.wm.edu/semeru/flat3/

FLAT³





Results are ranked based on their textual similarity to the query

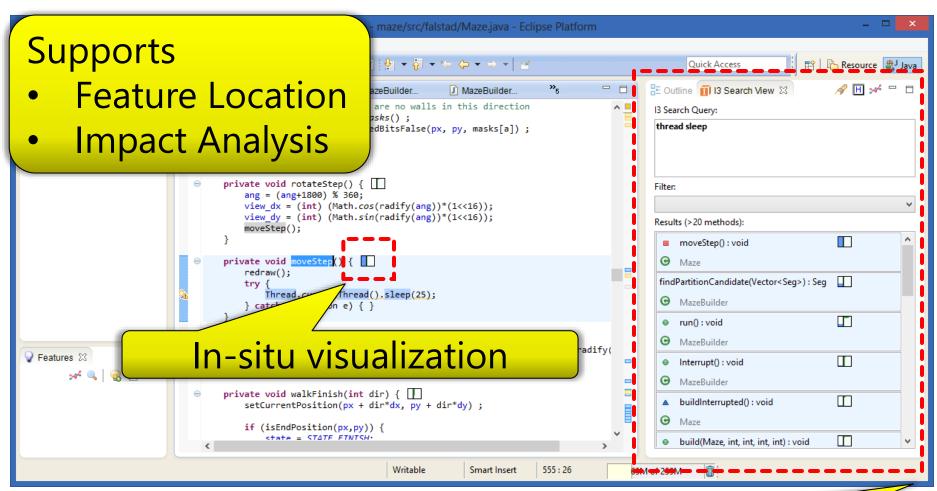
FLAT^{3:} Feature Location and Textual Tracing Tool

- http://www.cs.wm.edu/semeru/flat3/
 - Contains 5 minute demo video
- Eclipse plug-in eclipse (can be easily extended)
- Static feature location (uses Information Retrieval)

Dynamic feature location (uses execution traces)

