

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

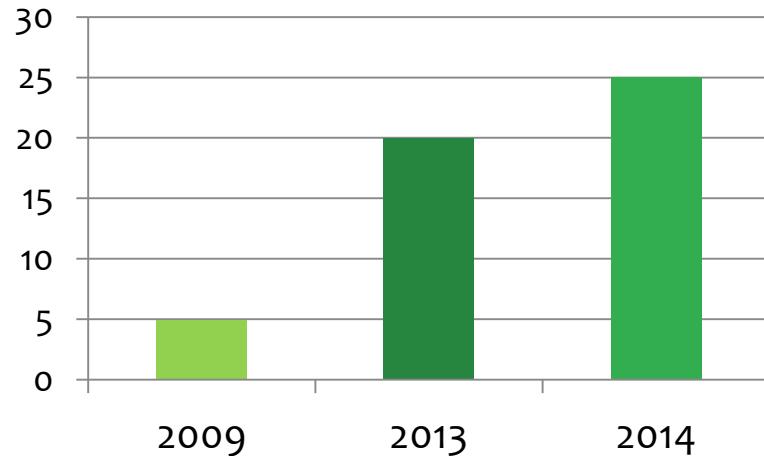
- Uberlândia: Strategic location
 - Population: 700,000
 - Reference in Logistics
- UFU
 - Relatively young institution
- Graduate Program in Computer Science
 - 2001 MSc 2011 PhD
- This talk is around giving shape to
 - **LA**boratory of **S**oftware **C**omprehension, **A**nalytics and **M**ining
 - 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)

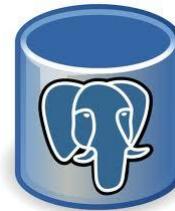


- In 2009, OSS estimated in ~US\$ 400 bi

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: Glassfish (Wildfly)
 - Cloud Platforms: CloudStack
 - Distributed Computing: Hadoop
- Libraries and Frameworks
 - Search Engines: Lucene
 - Frameworks: Spring
- IDEs and SE Tools
 - Eclipse, Netbeans
 - SonarQube

But, this is not the treasure!
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?



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 Projects

Search Projects...



