

How Disabled Tests Manifest in Test Maintainability Challenges?

Dong Jae Kim, Bo Yang,
Jinqiu Yang, Tse-Hsun (Peter) Chen

ESEC/FSE 2021: Proceedings of the 29th ACM Joint Meeting on
European Software Engineering Conference and Symposium on the
Foundations of Software Engineering August 2021 Pages 1045–1055

発表者：松尾 和輝

背景

- ソフトウェアのテストは，ソフトウェアの品質を保つためには不可欠な作業である
- テストに失敗すると修正を行う必要があるが，修正が困難な場合があり，メンテナンス作業の軽減のためにテストを無効化することがある
- 無効化されたテストは時代遅れになり，技術的負債の原因となって，長期的なメンテナンスに悪影響を及ぼす可能性があるが，テストの無効化がどのように普及し，進化し，維持されているかはわかっていない

目的

- 開発者がどのくらいの頻度でテストを無効にしているのか（その頻度と推移），また開発者がなぜテストを無効にしているのか（テストを無効にしたり，再び有効にしたりする動機）を調査する

対象リポジトリ

- **GitHub**に登録されているスター数の多い上位**1000**のJavaリポジトリから，コードの規模，星の数，コミット数の合計が上位**10%**であるリポジトリを選択
- **2020**年に1回もコミットを行っていないリポジトリは削除し，残った**15**のリポジトリを対象とする
- 調査期間は**2015年1月**から**2020年1月**まで

Systems	Test LOC		Source LOC		Total # Commits
	2015	2020	2015	2020	
Camel	355K	533K	289K	607M	22,584
Cassandra	28K	78K	177K	216K	4,336
Cloudstack	51K	80K	1M	519K	4,698
Druid	22K	147K	64K	242K	4,181
Flink	82K	371K	105K	407K	13,997
Hadoop	349K	660K	463K	810K	13,668
Hbase	168K	287K	433K	409K	7,271
Hive	104K	269K	524K	1M	8,221
Ignite	162K	500K	253K	577K	15,397
Incubator-pinot	1.6K	75K	11K	196K	6,179
Kafka	2K	133K	11K	132K	5,845
Maven	14K	17K	44K	48K	660
Openfire	1.2K	5.2K	179K	94K	2,097
Orientdb	74K	188K	140K	360K	8,452
Storm	2.7K	35K	61K	240K	4,554
Total	1.4M	3.3M	3.8M	5.9M	122K

リサーチクエスチョン

- RQ1：テストの無効化はどの程度行われているか？
- RQ2：無効化されたテストの変更パターンは何か？
- RQ3：なぜ開発者はテストを無効にする行為を行うのか？

RQ1 -方法-

- RQ1：テストの無効化はどの程度行われているか？
- アプローチ
 - 開発者がクラスレベルとメソッドレベルの両方でテストを無効にする回数を計測する
 - 有効なテストと無効なテストを検出するために、既製のJavaParser [1] を使用する
 - 比較のためにRefactoringMiner[26]を用いてリファクタリング活動を検出する

[1] <https://javaparser.org/>

[26] Nikolaos Tsantalis, Ameya Ketkar, and Danny Dig. 2020. RefactoringMiner 2.0. IEEE Transactions on Software Engineering (2020), 21. <https://doi.org/10.1109/TSE.2020.3007722>

RQ1 -結果-

- 下図はテストの無効化の変更とテストコードのリファクタリングの頻度を比較したもの
- メソッドレベルでは両者は同等であり、クラスレベルではリファクタリングよりもテストの無効化の変更の方が多く行われている

	Method Level		Class Level		Total	Total
	Total	Per Commit	Total	Per Commit	Changes	Commits
Test Disabling Changes	2,581	2.7	530	0.6	3,111	949
Refactoring	2,495	1.9	120	0.1	2,615	1,334
Δ % Percentage	+3.4%	+42%	+341%	+500%	+19%	-29%

RQ1 -結果-

- 下図はテストの無効化に関連する3種類の変更（テストの無効化，テストの再有効化，無効化されたテストの削除）の頻度を示したもの
- テストコードに加えられた変更のうち，62.5%が無効化のための変更であった
- 再有効化されたテストのうち，42%はテストコードに変更が加えられている

	Disabling Tests	Re-enabling Tests		Deleting Disabled Tests
		Modified	Unmodified	
Method	2,581	314	486	762
Class	530	115	96	87
Total	3,111 (62.5%)	429 (8.6%)	582 (11.7%)	849 (17.1%)

RQ2 -方法-

- RQ2： 無効化されたテストの変更パターンは何か？
- アプローチ
 - コミット履歴を用いて，無効化されたテストがどのように変化するか，最終的にどのようなになっているかを調査する