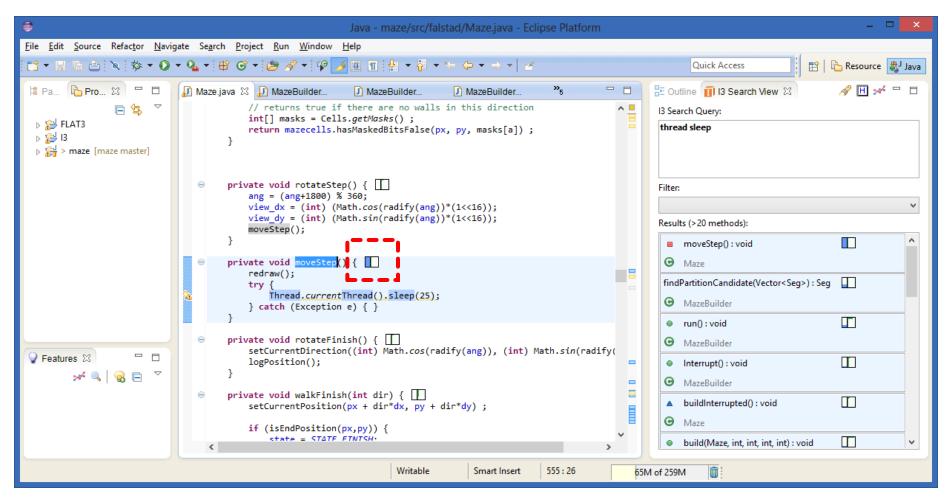
Feature annotation

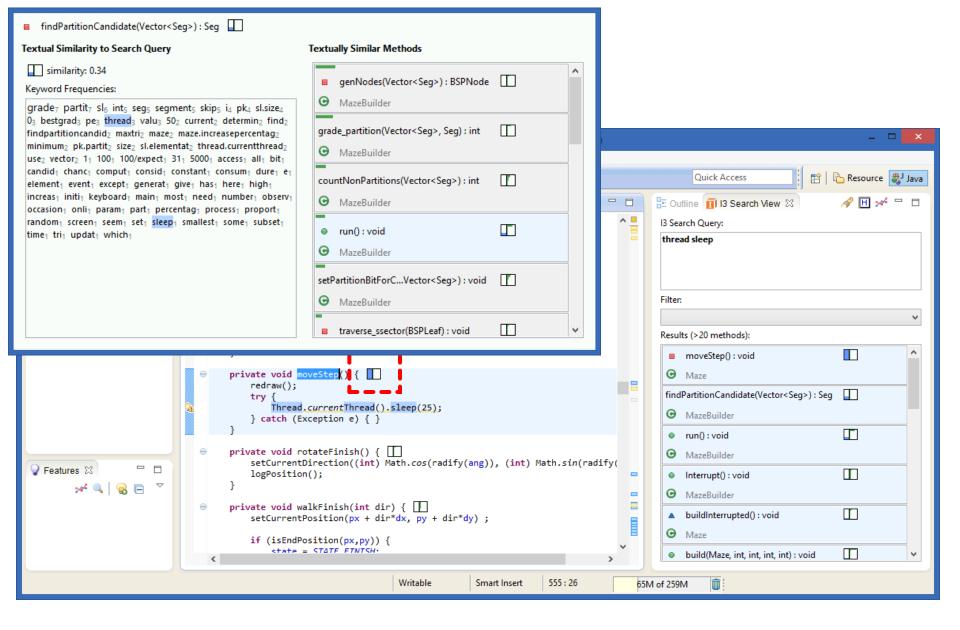
13

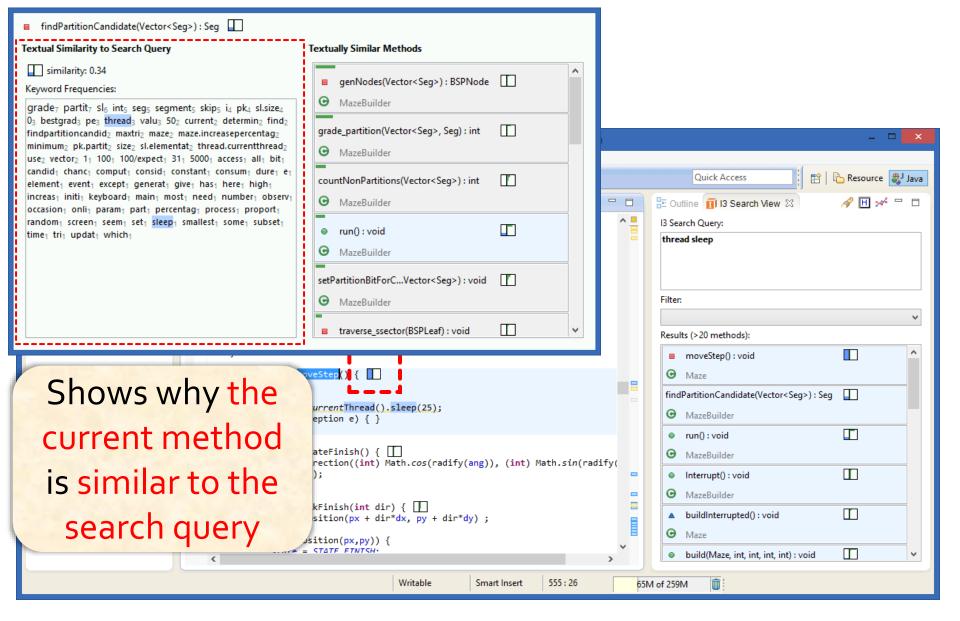


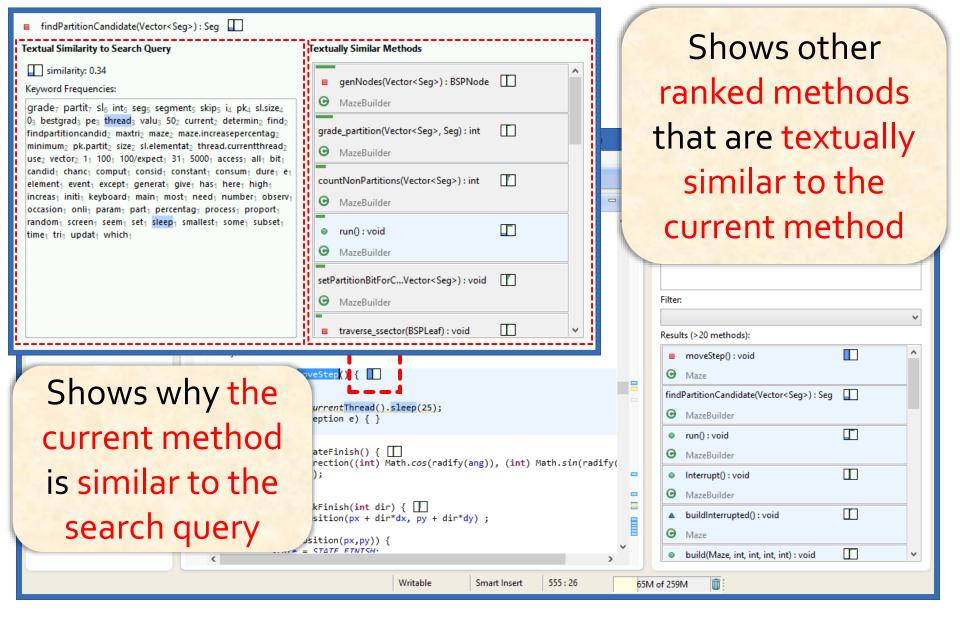
Focus on user interface

13

