

Elmar Juergens, Nils Göde

# Achieving Accurate Clone Detection Results

May 8<sup>th</sup>, 2010  
4th IWSC, Cape Town

# Motivation

## Lack of Accuracy

- Clone detection produce false positives
- [Kapsner & Godfrey 2004]: between 27% and 65% false positives
- [Tiarks, Koschke & Falke 2009]: up to 75% false positives

## Consequences

- Diluted research conclusions
- Hindered adoption of clone detection by industry

## Questions

- When are clones relevant for *maintenance*?
- How can we achieve accurate clone detection results?

# Clone Relevance

## Clone Coupling

- Change to one clone may also need to be performed to siblings
- Fundamental characteristic that causes problems for *maintenance*
- Independent of whether clone can be removed!

## Relevance Question

*If you modify a clone during maintenance, do you want to be informed about its siblings to be able to modify them accordingly?*

# Tailoring

## Assessment

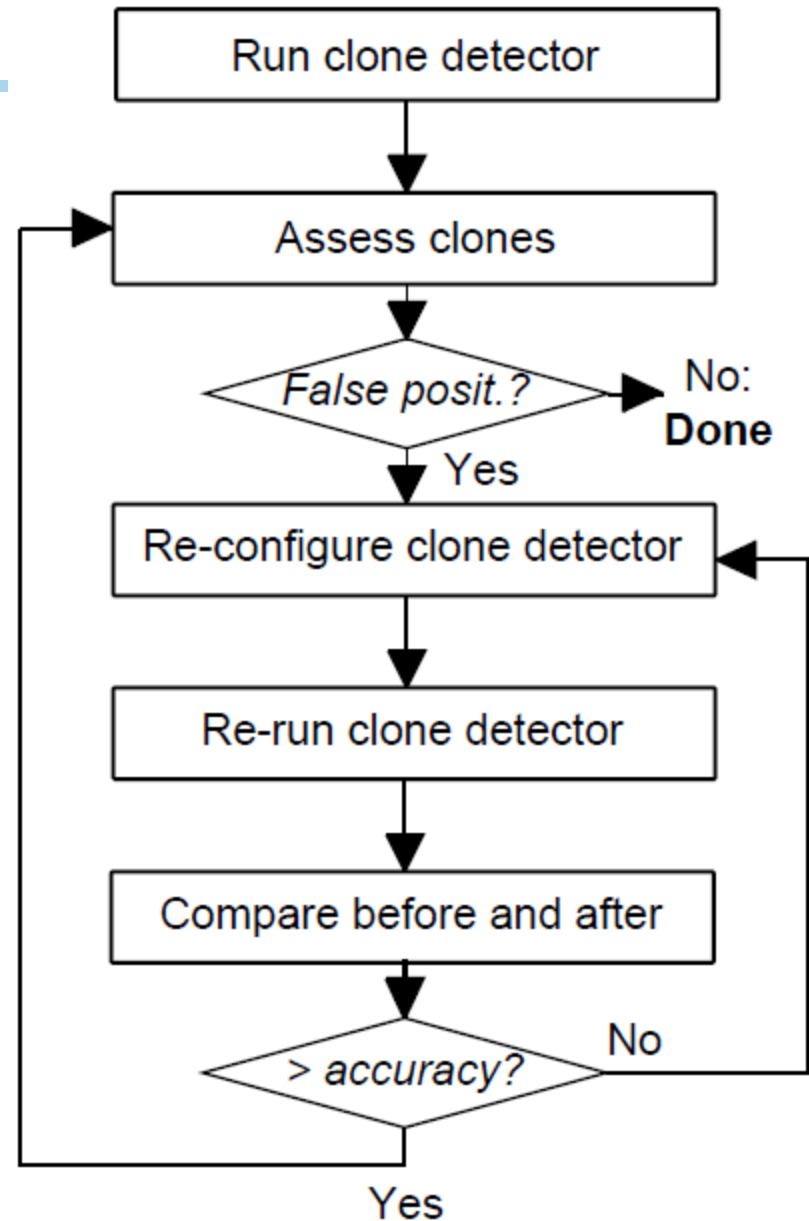
- Answer Relevance Question
- Random Sample

## Configuration

- Detector-dependent
- Code exclusion, shaping, Context-Sens. normalization...

## Assumptions

- Consistency
- Correctness
- Generalizability



# Research Questions

---

RQ1: Do developers estimate clone coupling *consistently*?

RQ2: Do developers estimate clone coupling *correctly*?

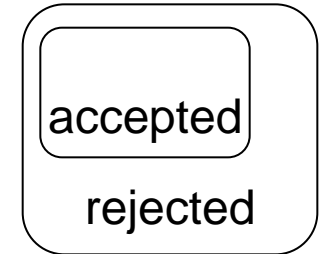
RQ3: Can coupling be generalized from a sample?

RQ4: How large is the impact of tailoring on clone detection results?

