

# Software Comprehension and Reuse supported by Analytics and Repository Mining

**Marcelo de Almeida Maia**

Computer Science Department

Federal University of Uberlândia

*marcmaia@facom.ufu.br*



**Laboratory of**  
*Software Comprehension, Analytics and Mining*

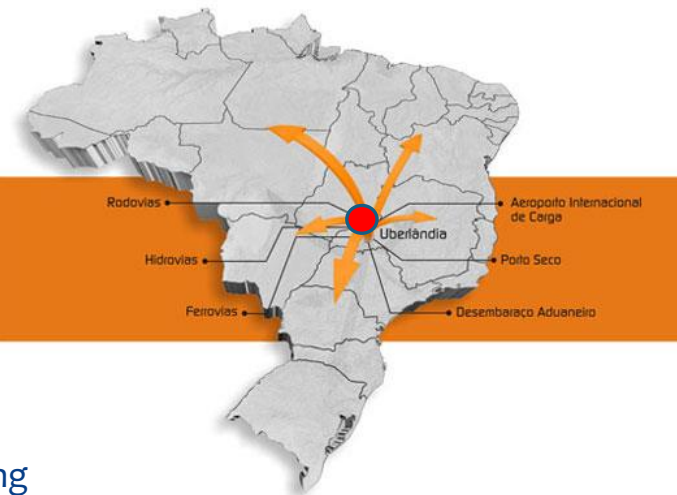


SEIF Brazil Workshop 2013



# A little bit about Uberlândia, UFU, LASCAM

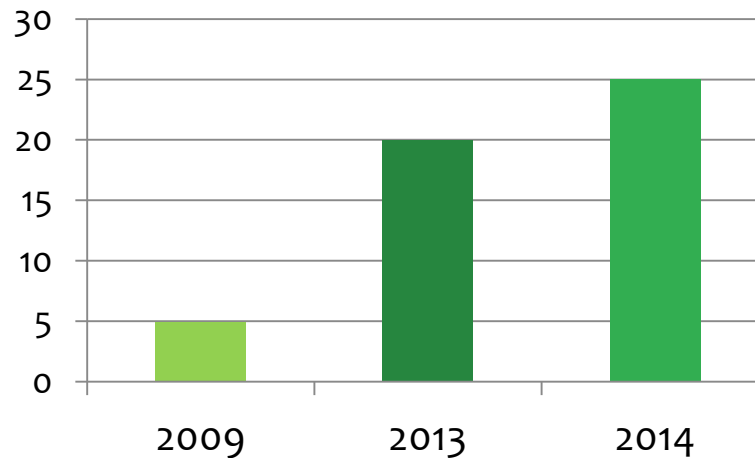
- Uberlândia: Strategic location
  - Population: 700,000
  - Reference in Logistics
- UFU
  - Relatively young institution
- Graduate Program in Computer Science
  - 2001 MSc    2011 PhD
- This talk is around giving shape to
  - **L**aboratory of **S**oftware **C**omprehension, **A**nalytics and **M**ining
    - 3 PhD students
    - 4 MSc students
    - 2 undergraduate students
- Agenda
  - Motivation
  - Some previous/current work
  - Some related work
  - Our vision



# Open source: a treasure or not?

- How much does open source code worth?
- In 2009, in a Computerworld column:

**LOC (billions)**



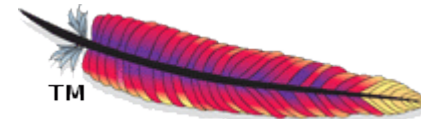
- In 2009, OSS estimated in ~US\$ 400 bi



# OSS and Software Development

## Few examples

- Infrastructure
  - DBMS: Postgres
  - Application Servers: Jboss (Wildfly)
  - Cloud Platforms: CloudStack
  - Distributed Computing: Hadoop
- Libraries and Frameworks
  - Search engines: Lucene
  - Frameworks: Spring
- IDEs and SE Tools
  - Eclipse, Netbeans
  - SonarQube



# OSS and Software Development

## Few examples

### ■ Infrastructure

- DBMS: Postgres
- Application Servers: JBoss (Wildfly)
- Cloud Platforms: CloudStack
- Distributed Computing: Hadoop

### ■ Libraries and Frameworks

- Search Engines: Lucene
- Frameworks: Spring

### ■ IDEs and SE Tools

- Eclipse, Netbeans
- SonarQube

GREAT!

