Confusion Detection in Code Reviews


Felipe Ebert


Fernando Castor


Nicole Novielli


Alexander Serebrenik

Universidade Federal de Pernambuco

Università degli Studi di Bari Aldo Moro

TU/e Technische Universiteit Eindhoven University of Technology

FACEPE Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco

CAPES

Confusion Detection in Code Reviews


Felipe Ebert


Fernando Castor


Nicole Novielli


Alexander Serebrenik

Universidade Federal de Pernambuco

Università degli Studi di Bari Aldo Moro

TU/e Technische Universiteit Eindhoven University of Technology

FACEPE Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco

CAPES

Confusion Detection in Code Reviews

Felipe Ebert

Fernando Castor

Nicole Novielli

Alexander Serebrenik

Universidade Federal de Pernambuco

Università degli Studi di Bari Aldo Moro

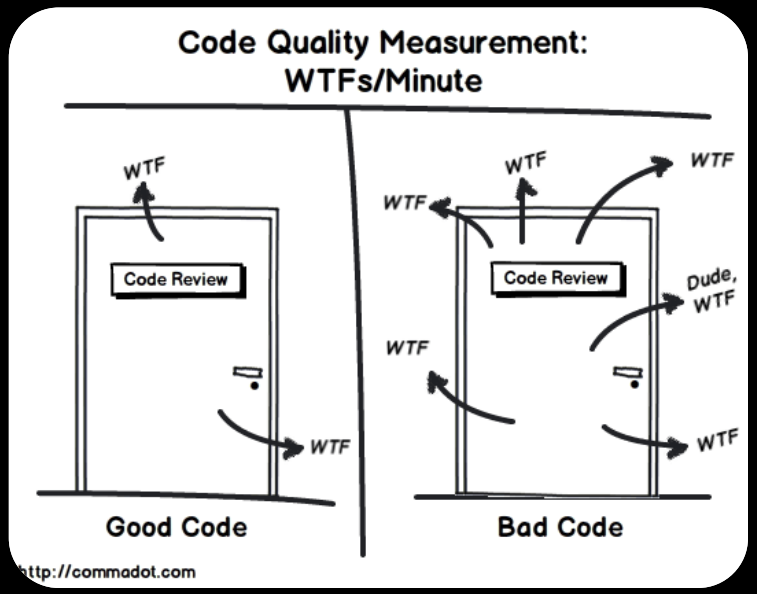
TU/e Technische Universiteit Eindhoven University of Technology

FACEPE Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco

CAPES

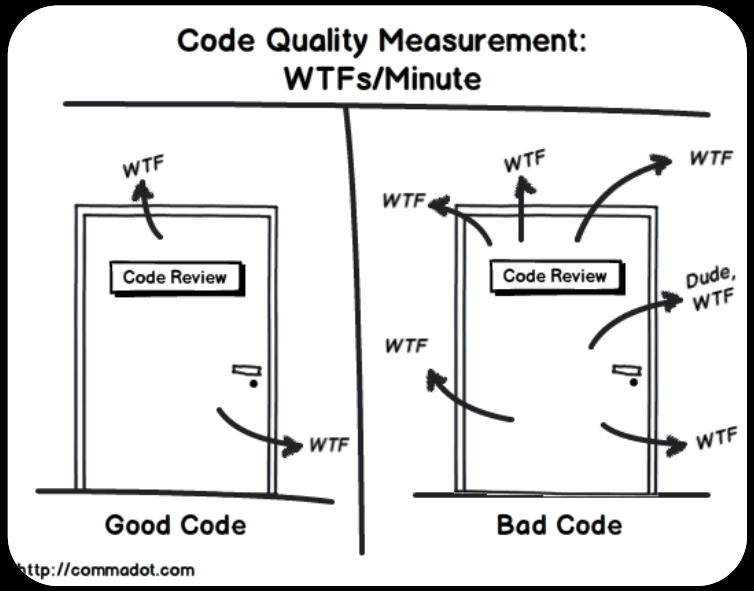
Confusion!!!

Why?



Confusion!!!

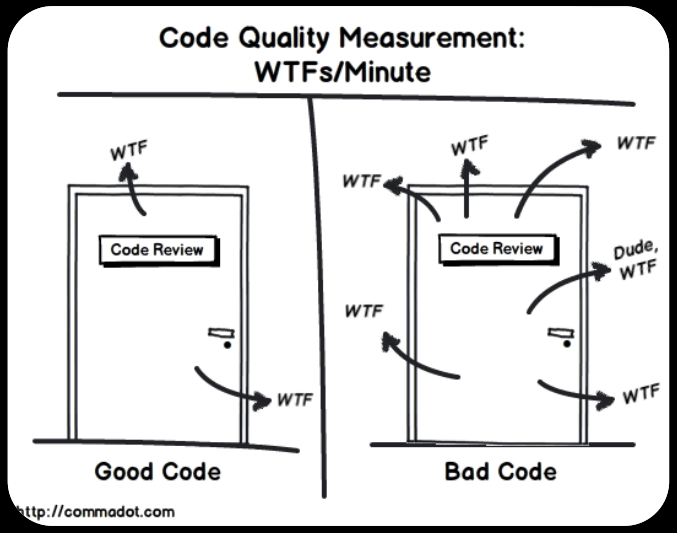
Why?



What?

"a situation in which people are ***uncertain*** *about what to do or are* ***unable to understand*** *something clearly"*

Merriam-Webster



Patch Set 2: Code-Review+2

*Though **I don't really understand** why ValueObject moved to runtime...*

https://android-review.googlesource.com/110347

Patch Set 1:

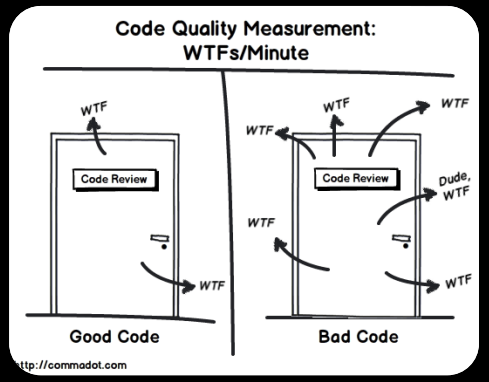
***What's the context?** Is this fixing/improving existing code? Could you use the assembler tests for it?*

https://android-review.googlesource.com/140403

***why do you need any pixels here?** as I understand, nullptr could be OK here, as this is an output, not input texture*

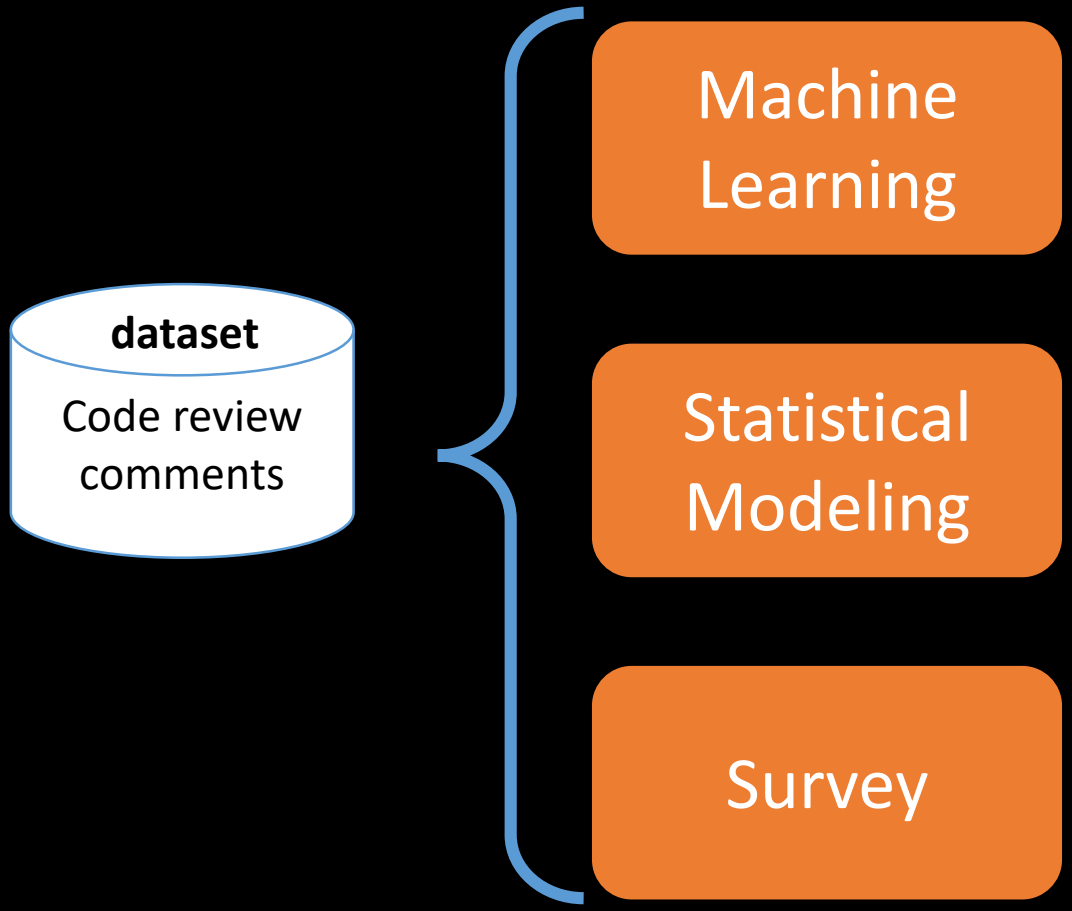
https://android-review.googlesource.com/291770