Run Query

Cancel

Options: Text-only results

Connecting 3,141,952 open source contributors

Click to see
example pro

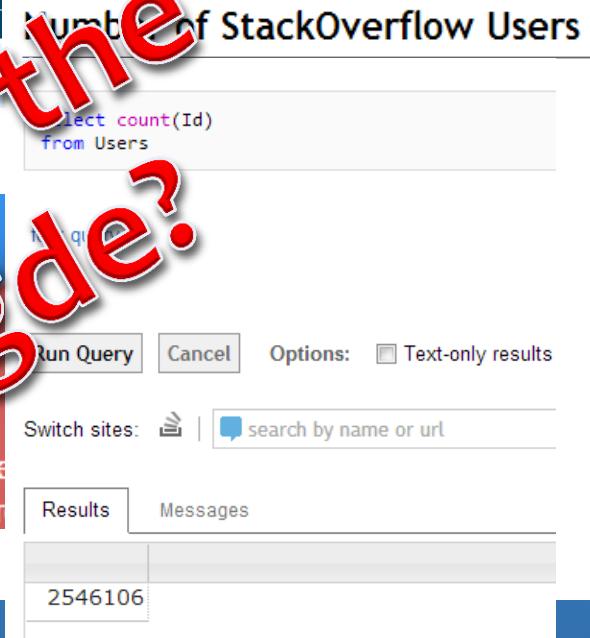


Comparing 3,141,952 open source contributors

Results

Messages

2546106

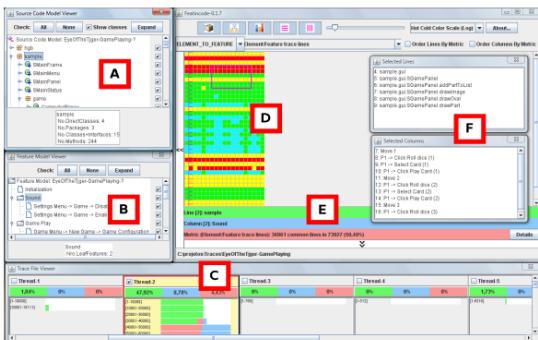


Some of our past work

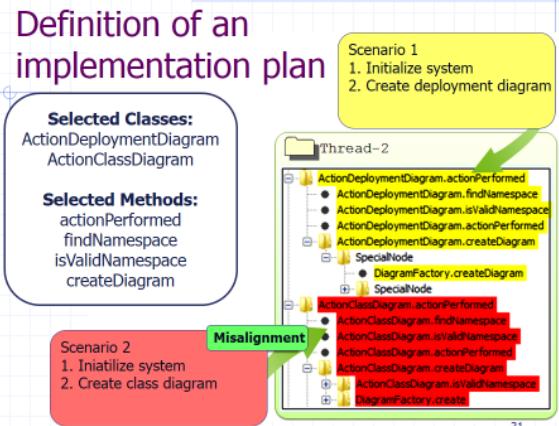
Execution traces

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

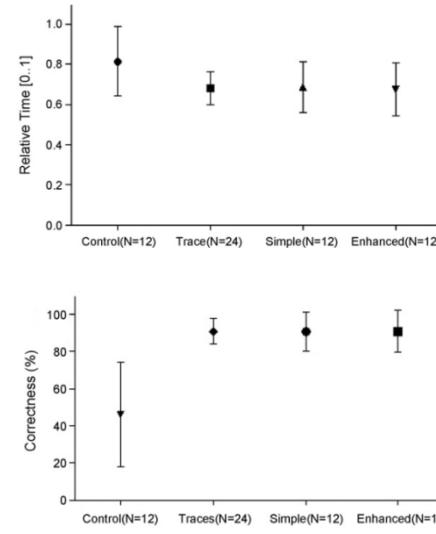
Tool
(WCRe2008)



Alignment
(CSMR2011)



Performance
assessment
(JSS2013)

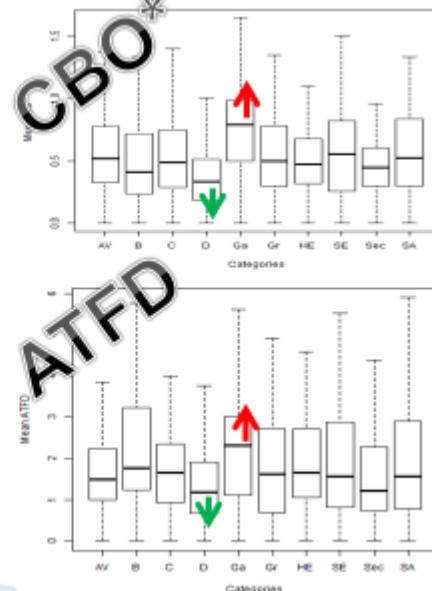
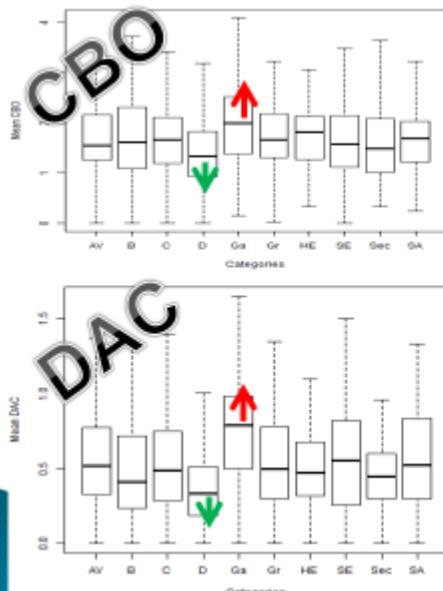