But, this is not the treasure\*  
we're seeking!  
At least not as a box!



\*If ever exists! 😊

# What is the “treasure”?

- “Raw” LOC !!??  
White-box Reuse !!??  
Fine-grained code elements !!??
- Treasure or Trash? Eye of the beholder...



# What is the “treasure”?

- Raw LOC !!!  
White-box Reuse !!!  
Fine-grained code elem

- Treasure or Trash?



60% of structure  
90% of furniture } = Trash





# Is it worth?

## ■ Gain

- Internal algorithms
- Internal concepts
- Templates
- Partial structure
- Architecture

## ■ Pain

- Coupling impact
- Lack of high-level doc
- Context understanding
- Quality assessment
- Adaptation/Restructuring...



# What is the “treasure”?

- Code or People?

Discover, Track and Compare Open Source

Search Projects...

Connecting 3,141,952 open source contributors

Click to see  
example project

Number of StackOverflow Users

```
select count(Id)  
from Users
```

Run Query

Cancel

Options: ☐ Text-only results

Switch sites:

☐ search by name or url

Results

Messages

2546106



How can we reuse the crowd knowledge?

Connecting 3,141,952 open source contributors

Results

Messages

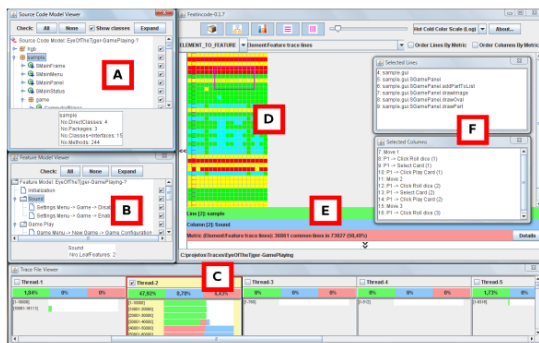
2546106

# Some of our past work

## Execution traces

- Using execution traces for comprehension and maintenance
  - Difficulties for large scale analysis of repositories*

Tool  
(WCRE2008)



Alignment  
(CSMR2011)

Definition of an  
implementation plan

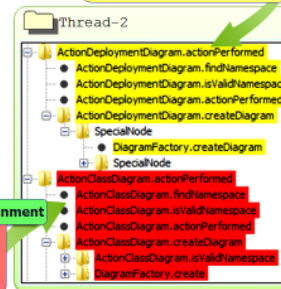
**Selected Classes:**  
ActionDeploymentDiagram  
ActionClassDiagram

**Selected Methods:**  
actionPerformed  
findNamespace  
isValidNamespace  
createDiagram

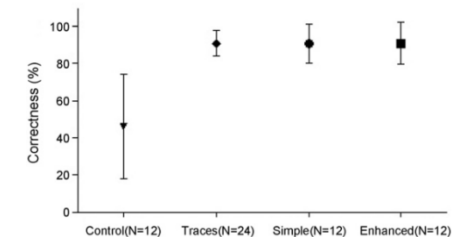
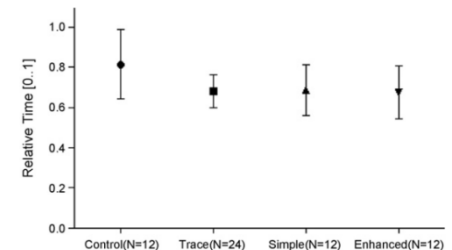
Scenario 2  
1. Initialize system  
2. Create class diagram

Scenario 1  
1. Initialize system  
2. Create deployment diagram

Misalignment



Performance  
assessment  
(JSS2013)