gerrit
Gerrit Code Review


Code Quality Measurement:
WTFs/Minute
WTF
Code Review
WTF
Good Code
WTF
WTF
WTF
Code Review
Dude, WTF
WTF
WTF
Bad Code
http://commadot.com


ACT NOW
Shift

To understand the reasons and consequences of confusion in code reviews



Provide the code documentation

Reviewers

Patch Set 2: Code-Review+2

*Though **I don't really understand** why ValueObject moved to runtime...*

Guidelines with best practices on coding and submitting for review

Authors

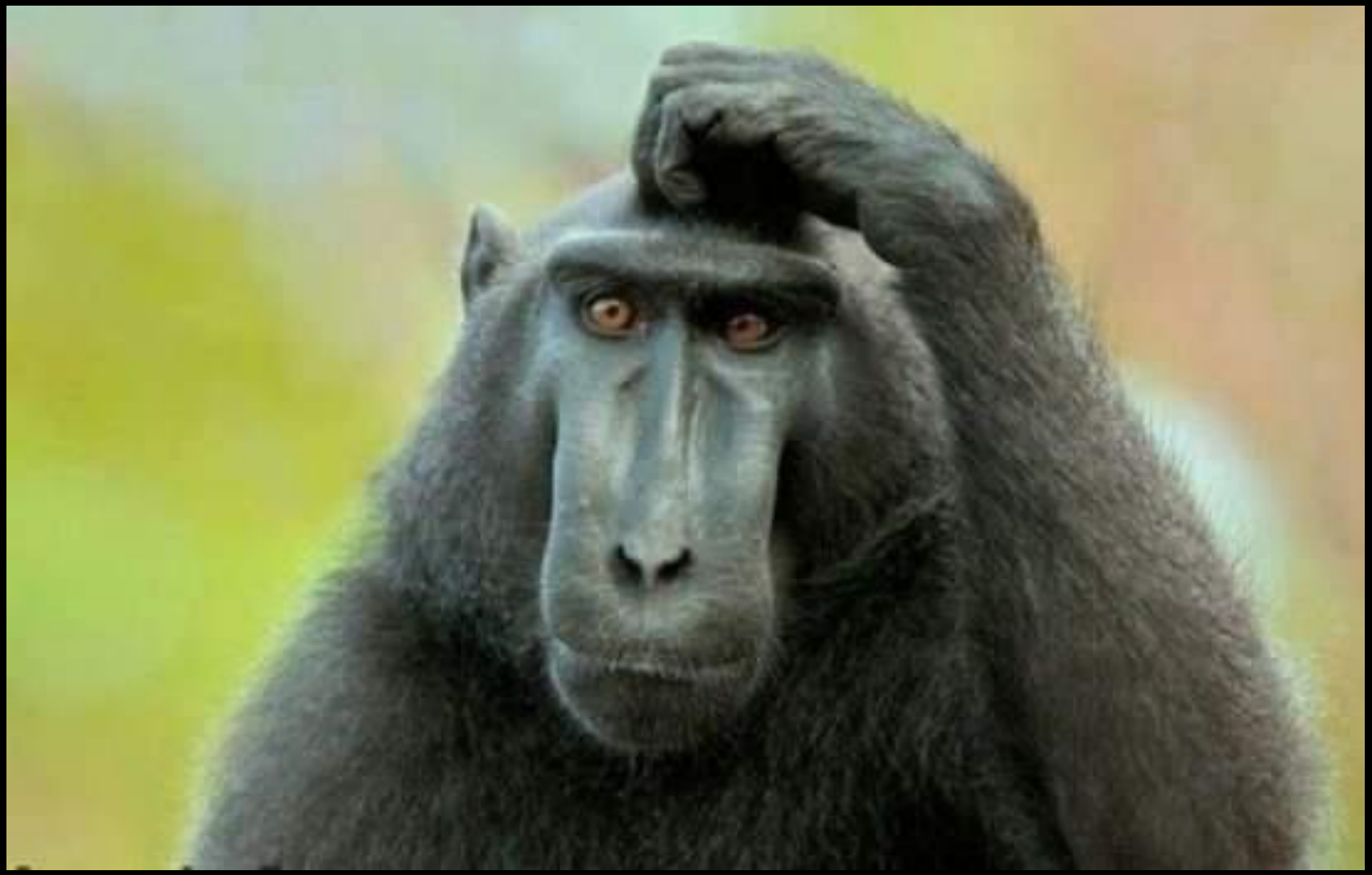
Patch Set 1:

***What's the context?** Is this fixing/improving existing code? Could you use the assembler tests for it?*

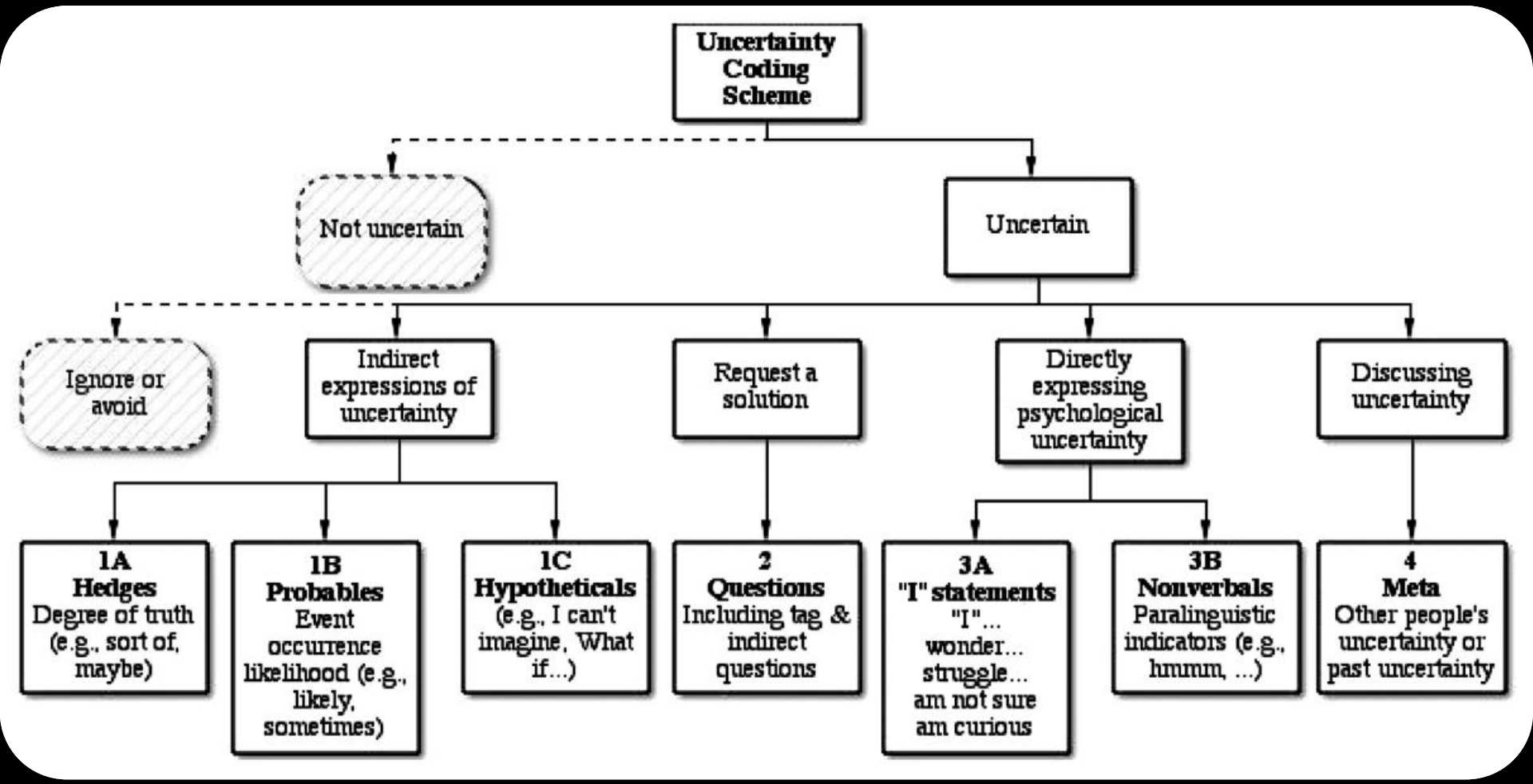
Provide other parts of the code

Reviewers

***why do you need any pixels here?** as I understand, nullptr could be OK here, as this is an output, not input texture*



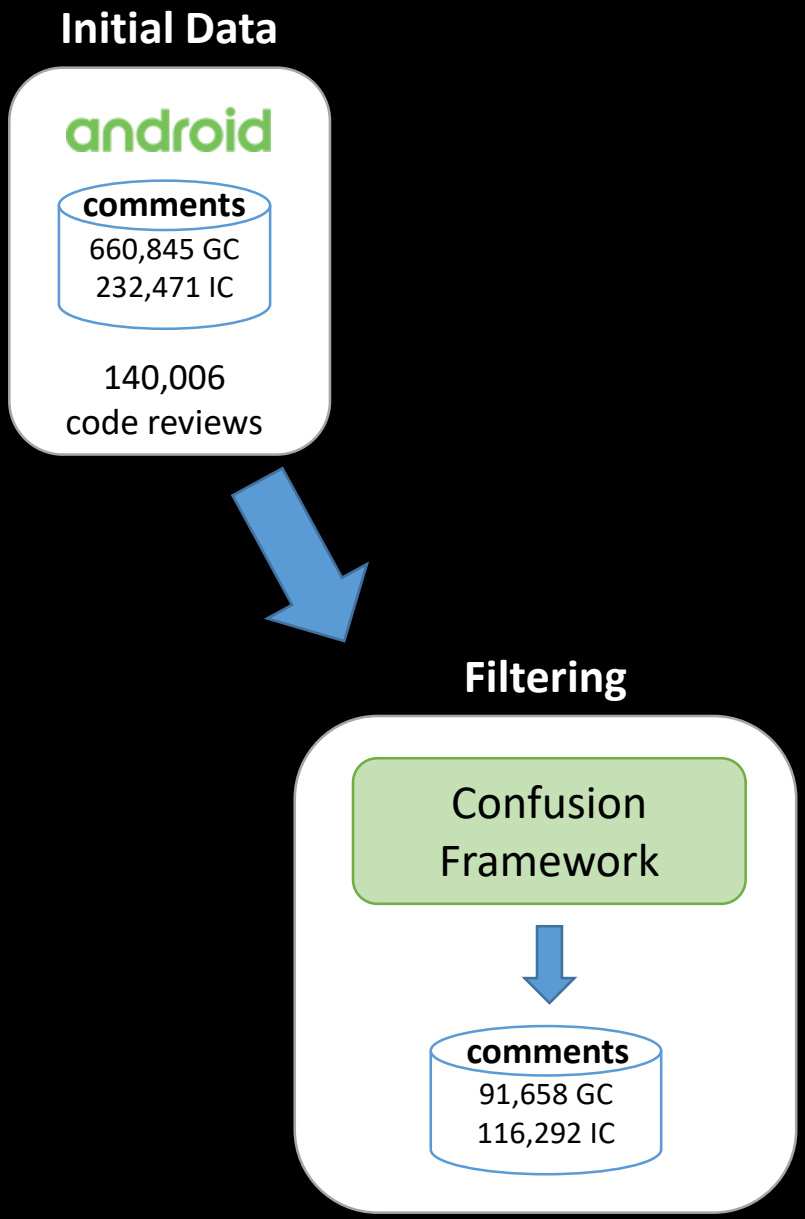
How do we identify and measure confusion?



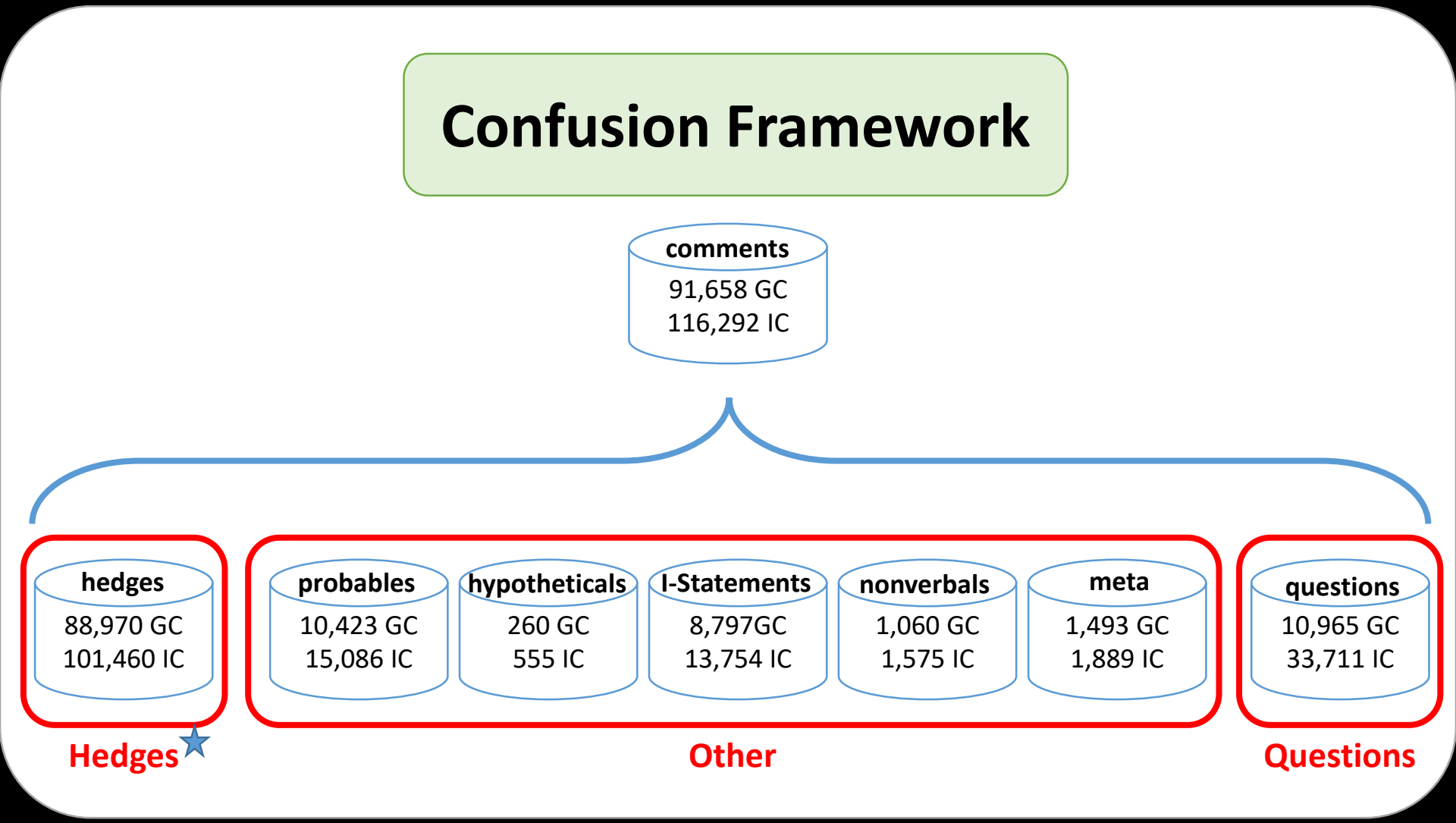
M. E. Jordan, D. L. Schallert, Y. Park, S. Lee, Y. hui Vanessa Chiang, A.-C. J. Cheng, K. Song, H.-N. R. Chu, T. Kim, and H. Lee, "Expressing uncertainty in computer-mediated discourse: Language as a marker of intellectual work," Discourse Processes, vol. 49, no. 8, pp. 660–692, 2012.