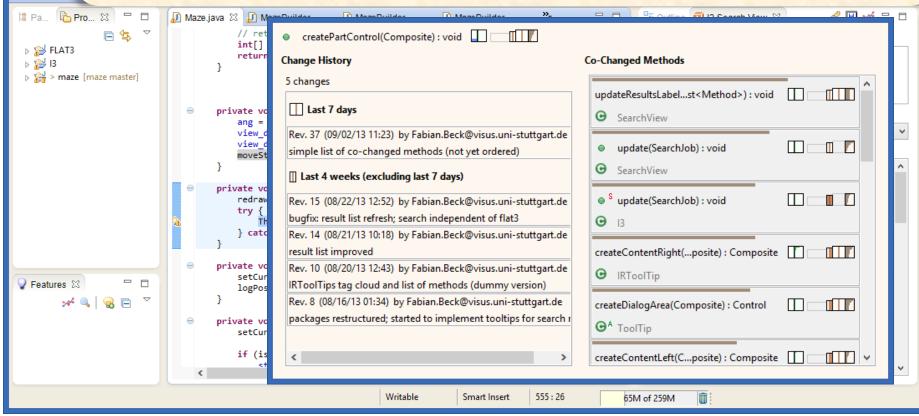




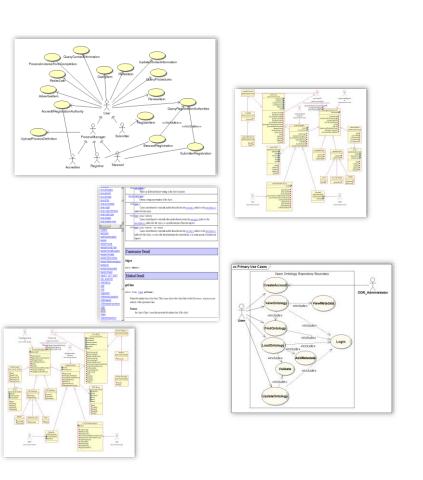




Analyzes historical information and shows a ranked list of methods that were changed frequently with the current method across multiple commits