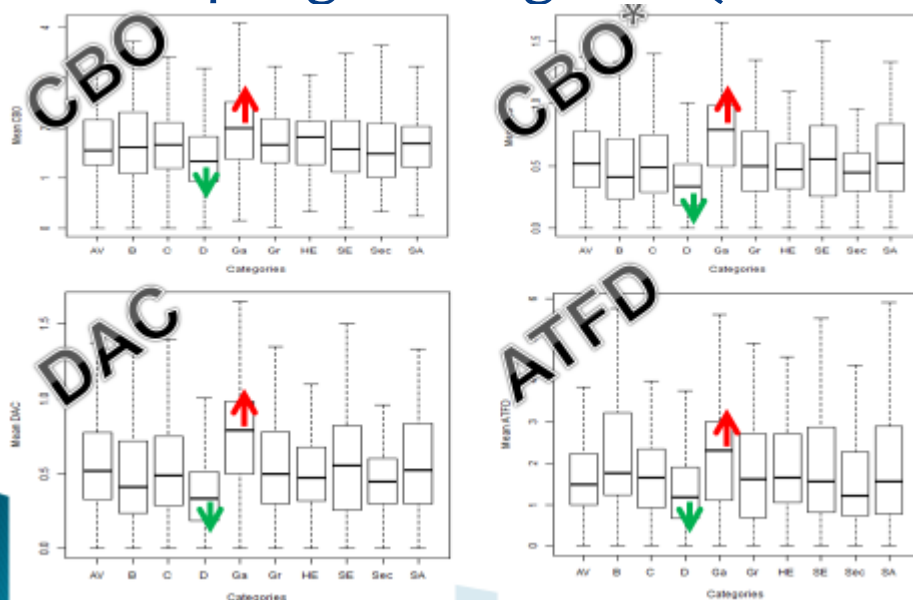


# Some of our past work

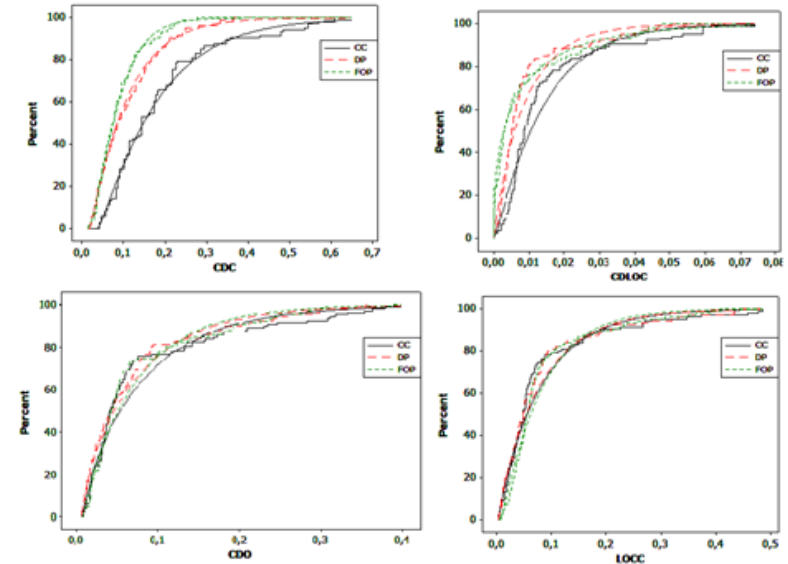
## Metrics

- Characterization of domains, PL constructions evaluation
  - *Difficulties for cause-effect understanding, measurement extraction for features.*