Initial Data

GC – General Comment
IC – Inline Comment


Initial Data
android
comments
660,845 GC
232,471 IC
140,006
code reviews
Filtering
Confusion
Framework
comments
91,658 GC
116,292 IC

Filtering


Confusion Framework
comments
91,658 GC
116,292 IC
hedges
88,970 GC
101,460 IC
probables
10,423 GC
15,086 IC
hypotheticals
260 GC
555 IC
I-Statements
8,797GC
13,754 IC
nonverbals
1,060 GC
1,575 IC
meta
1,493 GC
1,889 IC
questions
10,965 GC
33,711 IC
Hedges
Other
Questions

Initial Data

android

comments
660,845 GC
232,471 IC

140,006
code reviews

Filtering

Confusion Framework

comments
91,658 GC
116,292 IC

Maybe *write a comment with the XML format here*

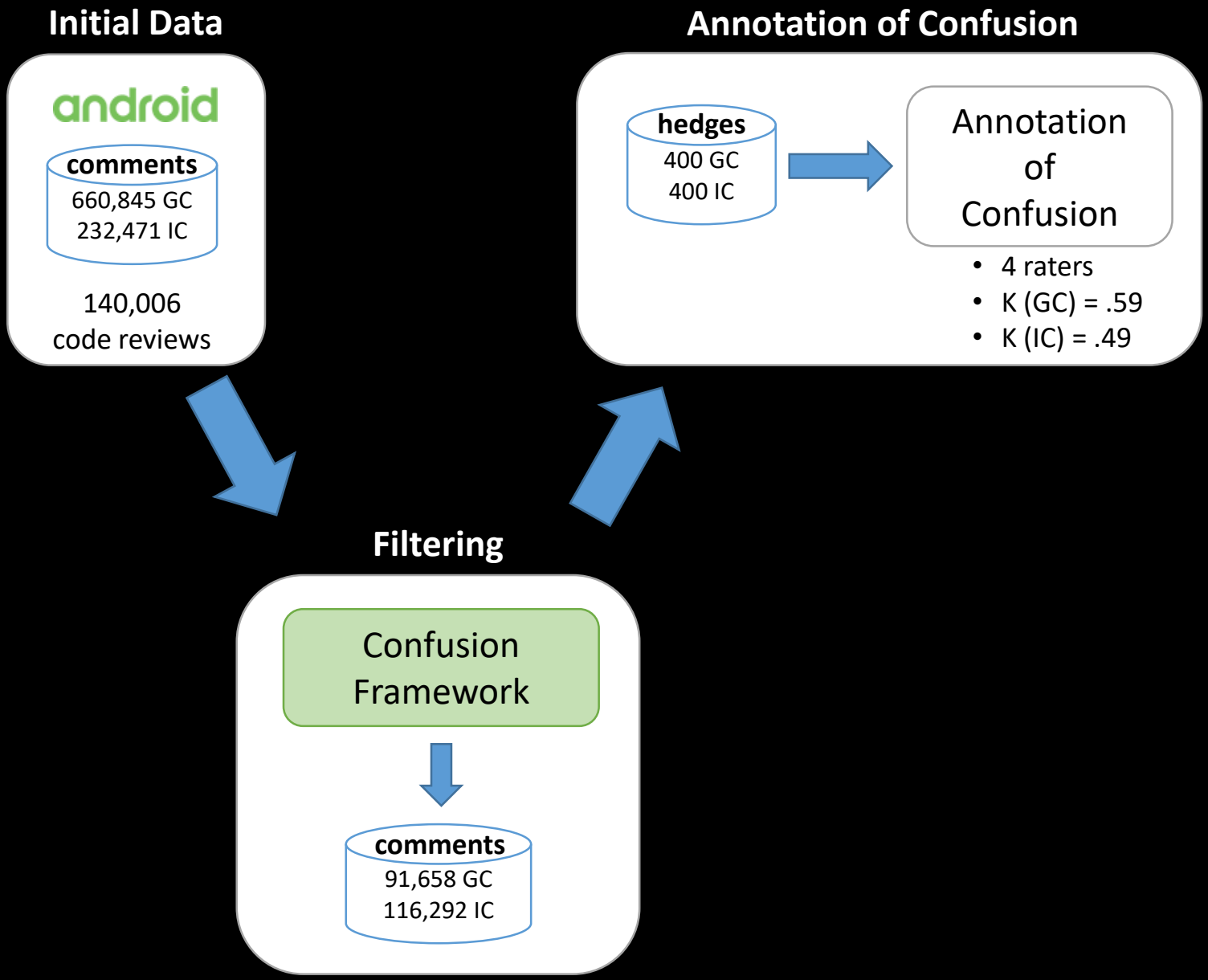
no confusion!

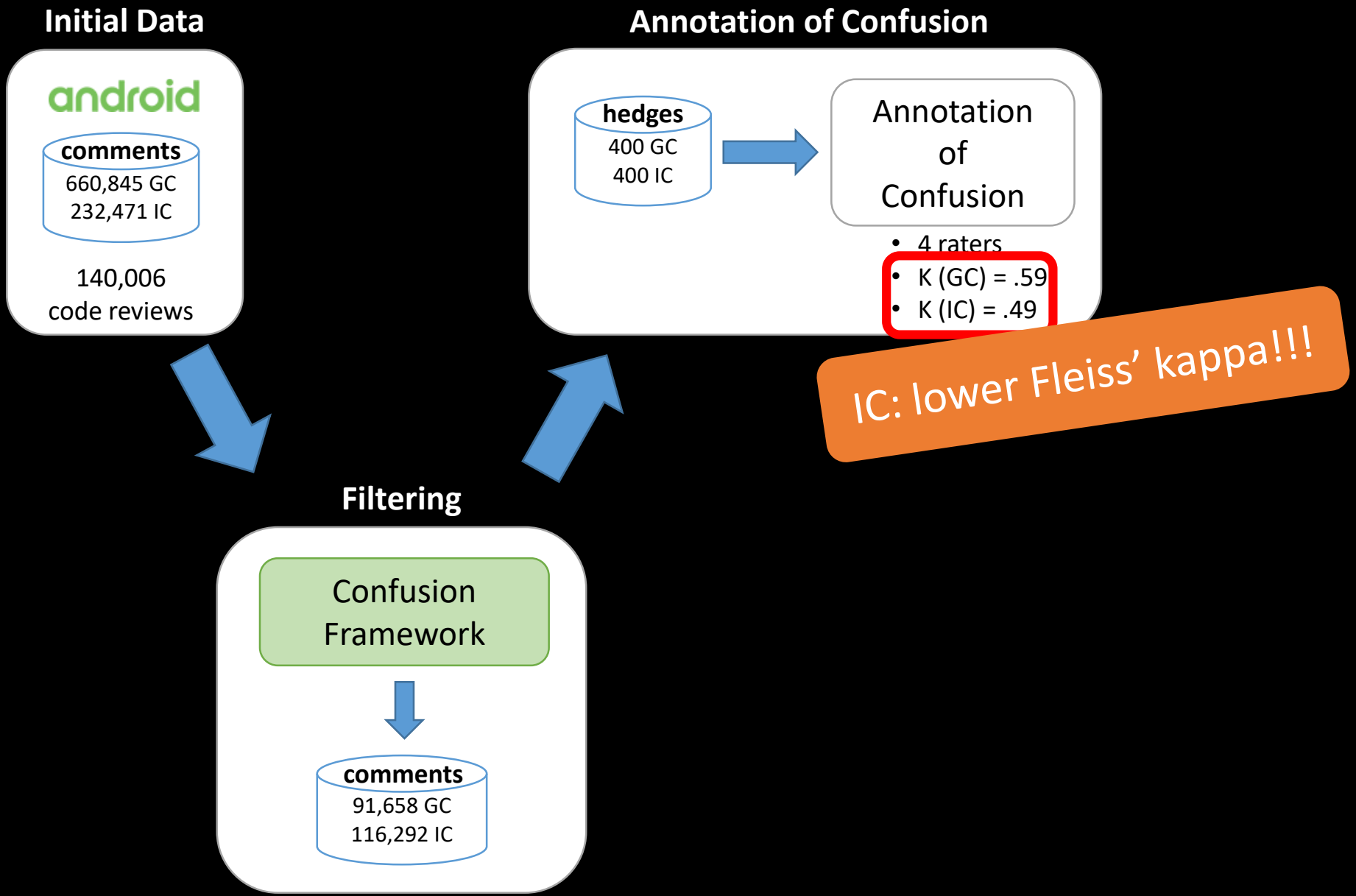
Patch Set 1: ***Could*** *anyone submit this?*

