

Domain Specific Warnings: Are They Any Better?

ICSM 2012

André Hora, Nicolas Anquetil, Stéphane Ducasse, Simon Allier

RMoD team

INRIA, Lille, France

andre.cavalcante.hora@inria.fr



Outline

- **Generic/Domain Specific Warnings**
- Assessing and Matching Warnings and Defects
- Research Questions
- Experiment Setting
- Experiment Results
- Conclusion

Generic Warnings

- Coding standard violation tools can ensure:
 - code quality, portability, common code patterns, support software maintenance, **prevent future bugs**



- Rule violations may be numbered by the thousands
 - False positives

Domain Specific (DS) Warnings

- H: Rules used are generic and are not focusing on domain-specific problems
- DS rules must be defined by an expert
- The efficiency of such **domain specific rules for defect prevention or reduction** needs therefore to be demonstrated



Generic vs. Domain Specific

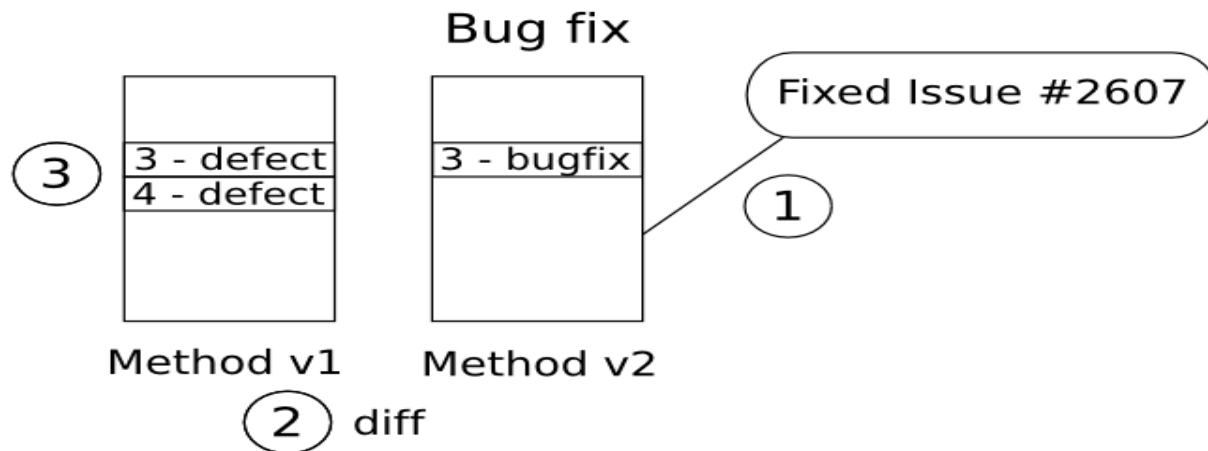
- Systematic study to verify the relation between **generic or domain warnings**, and observed **defects**
- Questions:
 - Is there a relation between generic **warnings** and **defects**?
 - Is there a relation between **DS warnings** and **defects**?

Outline

- Generic/Domain Specific Warnings
- **Assessing and Matching Warnings and Defects**
- Research Questions
- Experiment Setting
- Experiment Results
- Conclusion

Assessing and Matching Warnings and Defects (1)

- Lines of code level
- Identifying lines with:
 - Defects: (1) Mining commit messages, (2) diff, (3) mark
 - Warnings point to lines



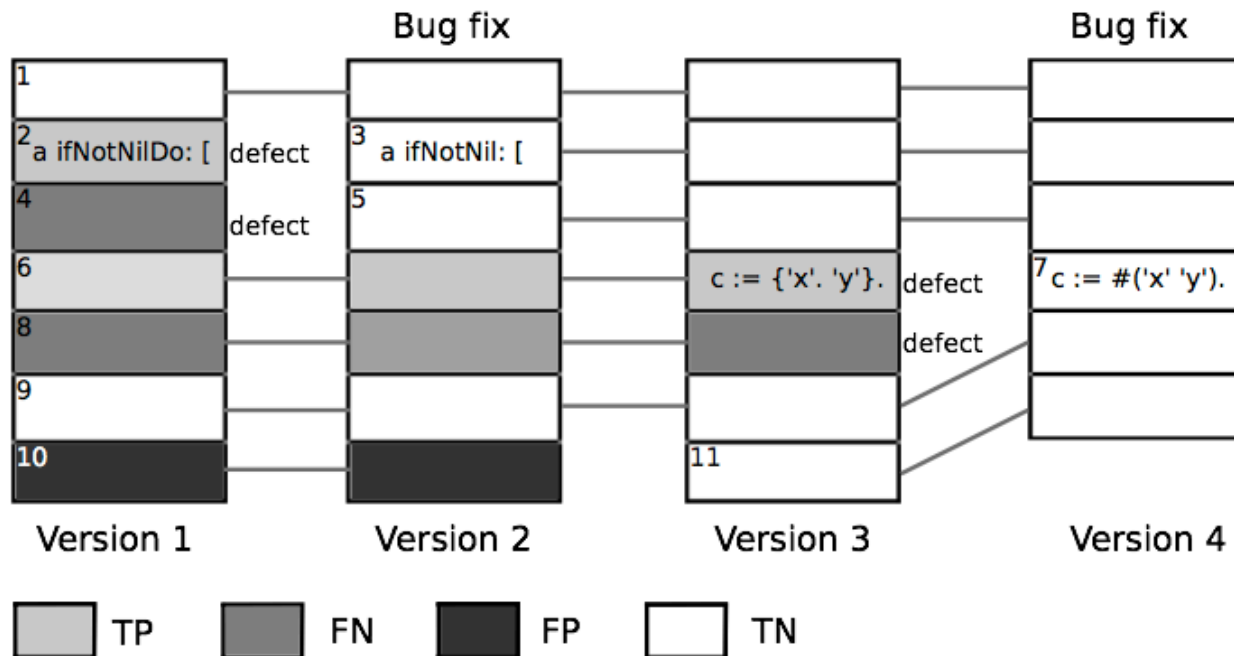
Assessing and Matching Warnings and Defects (2)