# RQ2+3: Correctness & Generalizability

## Study Object

- Munich Re, C#, 8 years, 360 kLOC, 4 developers (12 max)



## Study Design

- Tailoring partitions clone groups into *accepted* and *rejected*
- Change Coupling Ratio (CCR): *Probability that a change to one clone of a clone group should also be performed to at least one of its siblings.*
- H1: CCR for *accepted* is higher than for *rejected* clone groups.
- Determine CCR for both *accepted* and *rejected* clone groups on history

$$CCR = \frac{|consistent| + |inconsistent|}{|consistent| + |inconsistent| + |independent|}$$

- In all: 211 clone groups (1279 clones) manually assessed; 4 intervals in 1 year

# RQ2+3: Results

## Accepted

Int.	Consistent	Inconsistent	Independent	Coupling
1	15	3	3	0.857
2	11	1	10	0.545
3	31	6	13	0.740
4	1	0	0	1.000
1-4	58	10	26	0.723

## Rejected

Int.	Consistent	Inconsistent	Independent	Coupling
1	2	0	10	0.167
2	0	1	42	0.023
3	0	0	23	0.000
4	1	0	38	0.026
1-4	3	1	102	0.034

Developers do estimate clone coupling correctly

Coupling can be generalized from a random sample

# RQ4: Impact of Tailoring

## Study Design

- Compute cloning measures before and after tailoring; compare

## Results

	2006			2008		
	Untail.	Tail.	$\Delta$	Untail.	Tail.	$\Delta$
Clone Grps	598	332	-44%	2,558	1,028	-60%
Clones	2,118	1,005	-53%	12,675	3,558	-72%
Coverage	29.3%	18.3%	-38%	36.2%	19.4%	-46%
Blow-Up	27.8%	14.2%	-49%	41.2%	16.1%	-61%

- Substantial changes to metrics after tailoring
- 2006: Persistency hand-written; 2008: generated



# Summary & Future Work

## For the analyzed systems

- developers consistently estimate clone coupling
- estimation is correct and can be generalized from a sample
- Tailoring has a substantial impact on the results

## From our experience

- Results consistent with other industrial software (Java, Cobol, C/C++)
- Tailoring fundamental for adoption in industry

## Future Work

- Replication of study on further systems and other artifact than source code
- Further improve tailoring support

## Code exclusion

- Exclude code from detection
- Based on file extension / content

## Context sensitive normalization

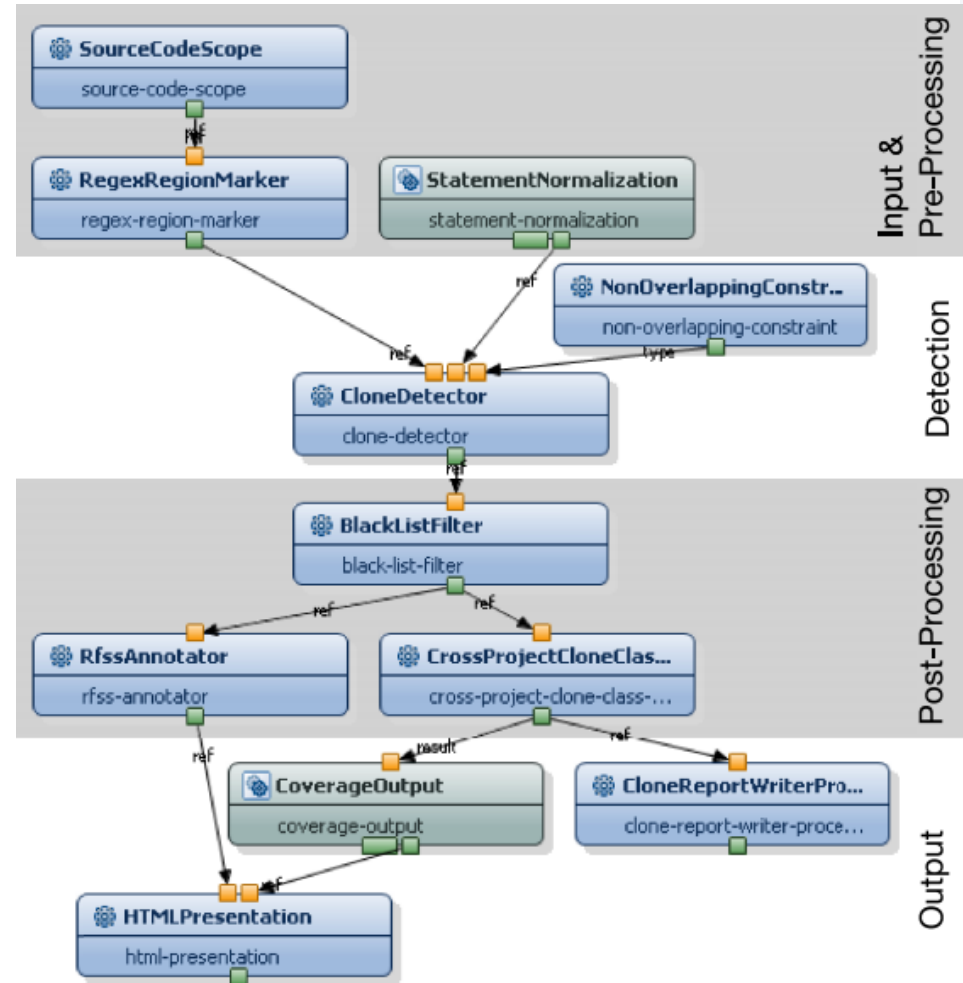
- Different notion of similarity for different regions of code

## Clone Shaping

- Cut to methods, blocks, ...

## Post-Detection Filtering

- Content-, gap-ratio-based, ...
- Blacklisting



# RQ1: Inter-Developer Consistency

## Study Object

- Munich Re; ABAP, 13 years, 442 kLOC, 10 developers (40 max)

## Study Design

- 3 developers inspect 48 (randomly chosen) clone pairs each
- Answer relevance question (accept; reject; undecided)
- Computation of inter-rater agreement: Cohen's Kappa

## Results

- Cohen's Kappa: 0.87  
“almost perfect agreement”
- Developers do estimate clone coupling consistently

Developers	Agreement
1 & 2	87.5%
1 & 3	85.4%
2 & 3	89.6%
1 & 2 & 3	81.3%
1 & 2 & 3 (w/o unrated)	88.1%