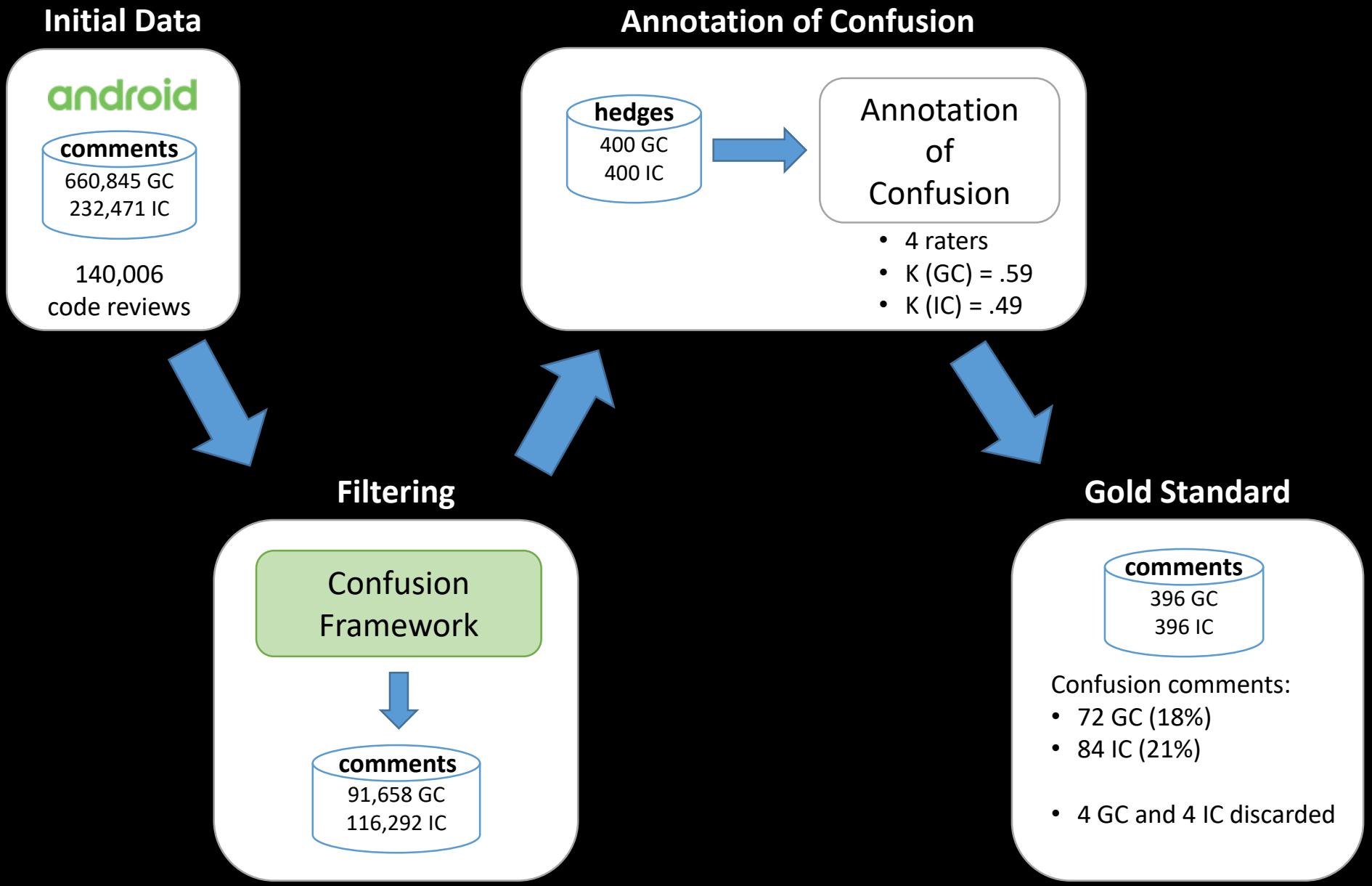
no confusion!

Patch Set 5: Svet: ***Could*** *you please review?*

no confusion!


Initial Data
android
comments
660,845 GC
232,471 IC
140,006
code reviews
Annotation of Confusion
hedges
400 GC
400 IC
Annotation
of
Confusion
• 4 raters
• K (GC) = .59
• K (IC) = .49
Filtering
Confusion
Framework
comments
91,658 GC
116,292 IC


Initial Data
android
comments
660,845 GC
232,471 IC
140,006
code reviews
Annotation of Confusion
hedges
400 GC
400 IC
Annotation
of
Confusion
4 raters
K (GC) = .59
K (IC) = .49
IC: lower Fleiss' kappa!!!
Filtering
Confusion
Framework
comments
91,658 GC
116,292 IC


Initial Data
android
comments
660,845 GC
232,471 IC
140,006
code reviews
Annotation of Confusion
hedges
400 GC
400 IC
Annotation
of
Confusion
• 4 raters
• K (GC) = .59
• K (IC) = .49
Filtering
Confusion
Framework
comments
91,658 GC
116,292 IC
Gold Standard
comments
396 GC
396 IC
Confusion comments:
• 72 GC (18%)
• 84 IC (21%)
• 4 GC and 4 IC discarded


12

Precision

		P	R	F
OneR	**GC**	**.875**	.194	.318
	IC	**.615**	.095	.165

Recall

		P	R	F
Multinomial Naive Bayes	**GC**	.209	**.944**	.342
	IC	.234	**.988**	.378

Precision and Recall

		P	R	F
JRip	**GC**	**.696**	**.542**	.609
Logistic	**IC**	**.434**	**.583**	.497

Precision

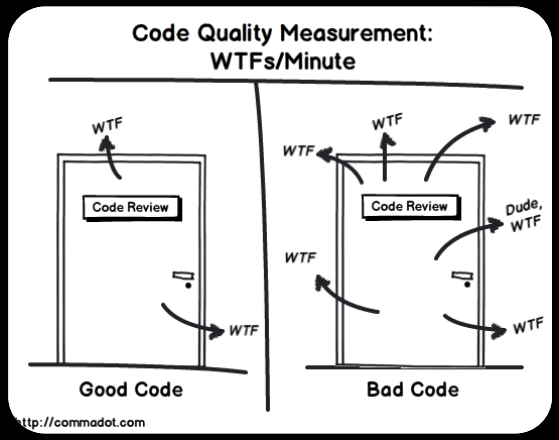
		P	R	F
OneR	**GC**	**.875**	.194	.318
	IC	**.615**	.095	.165

Recall

		P	R	F
Multinomial Naive Bayes	**GC**	.209	**.944**	.342
	IC	.234	**.988**	.378

Precision and Recall

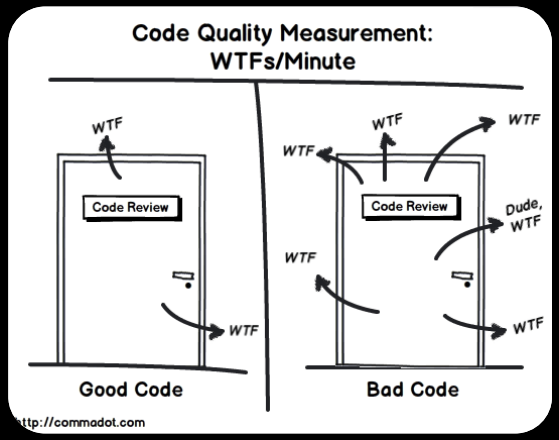
		P	R	F
JRip	**GC**	**.696**	**.542**	.609
Logistic	**IC**	**.434**	**.583**	.497

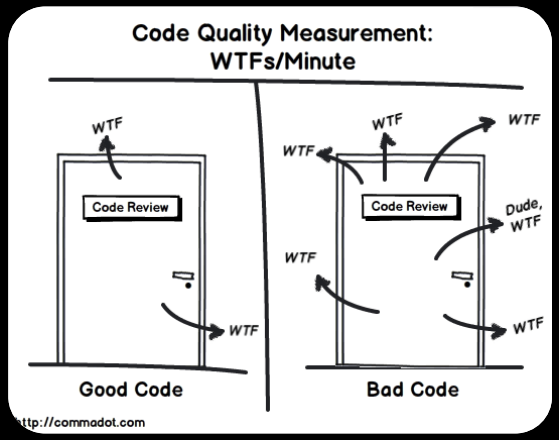

Code Quality Measurement:
WTFs/Minute
WTF
Code Review
WTF
Good Code
WTF
WTF
WTF
Code Review
Dude,
WTF
WTF
WTF
Bad Code
http://commadot.com

Inline comment

Do you really want a Java string here?
A ModifiedUTF8 one not enough?

confusion!