Some of our past work

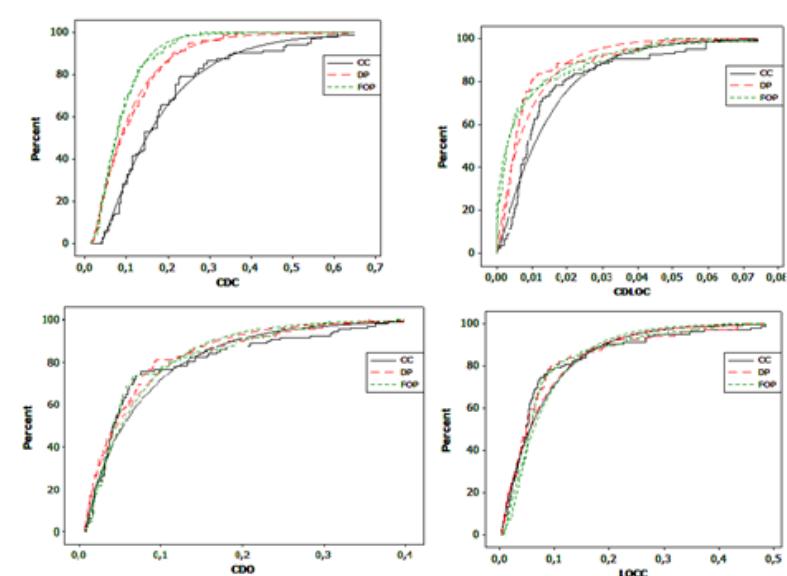
Metrics

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

Coupling x Categories (MSR2013)



FOP x SPLs (SCP2014)



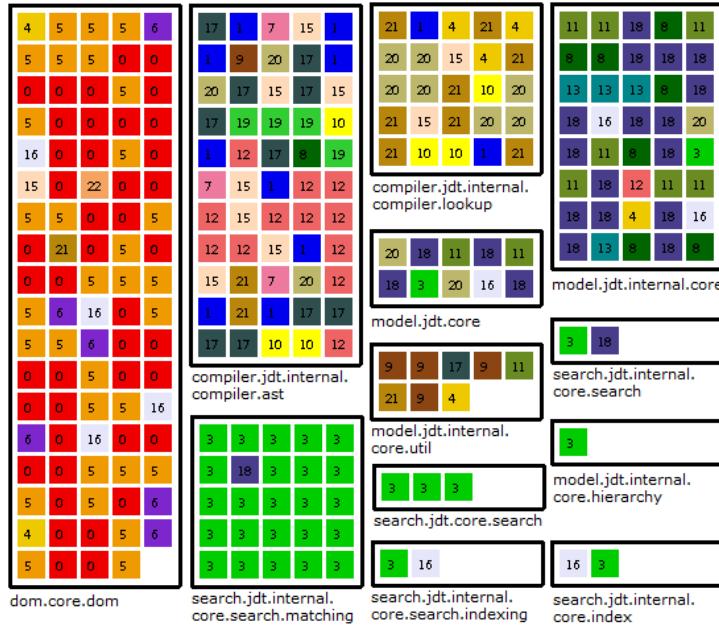
Some of our current work

Version history and Traces

Modularity Assessment and Comprehension

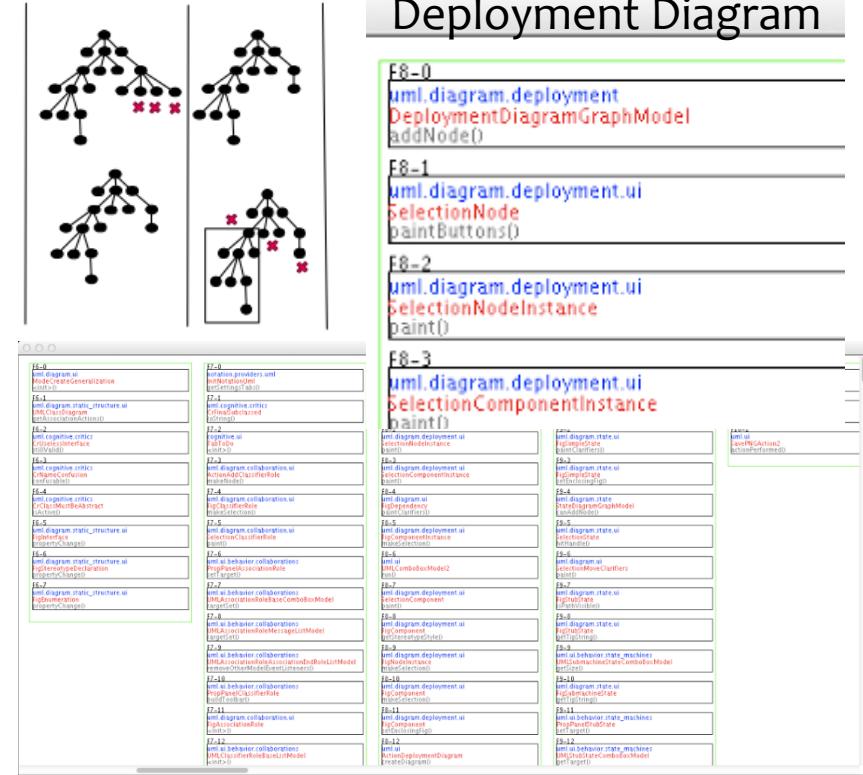
- Difficulties for cause-effect understanding, ground-truth evaluation
- Necessity to have large history of changes and high coverage traces