RQ2 -結果-

- 下図は無効化されたテストの変更パターンを示したもの
- 41%は最終的に無効化されていた
- 解決したケースでは、28%が削除され、30%が再有効化されている

Change Pattern	Frequency
<i>Change patterns for unresolved tests</i>	
DISABLED	1,229
DISABLED → RE-ENABLED → DISABLED	21
DISABLED → DELETED → DISABLED	1
Total	1,251
<i>Change patterns for resolved tests</i>	
DISABLED → RE-ENABLED	871
DISABLED → DELETED	824
DISABLED → RE-ENABLED → DISABLED → RE-ENABLED	46
DISABLED → RE-ENABLED → DISABLED → DELETED	24
DISABLED → RE-ENABLED → DISABLED → RE-ENABLED →	1
DISABLED → RE-ENABLED	
Total	1,766

RQ3 -方法-

- RQ3: なぜ開発者はテスト無効化の実践を利用するのか？
- アプローチ
 - 層化サンプリング手法を用いて各リポジトリをサンプリングする
 - 著者2名が、コミットメッセージ、テストコード、テストコード周辺のコメント、バグレポートなどの関連するソフトウェア成果物を手作業で調査し、最初の分類を導き出す
 - 著者2名は導き出された理由を統一し、進化パターンごとに割り当てられた理由を比較する
 - 意見の相違があった場合は、合意に達するまで議論する

RQ3 -結果-

- テストの失敗を隠すため (40%)
- 機能メンテナンスやリファクタリングを行うため (23%)
- 手動テストへの移行のため (11.2%)
- 依存関係の問題 (6.2%)
- テスト設計の問題 (5%)
- その他 (12.5%)

RQ3 -結果-

Table 7: Qualitative analysis result: a taxonomy of why developers disable tests.

Categories	Motivation	#Frequency
- Hiding Test Failures		133 (40.6%)
Disabling tests while working on bug fixes	Developers temporarily disable failing tests while working on fixing the bugs.	81
Flaky test	Developers may disable flaky tests to avoid occasional failures.	42
Won't-fix bugs in code/test	Developers disable tests to avoid failures as they will not work on fixing the failures due to difficulty.	7
Slow test	Test failures due to long running time, so developers disable such tests to avoid failures.	2
Library incompatibility	JUnit annotation (@Parameterized) was not supported by Ant in the used CI platform. Thus, the test was disabled.	1
- Precautions During Feature Maintenance		78 (23.8%)
New/Improved Features	In the process of implementing new or improving existing features, developers may temporarily disable relevant tests or introduce new tests that are disabled (e.g., commented-out tests) to avoid potential failures.	62
Deprecation	Developers may disable relevant tests in the process of deprecating features.	12
Refactoring	Developers may disable tests during refactoring.	4
- Diverting to Manual Testing		38 (11.6%)
Require manual input	Developers need to manually run tests that need to be manually configured (e.g., database setup and secret keys).	24
Expensive Test	Developers need to manually run tests that are expensive to run and may not have to be run all the time (e.g., performance and migration test).	13
Experimentation	Developers manually run tests that are under experimentation.	1
- Dependency Issue		21 (6.4%)
Difficulties in maintaining external dependencies	External dependency is hard to maintain due to various reasons. Developers may need to disable tests when there are bugs in the external dependency, unexpected version changes, or dependencies are hard to integrate or use.	17
Waiting for functionality update in external dependency	Developers disable tests while waiting for feature improvement in the external dependencies.	4
- Test Design Issues		16 (4.9%)
Selective test inheritance	Developers disable tests to disable unneeded inherited tests, while selectively reusing some inherited tests.	14
Redundant Test	Developers disable tests that are redundant and covered by a different test class.	2
- Other Reasons		41 (12.5%)
Unknown	Lacks of explicit mention of why tests are disabled (e.g., no relevant Jira issues and comments).	32
Obsolescence	Developers disable obsolete tests.	5
disable by mistake	Developers accidentally disable the test.	4

まとめ

- 開発者がテストを無効にする頻度，理由を調査
- 15個のリポジトリからテストコードを抽出，コミット履歴からテストコードの変更を特定
- テストのリファクタリングと同じくらい，テストの無効化が行われている
- 無効化されたテストの半数は，最終的には無効化されたままである
- 無効化する理由は様々だが，特に多いのはテストの失敗を隠すため

所感

- テストを無効化する変更が**62.5%**と、多く感じた
 - テストの失敗を隠すために無効化する人が多いみたいだけど、テストに失敗したままリリースしてもいいのか
- テスト作業の大変さを感じた
 - テストの無効化が多いということはそれだけバグが多いということ？