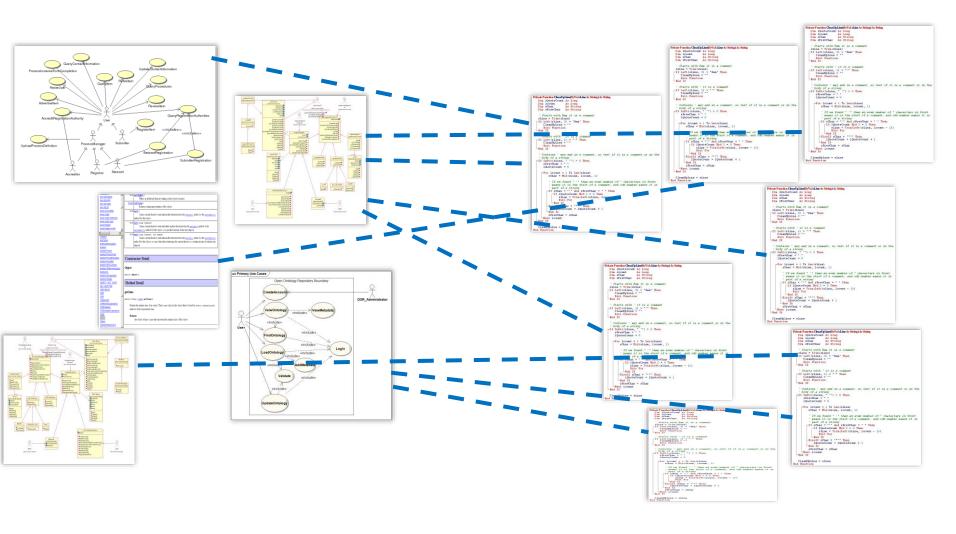
File Edit Sc





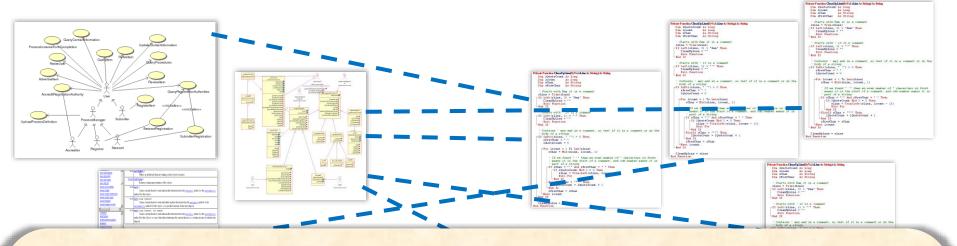
Design Documents

Source Code



Design Documents

Source Code



The Problem Traceability Link Recovery is trying to solve:

What are the documents/artifacts associated with a given source code component?

Traceability can be Between

- Requirements and code
- Design and code
- Requirements and design
- Requirements and test cases
- Design and test cases
- Bug report and maintainer
- Manual page to code

-....

Traceability Link Recovery "Definition"*

- The ability to describe and follow the life of a requirement, in both a forward and backward direction:
 - •from its origins, through its development and specification, to its subsequent deployment and use, and through periods of ongoing refinement and iteration in any of these phases.

Why Traceability?

Why Traceability?

- It is required or suggested by many standards:
 - MIL-STD-498, IEEE/EIA 12207 (Military)
 - DO178B, DO254 (Avionic), EN50128 (Railways)
 - Medical, financial, etc.
- Contractual agreements to ensure that:
 - all required functionalities are there
 - there is no EXTRA functionality
- Facilitates:
 - bottom-up and top-down program comprehension
 - Impact analysis
 - Identification of reusable software components
- Easy to implement with modern tools (e.g., GitHub)

The basic Assumption

- Developers use consistent naming conventions to
 - create identifiers
 - write comments
 - name artifacts
 - write manual pages or e-mails
- Developers use domain concepts and knowledge in a uniform and consistent way

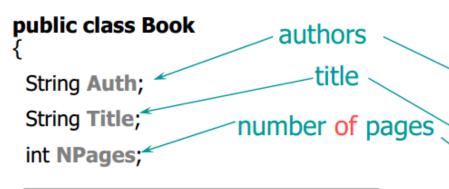
Example – A Requirement

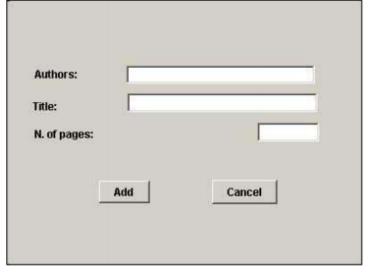
```
It shall be possible to insert a new book into the library. The user shall insert the following information:
```

- The **title** of the book;
- The names of the authors;
- The number of pages of the book.

```
After confirming the data inserted, data are stored into the database. In case of abort, nothing shall happen.
```

Example of Mapping





public class AddBook extends JFrame {

```
JPanel contentPane;
GridBagLayout gridBagLayout1 = new
GridBagLayout();
TextField Auth = new JTextField();
JLabel iLabel1 = new JLabel();
JTextField T = new JTextField();
JLabel jLabel2 = new JLabel();
JTextField Npages = new JTextField();
JLabel jLabel3 = new JLabel();
JButton Cancel = new JButton();
JButton Add = new JButton();
Border border1;
```

....