### Coupling x Categories (MSR2013)



### FOP x SPLs (SCP2014)



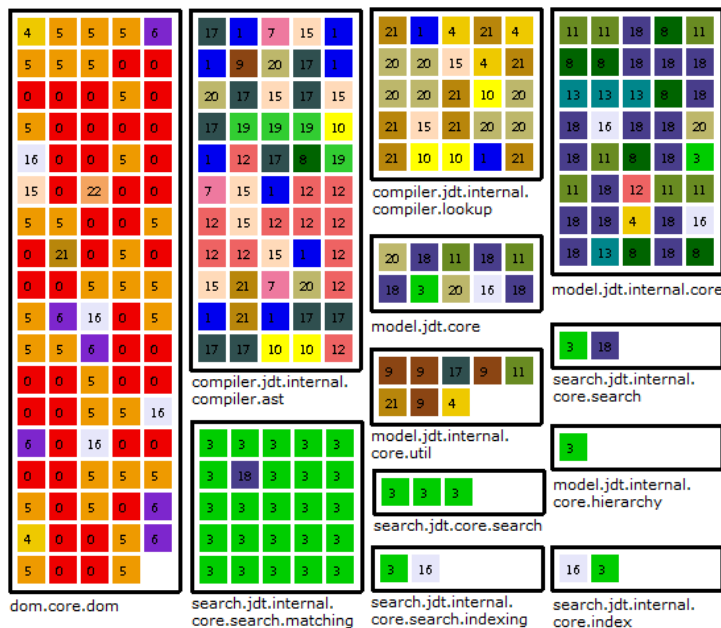
# Some of our current work

## Version history and Traces

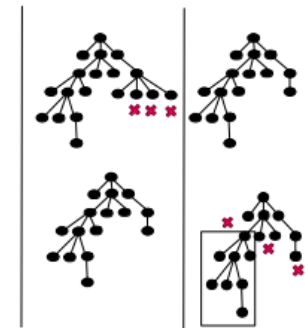
### ■ Modularity Assessment and Comprehension

- Difficulties for cause-effect understanding, ground-truth evaluation
- Necessity to have large history of changes and high coverage traces

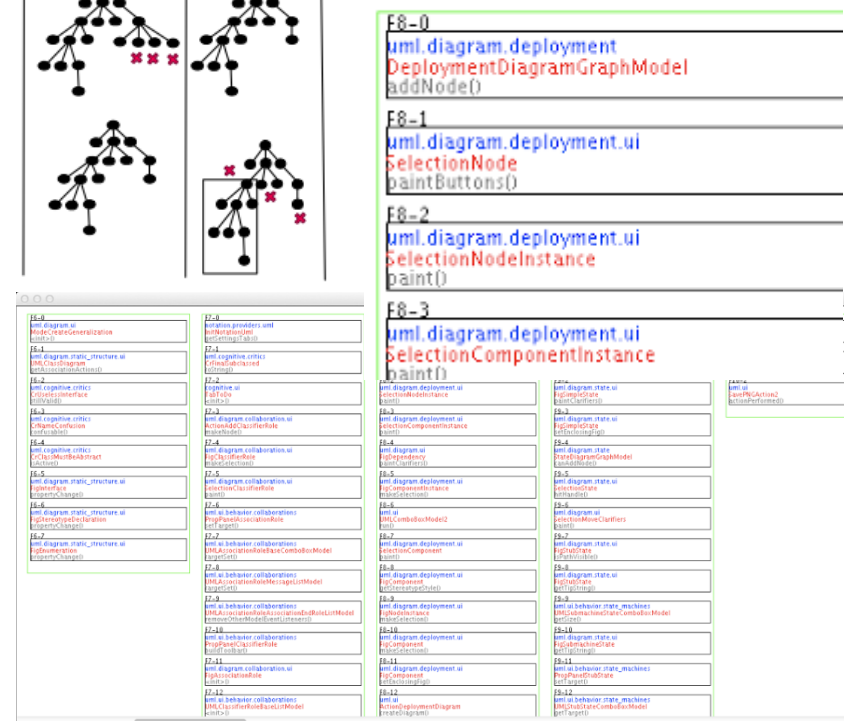
### Co-change clusters (submitted)