Inline comment

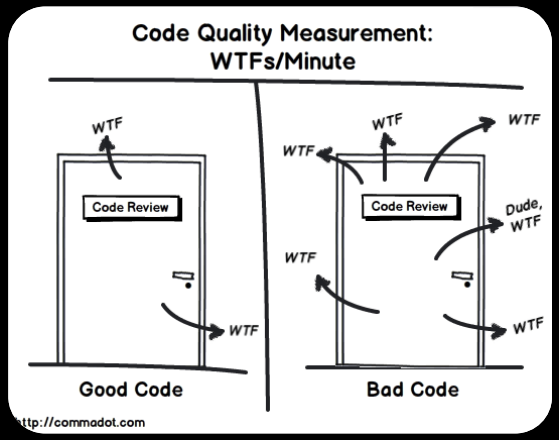
Do you really want a Java string here?
A ModifiedUTF8 one not enough?

confusion!

Inline comment

Maybe write a comment with the XML format here

no confusion!



Inline comment

Do you really want a Java string here?
A ModifiedUTF8 one not enough?

confusion!

Inline comment

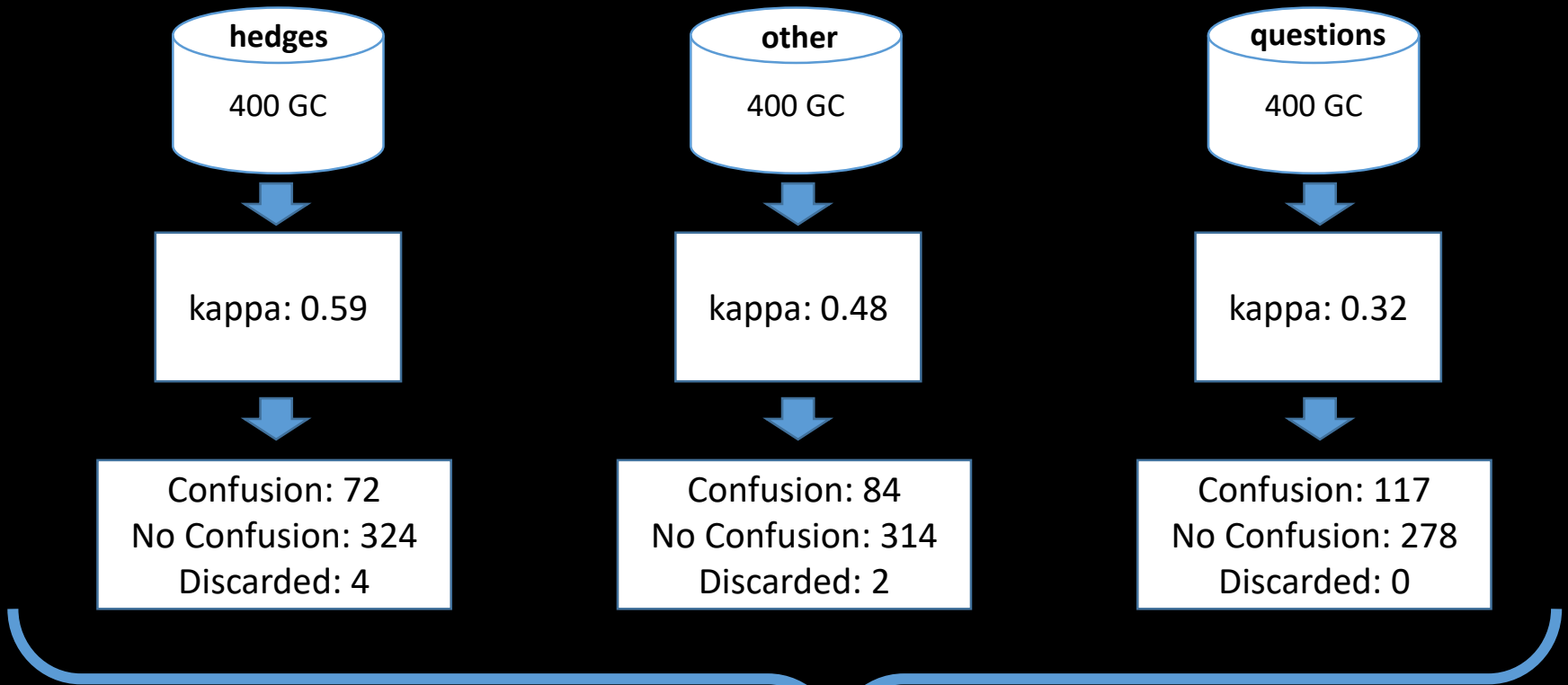
Maybe write a comment with the XML format here

no confusion!

