

Comparison of Similarity Metrics for Refactoring Detection

[Benjamin Biegel](#), [Quinten David Soetens](#), [Willi Hornig](#),
[Stephan Diehl](#), and [Serge Demeyer](#)



Mining Software Repositories

Honolulu, Hawaii. May 21, 2011.

 **Universität Trier**

 **Universiteit
Antwerpen**

Introduction

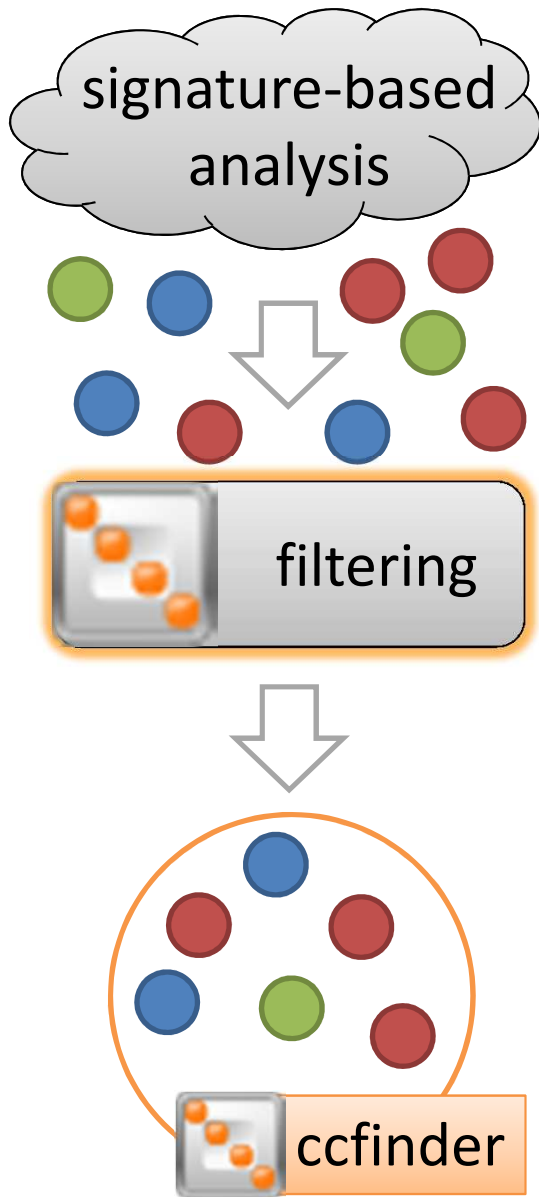
- Why is it important to detect refactorings?
- Replication experiment
 - Existing refactoring detection approach
 - **“Identifying Refactorings from Source-Code Changes”**
P. Weißgerber et al. (best paper, ASE’06)
 - But replaced the similarity metric (clone detector)
 - Non-identical replication (invited talk MSR’10)