### YASAR (WIP)



### Deployment Diagram

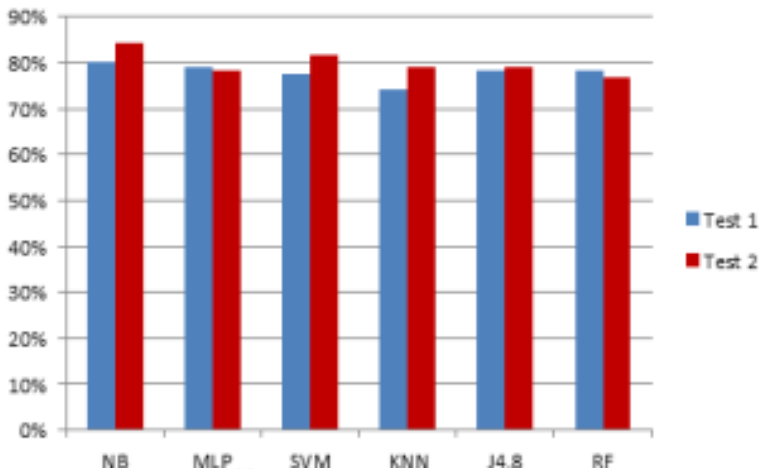


# Some of our current work

## Q&A mining

- Structured Information Extraction for Recommendation Systems using machine learning and IR (WIP)
  - Subjective nature of unstructure data

Post classification  
based on intent



Cookbook extraction 



ido ao Livro de Receitas !

19 : user list menu page item web select displai grap  
[convert a JTS-Geometry into an AWT-Shape?](#)

17 : paint draw method graphic compon call object :  
[a Java Component to an image file](#)  
[onent not called](#)  
[onent\(\) vs paint\(\) and JPanel vs Canvas in a paintbrush-type G](#)  
[image from a non-visible AWT Component?](#)  
[image from a non-visible AWT Component?](#)

18 : java run applic program applet file system instal  
[g and drop files from a directory in java](#)  
[low Close Listener/Event](#)  
[display?](#)  
[display?](#)

16 : mous kei click event press want listen user movi  
[cursor automatically move from one TextField to other](#)  
[tab key event](#)  
[take the icon alive in my AWT button? On click the icon goes](#)  
[add "Default Button" functionality to two buttons in Swing?](#)  
[mer on JLabel or JTable cell](#)  
[\(TextField limit only to 0 or 1](#)  
[sword field in AWT/SWT?](#)  
[quest focus synchronously in Swing?](#)

how to fire tab key event

Tags: java swing awt

How do we fire a "tab" key pressed event deliberately in java.?  
event programatically in java. Though this a bit lame question,p  
help..

The following example shows how to simulate mouse and key presses in J

```
try {  
    Robot robot = new Robot();  
  
    // Simulate a mouse click  
    robot.mousePress(InputEvent.BUTTON1_MASK);  
    robot.mouseRelease(InputEvent.BUTTON1_MASK);  
  
    // Simulate a key press  
    robot.keyPress(KeyEvent.VK_SHIFT);  
    robot.keyPress(KeyEvent.VK_TAB);  
    robot.keyRelease(KeyEvent.VK_TAB);  
    robot.keyRelease(KeyEvent.VK_SHIFT);  
} catch (AWTException e) {  
    e.printStackTrace();  
}
```

# Some of related state-of-the-art work

## Poshyvanyk and colleagues

- ASE Conf. 2013      Detecting Bad Smells in Source Code  
                                 Using Change History Information
  - *Necessity to have large history of changes*