Challenges for Traceability Link Recovery Techniques

- High level documents mostly in natural language
 - source code
 - acronyms
 - abbreviations
 - etc.

Automatic-generated code

Challenges for Traceability Link Recovery Techniques

- Conceptual distance between different artifacts
 - High level requirement vs. code or test cases
- Vocabulary inconsistency
 - •fault, defect, bug, issue, ...
- Text sparseness
 - there is no better data than more data

How will Traceability be Implemented in CS471/CS481?

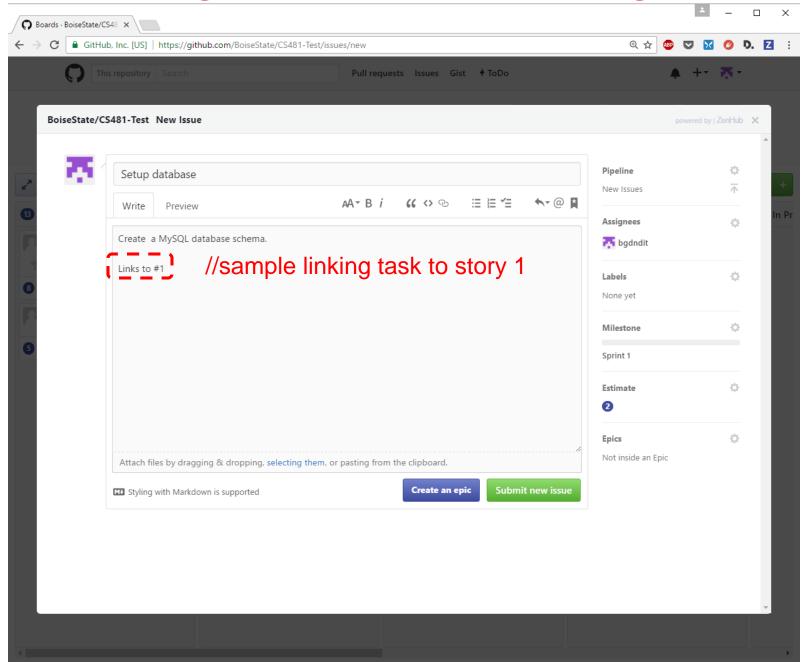
•Always link a task to a story (e.g., #123), where 123 is the story ID

- •Always link a task to a story (e.g., #123), where 123 is the story ID
 - but my task doesn't have a story to link...

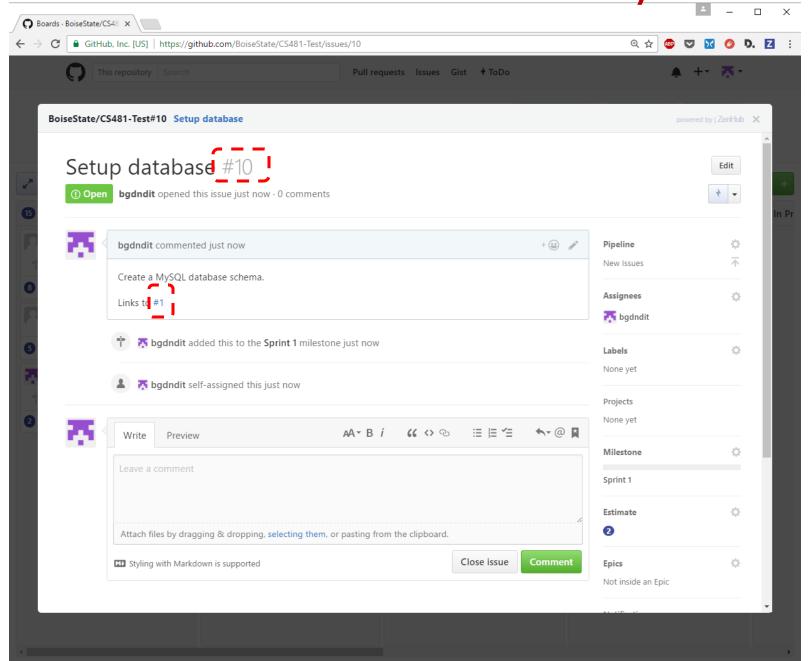
- •Always link a task to a story (e.g., #123), where 123 is the story ID
 - but my task doesn't have a story to link...
 - Did you forget to create the story?