Future work

- Other categories + new classifiers
- Statistical modeling
- Surveys

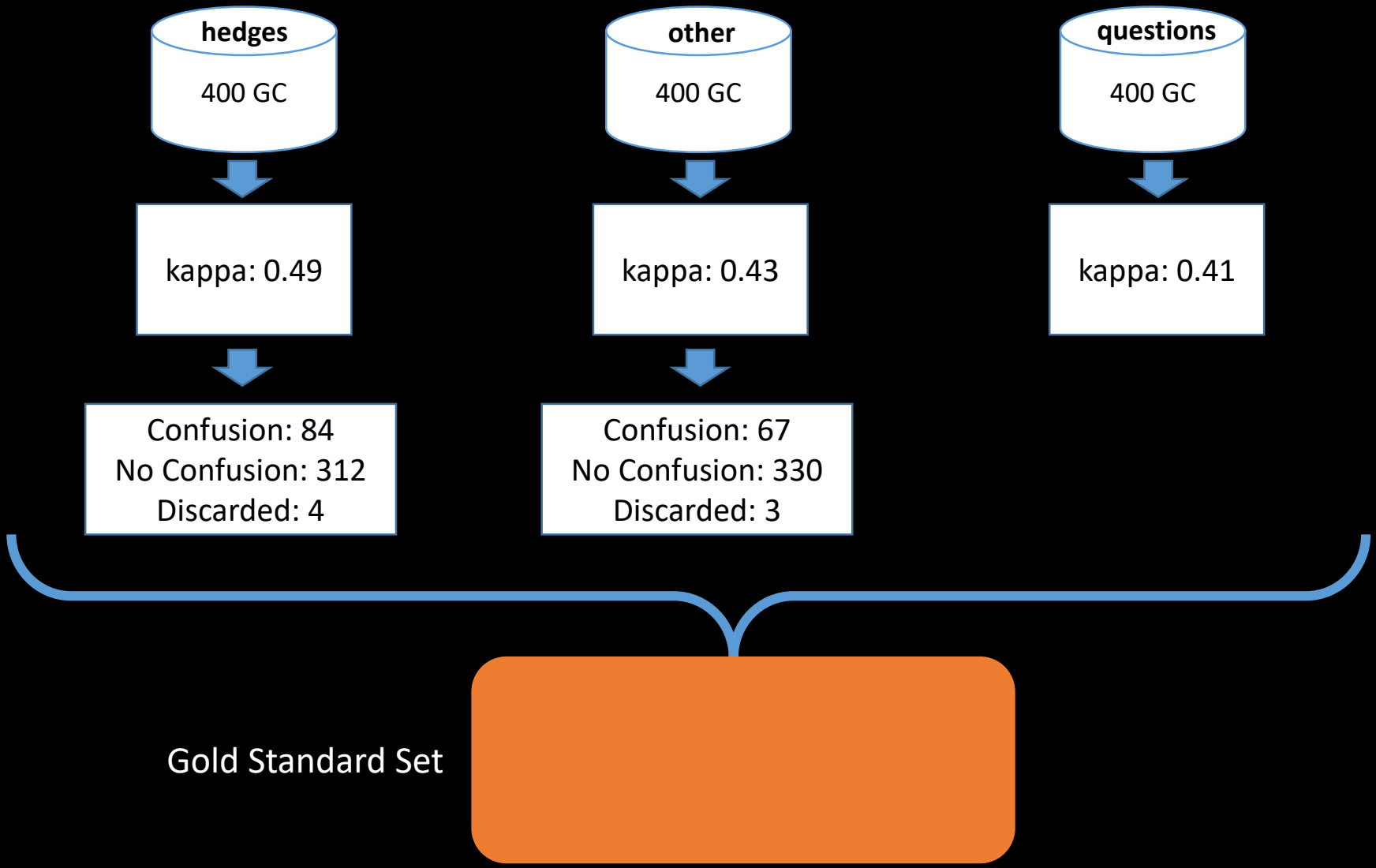
Manual Annotation - GC



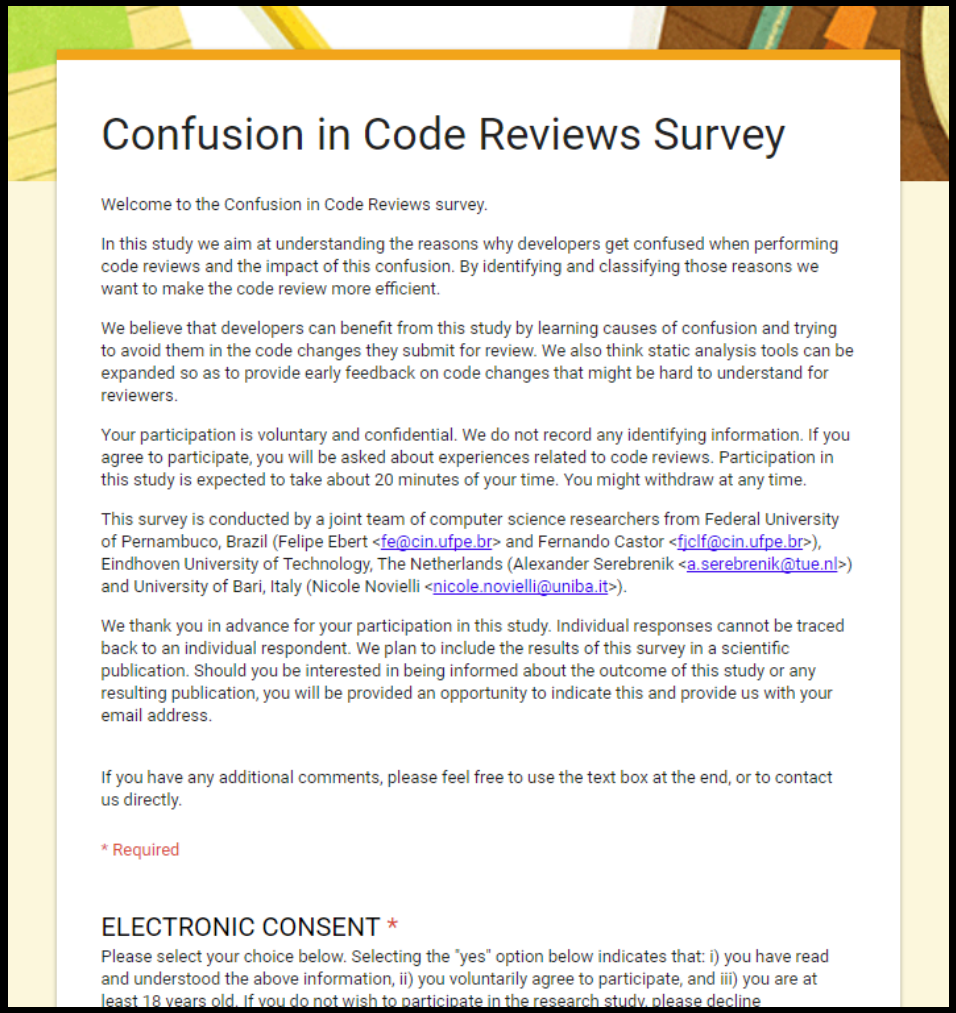
Gold Standard Set
(1,136 code reviews)

Confusion	273	23%
No Confusion	916	77%
Total	1,189	100%

Manual Annotation - IC



Survey



Confusion in Code Reviews Survey

Welcome to the Confusion in Code Reviews survey.

In this study we aim at understanding the reasons why developers get confused when performing code reviews and the impact of this confusion. By identifying and classifying those reasons we want to make the code review more efficient.

We believe that developers can benefit from this study by learning causes of confusion and trying to avoid them in the code changes they submit for review. We also think static analysis tools can be expanded so as to provide early feedback on code changes that might be hard to understand for reviewers.

Your participation is voluntary and confidential. We do not record any identifying information. If you agree to participate, you will be asked about experiences related to code reviews. Participation in this study is expected to take about 20 minutes of your time. You might withdraw at any time.

This survey is conducted by a joint team of computer science researchers from Federal University of Pernambuco, Brazil (Felipe Ebert <fe@cin.ufpe.br> and Fernando Castor <fjclf@cin.ufpe.br>), Eindhoven University of Technology, The Netherlands (Alexander Serebrenik <a.serebrenik@tue.nl>) and University of Bari, Italy (Nicole Novielli <nicole.novielli@uniba.it>).

We thank you in advance for your participation in this study. Individual responses cannot be traced back to an individual respondent. We plan to include the results of this survey in a scientific publication. Should you be interested in being informed about the outcome of this study or any resulting publication, you will be provided an opportunity to indicate this and provide us with your email address.

If you have any additional comments, please feel free to use the text box at the end, or to contact us directly.

* Required

ELECTRONIC CONSENT *

Please select your choice below. Selecting the "yes" option below indicates that: i) you have read and understood the above information, ii) you voluntarily agree to participate, and iii) you are at least 18 years old. If you do not wish to participate in the research study, please decline

- Emails sent: 4,645
  - Deliverable: 3,765
  - Undeliverable: 880
- Responses: 16 (0.4%)

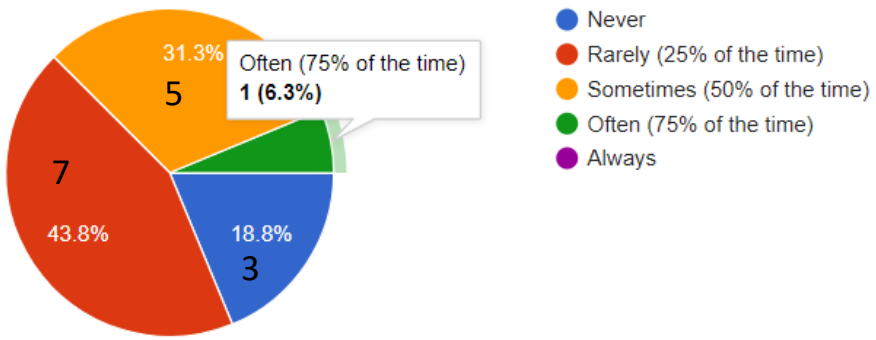
Survey

- How often did you feel confused
  - when reviewing code changes?
  - when your code has been reviewed?
- What usually makes you confused...?
- What is the impact of confusion…?
- What do you usually do to overcome confusion…?

When reviewing code changes

Developers might feel confused or think that they do not understand the code they review. How often did you feel this way when reviewing code changes?

16 responses



When reviewing code changes

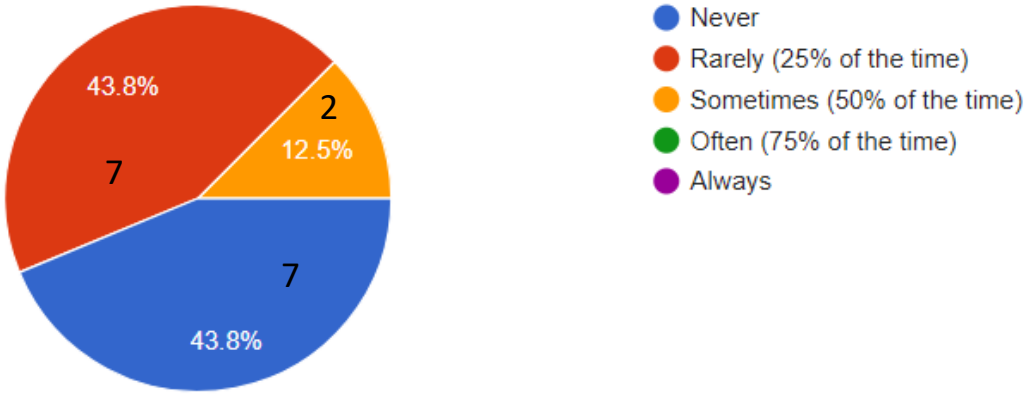
Developers might f
code they review
changes?