Main Research Questions

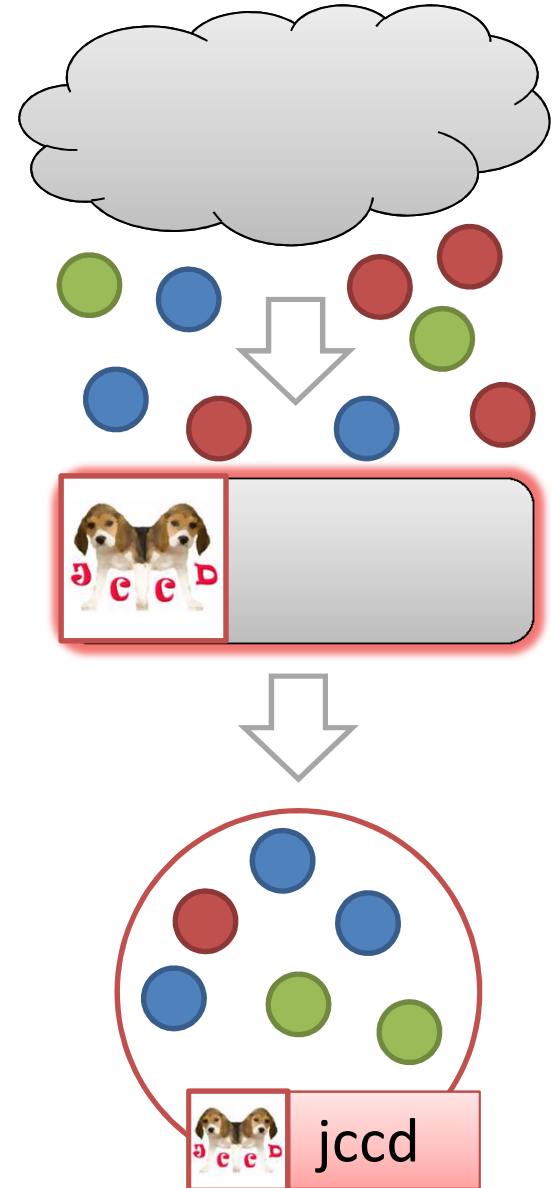
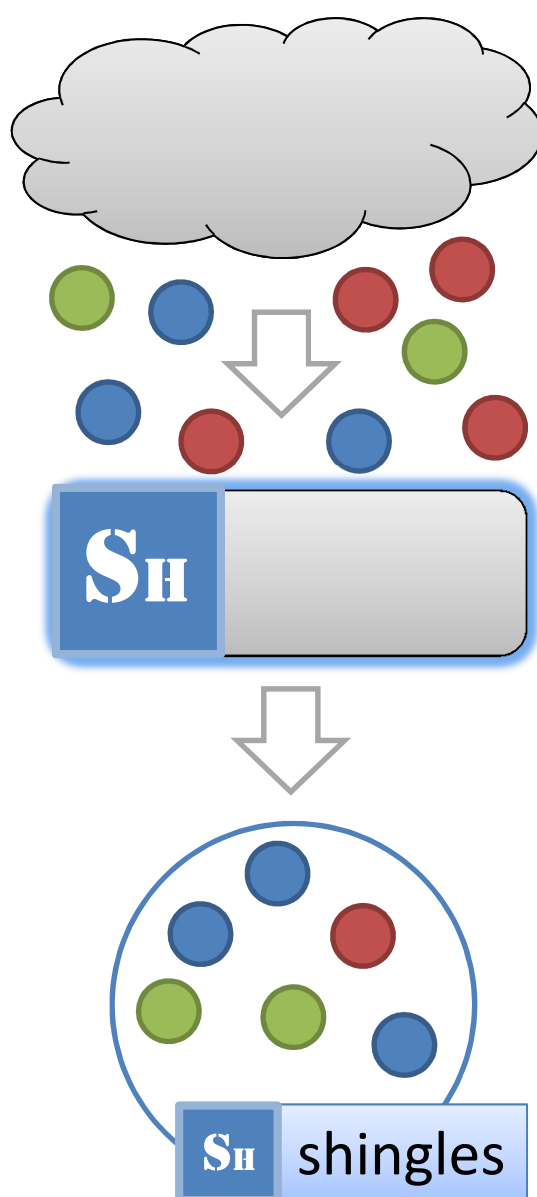
- I. How significant is the **influence** of each similarity metric?
- II. What is the **impact** of these metrics on the detected refactorings?