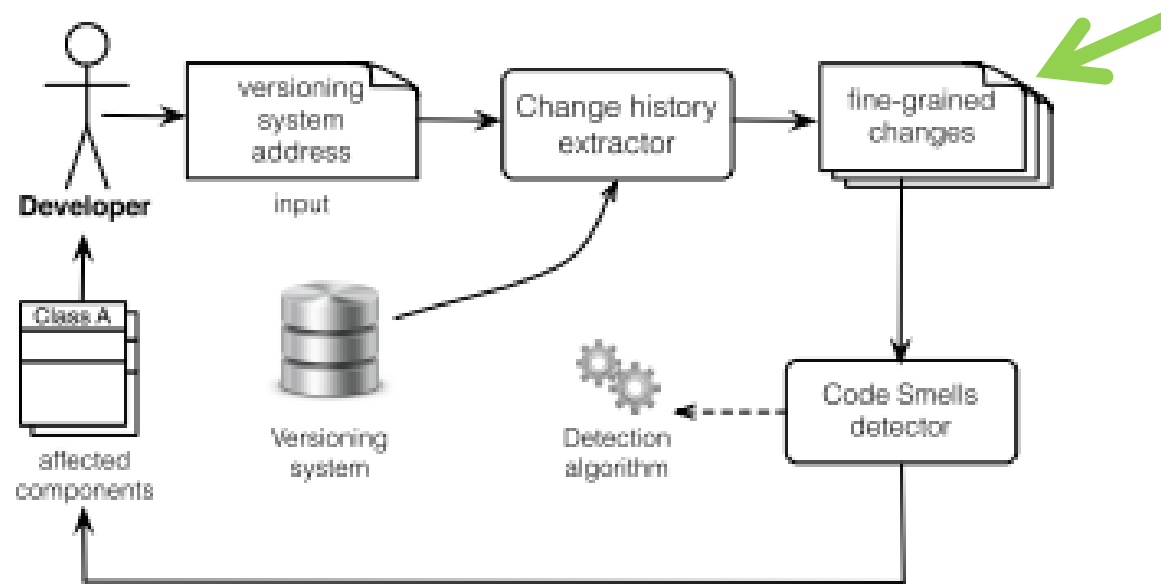


Fig. 1. HIST: The proposed code smell detection process.



# Some of related state-of-the-art work

## Lanza and colleagues

### ■ CSMR2013 - Seahawk

- formulates queries from the current context in the IDE, and presents a ranked and interactive list of results
- Low NDCG: 9.07%