- •Always link a task to a story (e.g., #123), where 123 is the story ID
 - but my task doesn't have a story to link...
 - Did you forget to create the story?
 - working on an "independent" task (i.e., not linked to a story) is a good sign you're going off on a tangent...so stop it and reassess

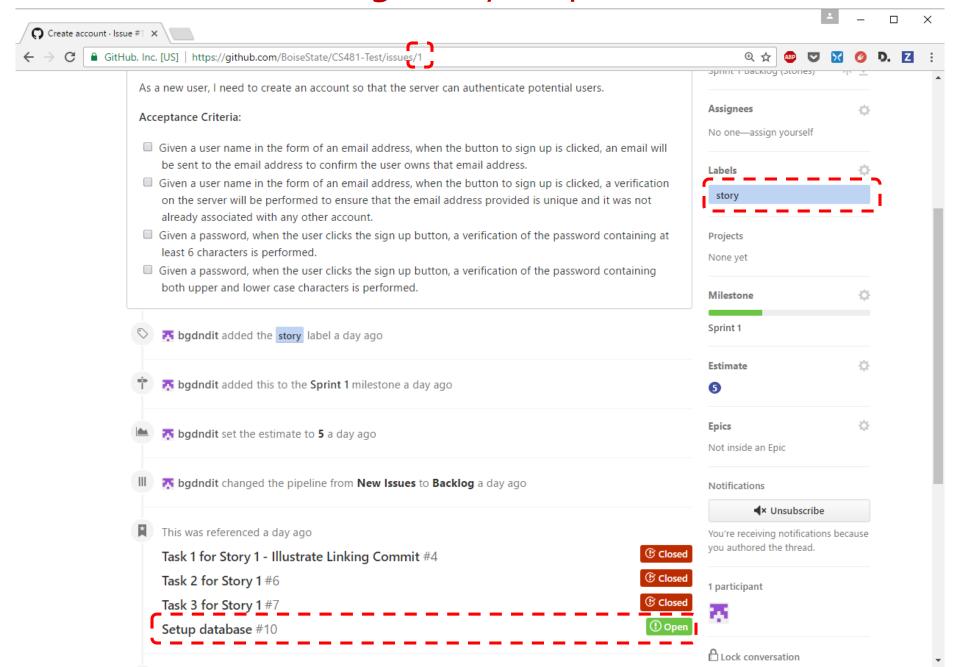
Referencing a new task to existing story #1