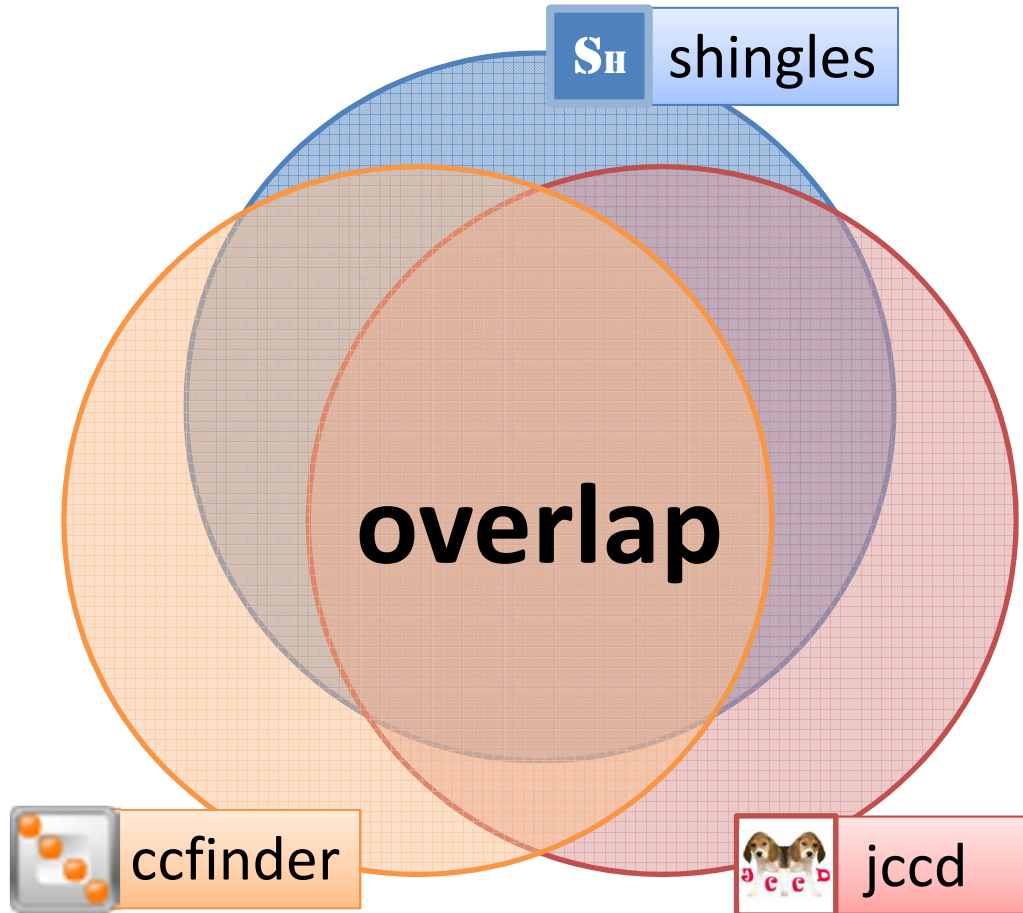
EXPERIMENTAL SETUP

original



replication





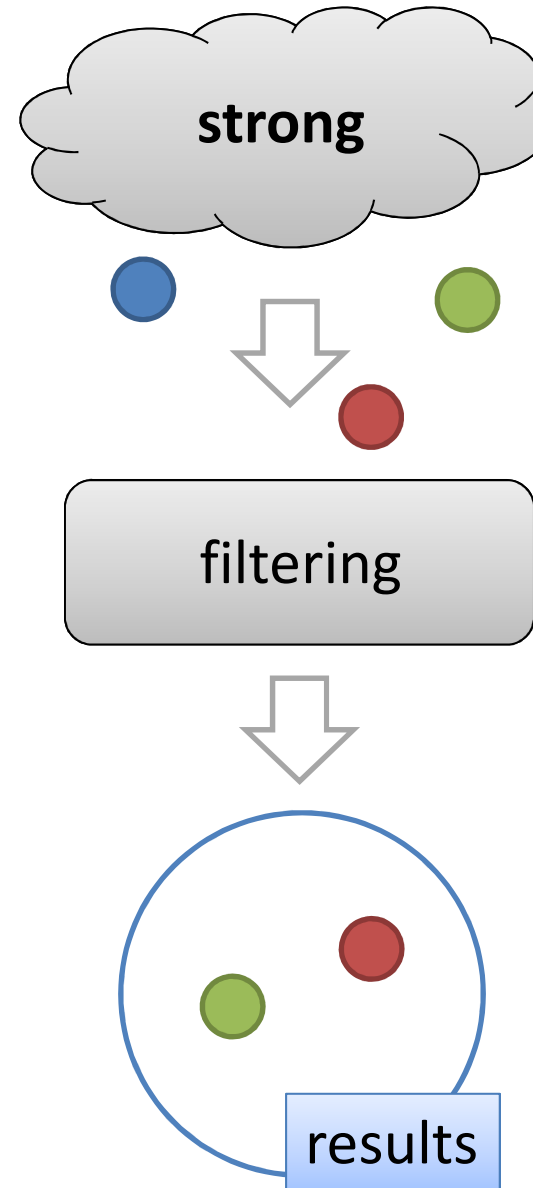
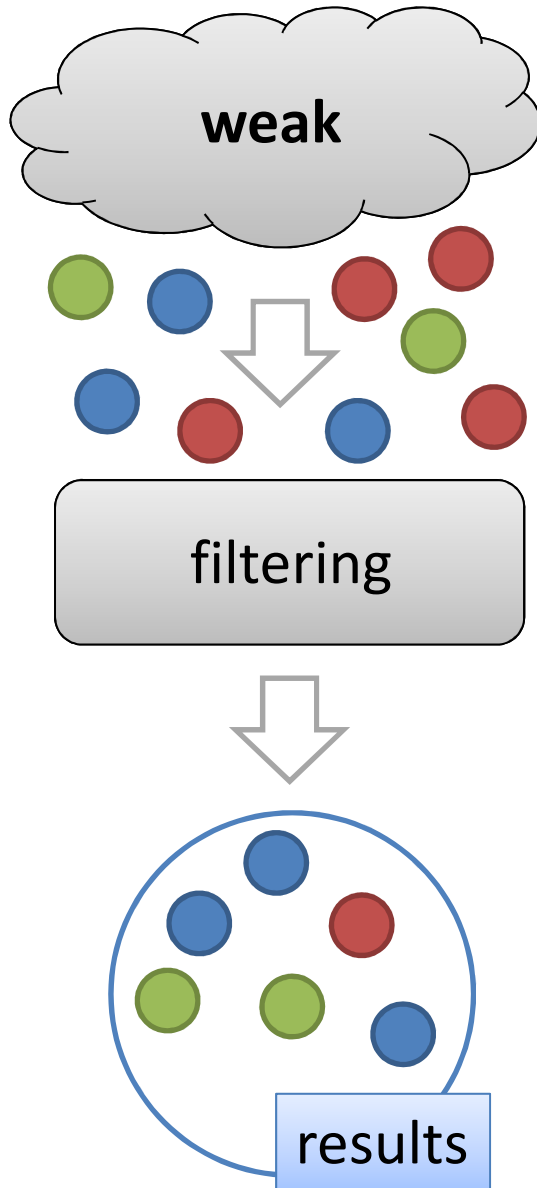
Evaluated Projects

- 4 of 7 projects of the original experiment
 - **Azureus**, 10664 transactions
 - **Jfreechart**, 2412 transactions
 - **JFtp**, 209 transactions
 - **Tomcat 3**, 4158 transaction

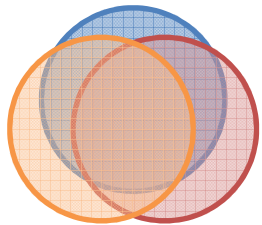
FILTERING STRATEGIES AND RESULTS

Filtering Strategy 1

weak and strong



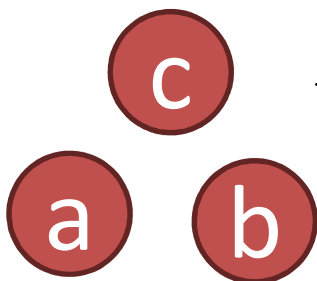
Overlap	
weak 2.1% - 22.8%	strong 51.5% - 81.2%