Co-change clusters (submitted)



YASAR (WIP)

Deployment Diagram

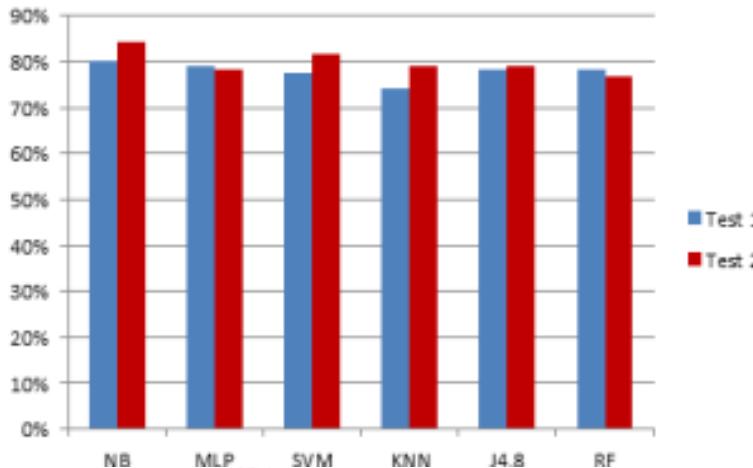


Some of our current work

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 Q. A.

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,pl 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*

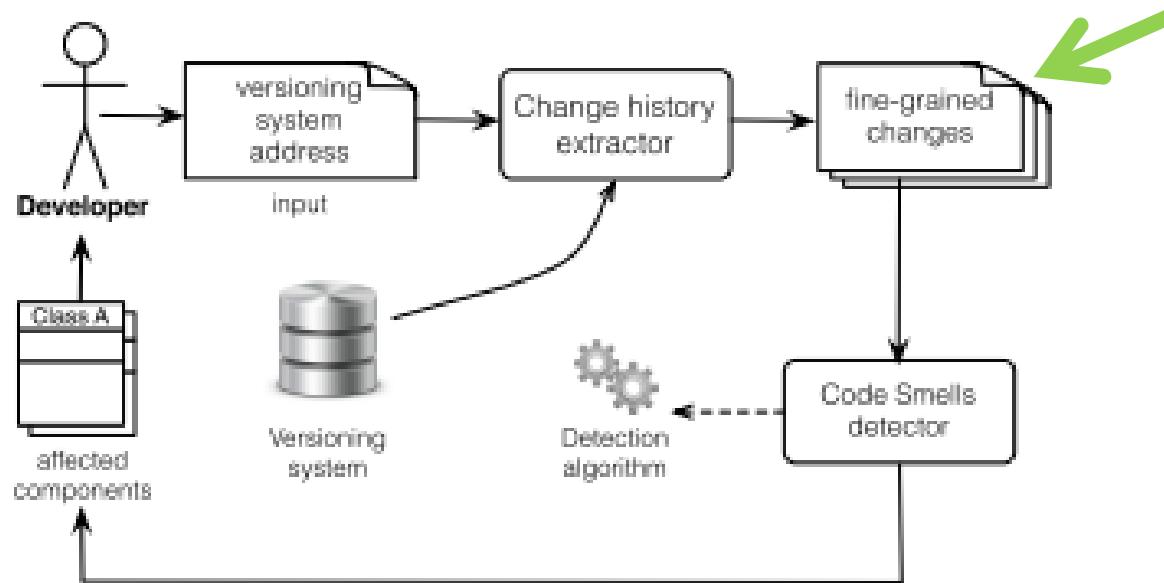


Fig. 1. HIST: The proposed code smell detection process.