16 responses

When authoring code changes

Developers who authored code changes might feel confused or think that they do not understand something when their code is being reviewed. How often did you feel this way when your code has been reviewed?

16 responses

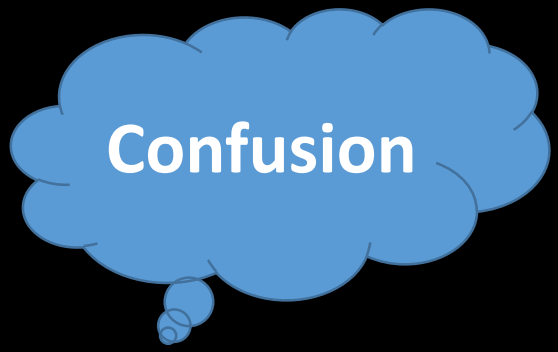


Ultimate Goal!

Patch size

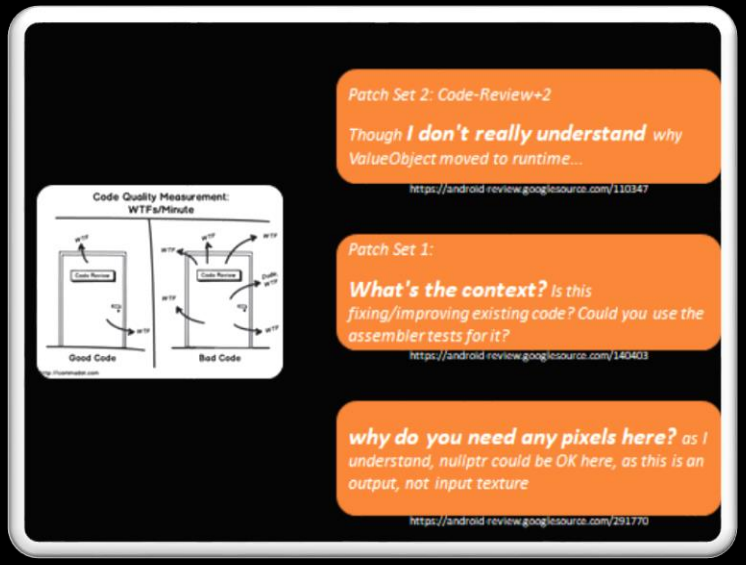
patch sets

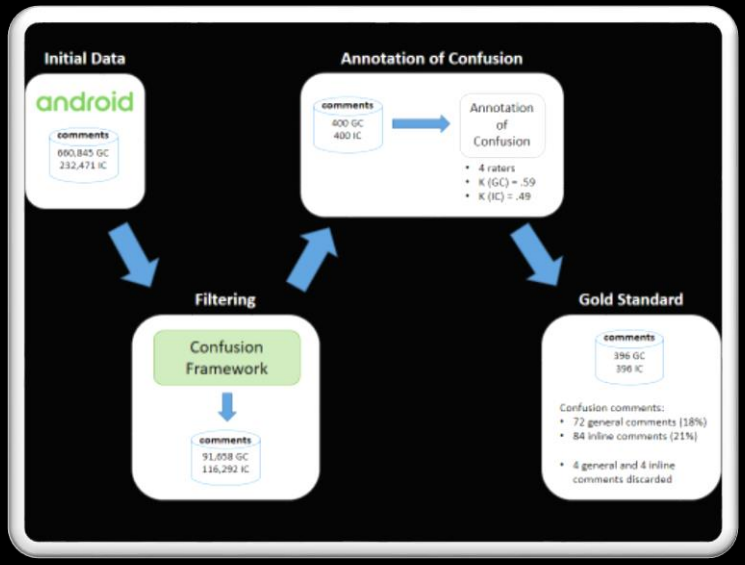
Reviewers experience

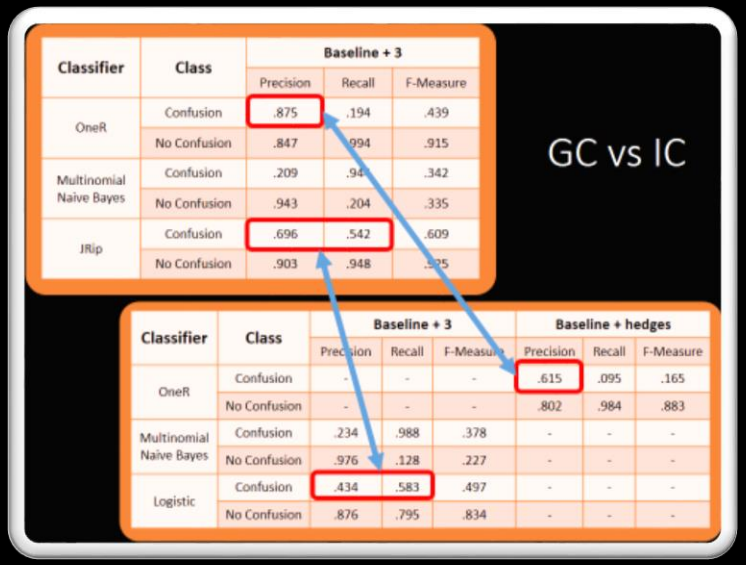


Code review

- Outcome
- Duration

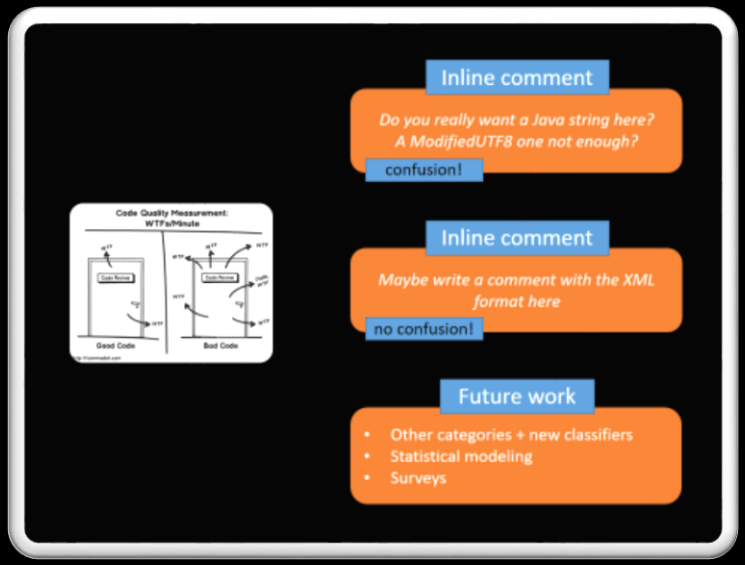






Classifier	Class	Baseline + 3		
		Precision	Recall	F-Measure
OneR	Confusion	.875	.194	.439
	No Confusion	.847	.994	.915
Multinomial Naive Bayes	Confusion	.209	.94	.342
	No Confusion	.943	.204	.335
JRip	Confusion	.696	.542	.609
	No Confusion	.903	.948	.925

Classifier	Class	Baseline + 3			Baseline + hedges		
		Precision	Recall	F-Measure	Precision	Recall	F-Measure
OneR	Confusion	-	-	-	.615	.095	.165
	No Confusion	-	-	-	.802	.984	.883
Multinomial Naive Bayes	Confusion	.234	.988	.378	-	-	-
	No Confusion	.976	.128	.227	-	-	-
Logistic	Confusion	.434	.583	.497	-	-	-
	No Confusion	.876	.795	.834	-	-	-



Felipe Ebert (fe@cin.ufpe.br), Fernando Castor (castor@cin.ufpe.br)
Nicole Novielli (nicole.novielli@uniba.it), Alexander Serebrenik (a.serebrenik @tue.nl)