- Lines marked as warning and/or defect related

Lines of code...	with defect	without defect
with warning	TP	FP
without warning	FN	TN

Assessing and Matching Warnings and Defects (3)

- Unique lines of code (ULoC)



Outline

- Generic/Domain Specific Warnings
- Assessing and Matching Warnings and Defects
- **Research Questions**
- Experiment Setting
- Experiment Results
- Conclusion

Research Questions

- Evaluating generic and DS rules
 - **RQ1** Is there a relation between **generic warnings** and **defects**?
 - **RQ2** Is there a relation between **DS warnings** and **defects**?

Outline

- Generic/Domain Specific Warnings
- Assessing and Matching Warnings and Defects
- Research Questions
- **Experiment Setting**
- Experiment Results
- Conclusion

Experiment Setting (1)

- Context: real systems, commit logs linked to an issue database, generic and DS rules
- **Seaside**: open-source web application framework written in Smalltalk



Number of snapshots	943
Average classes per snapshot	216
Average methods per snapshot	1,592
Average LOC per snapshot	6,428

- Generic (SmallLint) rules vs. Domain specific rules (Slime)

Experiment Setting (2)

- Is there a relation between **warnings** and **defects**?
 - **H0** Warnings and defects are **independent**
 - **Ha** Warnings and defects are **related**
 - Chi square test

Experiment Setting (3)

- Defects: 14,416 ULoCs, 664 with defects (4.6%)
- Warnings: used rules
 - **91 generic rules:** Unnecessary code, Spelling, Possible bugs, Pharo bugs, Bugs, Miscellaneous, Intention revealing
 - **29 DS rules:** Portability, Compatibility, Possible bug

Outline

- Generic/Domain Specific Warnings
- Assessing and Matching Warnings and Defects
- Research Questions
- Experiment Setting
- **Experiment Results**
- Conclusion

Experiment Results - All rules

- **RQ1** Is there a relation between generic warnings and defects?
 - **H0: Generic warnings and defects are independent**
 - Ha: Generic warnings and defects are related
 - $p\text{-value} = 0.65$, **residuals close to zero**, precision = 5%

Lines of code...	with defect	without defect	total
with warning	55 (-0.48)	1,063 (-0.10)	1,118
without warning	609 (-0.14)	12,689 (0.03)	13,298
total	664	13,752	14,416

Experiment Results - All rules

- **RQ2** Is there a relation between DS warnings and defects?
 - H0: DS warnings and defects are independent
 - **Ha: DS warnings and defects are related**
 - *p-value* < 0.001, **residual over-represented**, effect-size = 5.1%, precision = 12 %

Lines of code...	with defect	without defect	total
with warning	37 (5.97)	275 (-1.31)	312
without warning	627 (-0.88)	13,477 (0.19)	14,104
total	664	13,752	14,416

Experiment Results - Top rules

- Until now:
 - Generic warnings and defects: **independent**
 - DS warnings and defects: **related**
- Warnings: top rules
 - Best correlation with defects (**best precision**)
 - Rules performed better than random predictor

Experiment Results - Top rules

- **RQ4** Is there a relation between top generic warnings and defects?
 - H0: Top generic warnings and defects are independent
 - **Ha: Top generic warnings and defects are related**
 - $p\text{-value} < 0.001$, **residual over-represented**, ES = 4.2%, precision = 11%

Lines of code...	with defect	without defect	total
with warning	32 (4.91)	267 (-1.07)	299
without warning	632 (-0.71)	13,485 (0.15)	14,117
total	664	13,752	14,416

Experiment Results - Top rules

- **RQ5** Is there a relation between top DS warnings and defects?
 - H0: Top DS and defects are independent
 - **Ha: Top DS warnings and defects are related**
 - *p-value* < 0.001, **residual over-represented**, ES = 7.6 %, precision = 19%

Lines of code...	with defect	without defect	total
with warning	32 (8.85)	133 (-1.94)	165
without warning	632 (-0.95)	13,619 (0.20)	14,251
total	664	13,752	14,416

Experiment Results - Top rules

- **RQ6** Are top DS warnings more likely to point to defects than top generic warnings?
 - H0: Top DS and top generic warnings have the same precision
 - **Ha: Top DS warnings precision is higher**
 - 77 methods with at least one top generic warning and 67 with at least one top DS warning
 - *p-value* = 0.047, effect-size = 14%

Conclusion

- **Generic rules** are **not effective** enough to be used for defect prevention
- **DS rules** give more relevant information to avoid bugs
 - **Effective** to be used for defect prevention
- **Top DS rules** are **more effective** to be used for defect prevention than top generic rules
- We expect DS rules to be created and used by developers in complement to generic ones
- **Future work:** compare DS rules with rules extracted from project history, compare by category, more case studies

Domain Specific Warnings: Are They Any Better?

ICSM 2012

André Hora, Nicolas Anquetil, Stéphane Ducasse, Simon Allier

RMoD team

INRIA, Lille, France

andre.cavalcante.hora@inria.fr