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%

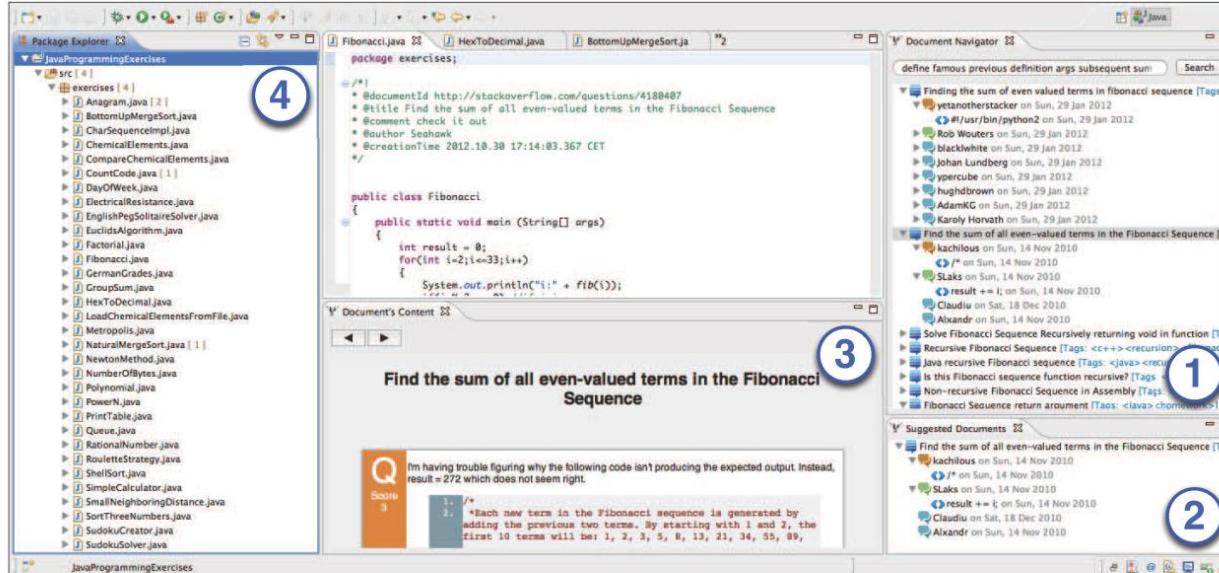
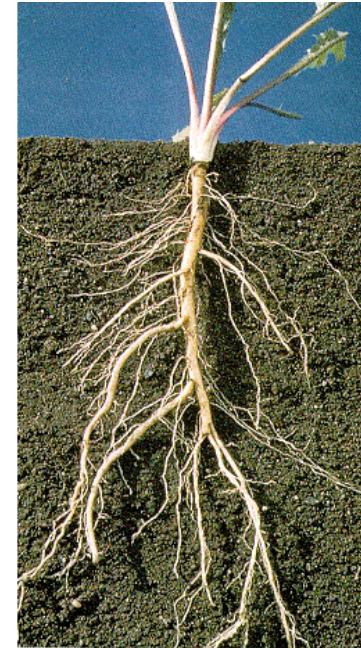
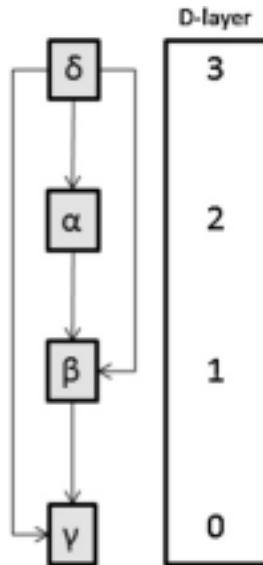
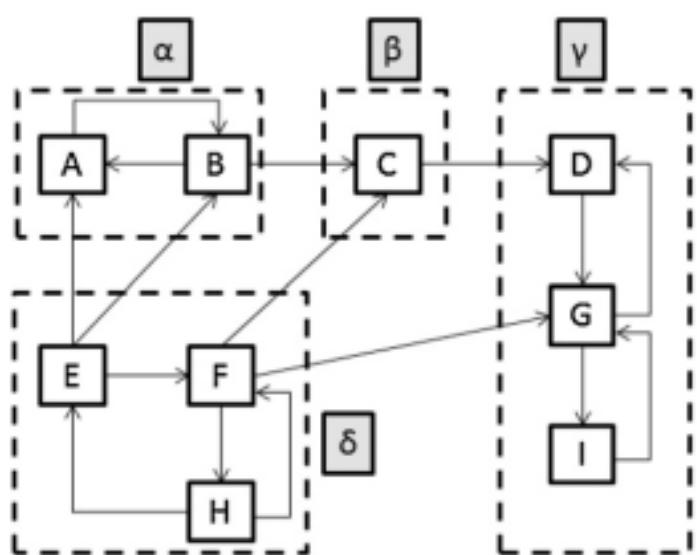