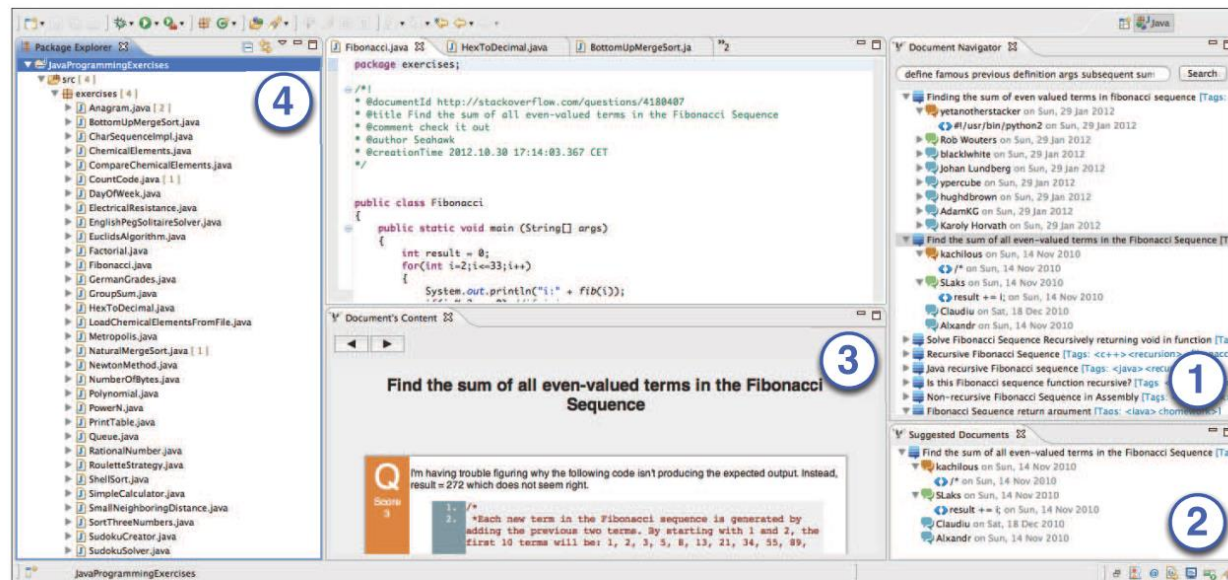


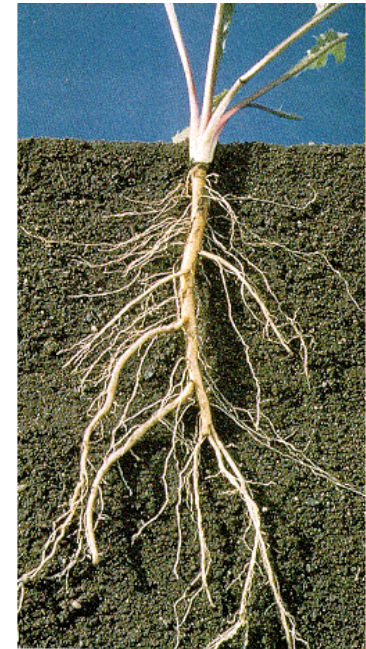
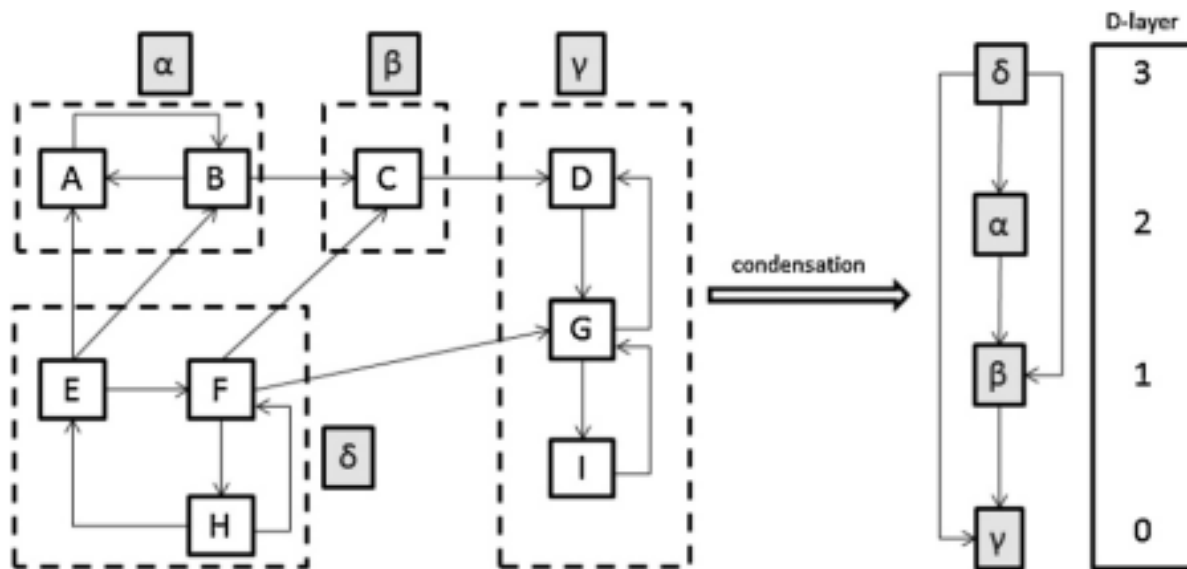
Figure 1: Seahawk User Interface



# Some of related state-of-the-art work

## Kakarontzas and colleagues

- JSS2013 – FWBR Metric (Facilitating White-Box Reuse)
  - Based on D-layer... regression D-layer -> CK function
  - **Lack of external validation**



# Envisioning our goals

## Gluing the pieces together



Code	Changes	Issues	Traces	Posts
------	---------	--------	--------	-------

**Data Warehouse: Extraction, Transform, Load**

**Data Analysis: Statistics, Data Mining, Machine Learning, IR**

**Recommendation Systems Integrated into IDEs**

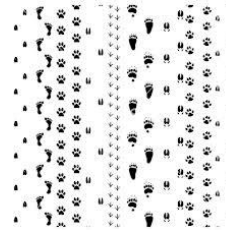
**Reuse  
in Repos**

**Modular  
Rejuvenation**

**Contextual  
How-tos**

# At last, but not least

Acknowledgement to MS Research for SEIF-Brazil



**Code**

**Changes**

**Issues**

**Traces**

**Posts**

**Data Warehouse: Extraction, Transform, Load**

**Data Analysis: Statistics, Data Mining, Machine Learning, IR**



**Visual Studio®**

**Reuse  
from Codeplex**

**Modular  
Rejuvenation**

**Contextual  
How-tos**

# QUESTIONS????

## Software Comprehension and Reuse supported by Analytics and Repository Mining

**Marcelo de Almeida Maia**

Computer Science Department

Federal University of Uberlândia

*marcmaia@facom.ufu.br*



**Laboratory of**  
*Software Comprehension, Analytics and Mining*



SEIF Brazil Workshop 2013