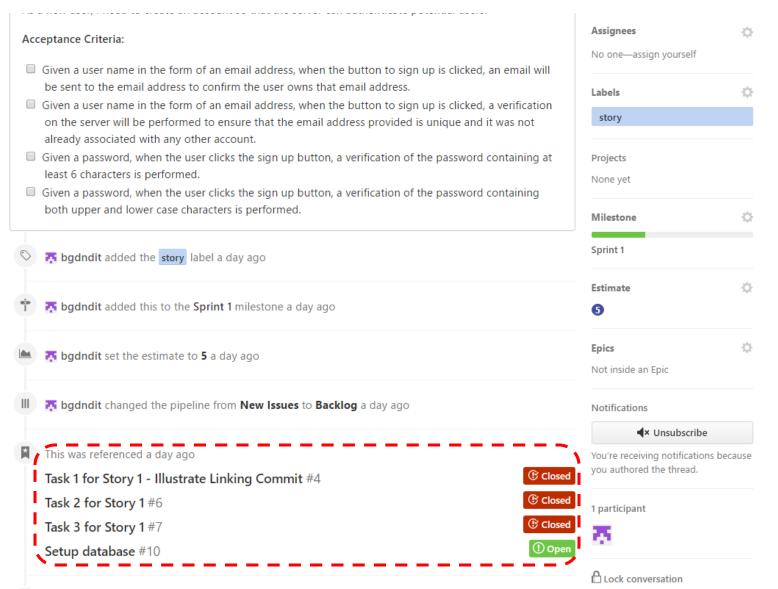
Task #10 linked to Story #1



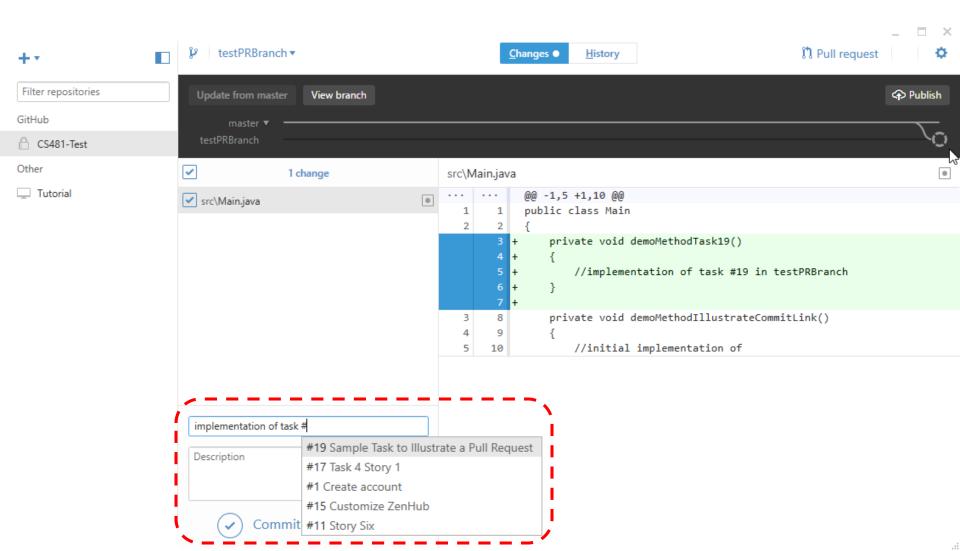
Benefits of Linking (story #1 points to task #10)



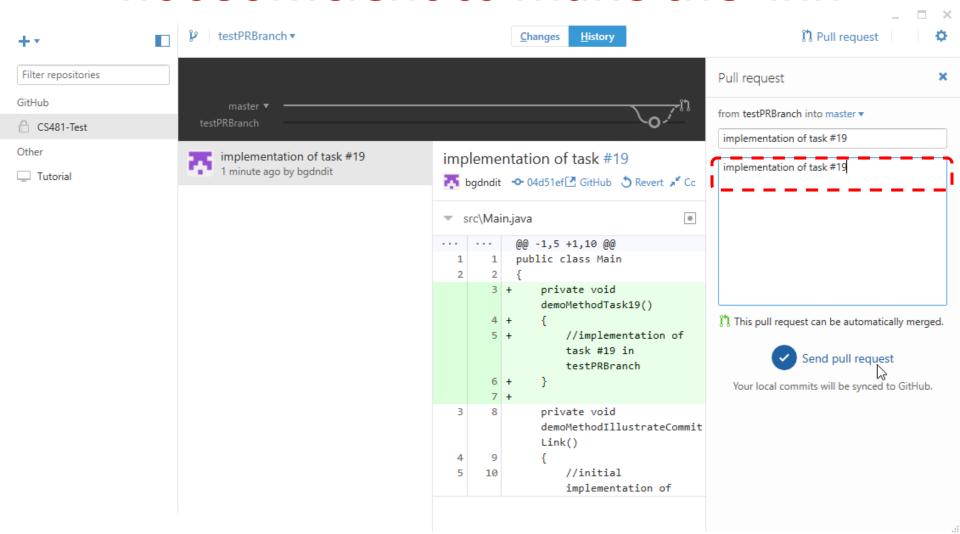
A story is split into multiple tasks (each with its owner): It's easy to track progress when using linking between software artifacts



In the commit log message specify the task/bug ID to associate the implementation with the ID



Description of PR should include a reference to the task ID. Title only is not sufficient to make the link



How will Traceability be Implemented in CS471/CS481?

Link tasks to stories

Link commit log messages to tasks/bugs

Link Pull-Requests to tasks