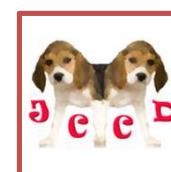
- the **strong** filter is very **selective**
 - much **smaller influence** of the clone detectors

Filtering Strategy 2

ambiguous candidates



best candidates



a

a

b

b

c

a

c

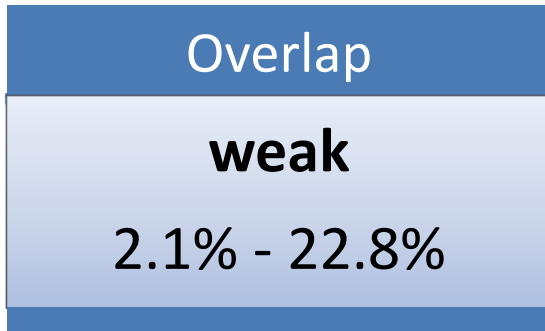
b

c

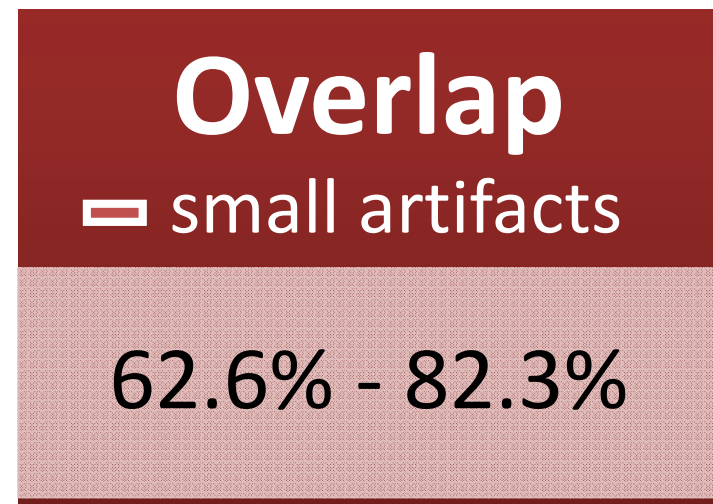
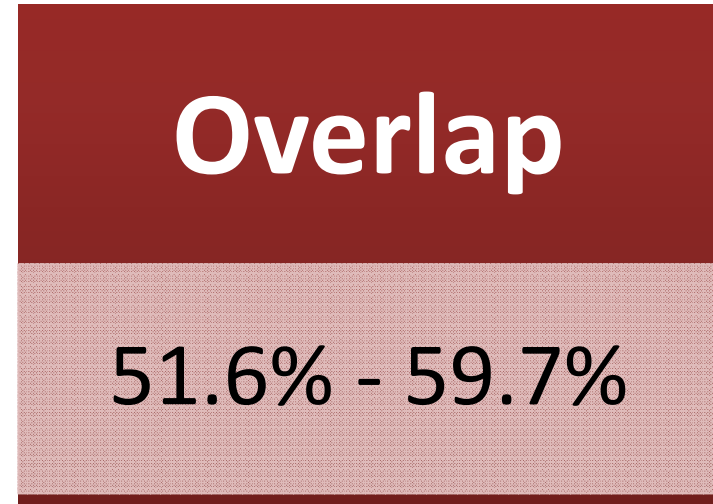
ranking

Filtering Strategy 2

ambiguous candidates



- still a large **common basis**
- but **different** rankings
- **less differences** by ignoring smaller artifacts



Conclusions

- I. How significant is the **influence** of each similarity metric?
- II. What is the **impact** of these metrics on the detected refactorings?
 - Specific characteristics of the clone detectors are mostly not represented in the detected refactoring candidates
 - The result sets have a comparable quality
 - Overlap has a pretty high recall

LESSONS LEARNED

Our Experience with the Replication

- Original raw data, processed data, scripts, tool, and documentation was available
- But still a steep learning curve
- Expert support important

Contact the original author first



Mahalo!