Figure 1: Seahawk User Interface

Some of related state-of-the-art work

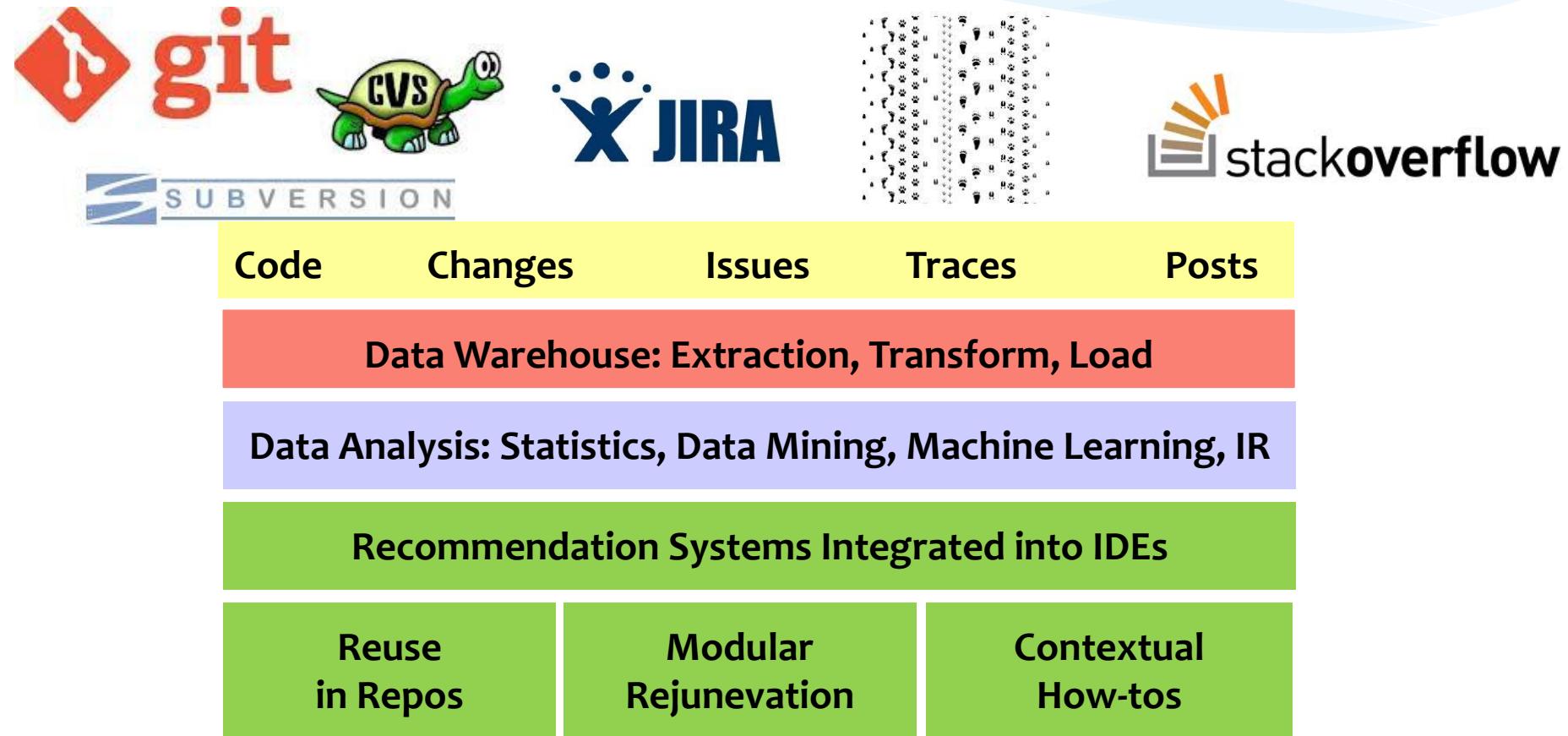
Kakarontzas and colleagues

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



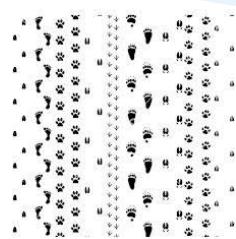
Envisioning our goals

Gluing the pieces together



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



Laboratory of
Software Comprehension, Analytics and Mining

UFU

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

