

COMPSUITE: A Dataset of Java Library Upgrade Incompatibility Issues

Xiufeng Xu
xiufeng001@e.ntu.edu.sg
Nanyang Technological University
Singapore

Chenguang Zhu
cgzhu@utexas.edu
The University of Texas at Austin
USA

Yi Li
yi_li@ntu.edu.sg
Nanyang Technological University
Singapore

ABSTRACT

Modern software systems heavily rely on external libraries developed by third-parties to ensure efficient development. However, frequent library upgrades can lead to compatibility issues between the libraries and their client systems. In this paper, we introduce COMPSUITE, a dataset that includes 123 real-world Java client-library pairs where upgrading the library causes an incompatibility issue in the corresponding client. Each incompatibility issue in COMPSUITE is associated with a test case authored by the developers, which can be used to reproduce the issue. The dataset also provides a command-line interface that simplifies the execution and validation of each issue. With this infrastructure, users can perform an inspection of any incompatibility issue with the push of a button, or reproduce an issue step-by-step for a more detailed investigation. We make COMPSUITE publicly available to promote open science. We believe that various software analysis techniques, such as compatibility checking, debugging, and regression test selection, can benefit from COMPSUITE.

KEYWORDS

Incompatibility issue, software libraries, dataset

1 INTRODUCTION

Modern software systems are becoming increasingly complex due to the need for integrating various components developed by different teams or organizations. These components are often subject to continuous evolution, and as a result, ensuring that new upgrades to third-party libraries do not cause any compatibility issues with the existing software system is a challenging task. The complexity of these systems and the number of dependencies involved make it difficult to anticipate and identify incompatibilities that may arise from updates to external components. Incompatibility issues resulting from upgrades to external components can compromise the reliability of software systems, potentially leading to significant financial losses for the organizations that rely on these systems.

Many techniques have been proposed to address third-party library compatibility issues, including regression testing [19, 20], static analysis [3], dependency conflict detection [28], and client-specific compatibility checking [21, 35]. These techniques address library compatibility issues in different dimensions and have been evaluated with their own isolated datasets.

An excellent dataset has the potential to serve as a valuable reference for future research in this field. However, composing the dataset requires intricate manual validation, e.g., confirming whether the cause of a test failure is due to runtime exception, assertion violation, or other reasons. Therefore, we propose COMPSUITE, the first incompatibility issue dataset focusing on library behavioral

incompatibility with concrete reproducible test cases. Each test case is isolated and validated, enabling the direct manifestation of the incompatibilities.

COMPSUITE comprises 123 real-world Java client-library pairs such that upgrading any library results in incompatibility issues for the corresponding client. Every incompatibility issue in COMPSUITE contains a test case created by developers, allowing for the reproduction of the issue. On top of this dataset, we also developed an automated command-line interface, which streamlines all processes of the reproduction, such as downloading and compiling a projects, running target tests and re-running the tests after a library upgrade. With this infrastructure, users may reproduce an incompatibility issue programmatically with minimal efforts.

Contribution. To summarize, we make the following contributions in this paper:

- (1) We construct a dataset, COMPSUITE, including 123 reproducible, real-world client-library pairs that manifest incompatibility issues when upgrading the library. These data points originate from 88 clients and 104 libraries.
- (2) We created an automated command-line interface for the dataset. With this interface, users are able to programmatically replicate an incompatibility issue from the dataset with a single command. The interface also offers separate commands for each step involved in the reproduction of incompatibility issues.

We envision that COMPSUITE to be used to evaluate various program analysis techniques, including compatibility checking, module-level regression testing selection, and debugging techniques. More detailed information can be found in Section 4.

The dataset and tool are available at: <https://github.com/compsuite-team/compsuite>.

2 DATASET CREATION

In this section, we outline the methodology and process employed to create the COMPSUITE dataset.

2.1 Subjects Selection

To ensure the representativeness and reproducibility of the COMPSUITE dataset, we focus on including high-quality and popular client projects and libraries. The selection of client projects was sourced from GitHub [4], a widely recognized online community for hosting open-source codebases. To ensure the inclusion of the most popular projects, we systematically sorted all the available projects in descending order based on their number of stars on GitHub and selected the target clients from the top of the list. The selection of libraries was sourced from Maven Central [17], which hosts 33.5M of Java libraries and their associated binaries, making it a widely used repository of libraries for Java API and library

Table 1: Details of clients and libraries included in COMPSUITE.

Client	#LoC	#Star	Library	#Maven Usage
retrofit	29.7K	41.5K	org.slf4j:slf4j-api	62.5K
apollo	61.3K	28K	com.google.guava:guava	34.4K
druid	441.9K	26.8K	org.scala-lang:scala-library	34K
webmagic	17.4K	10.8K	com.fasterxml.jackson.core:jackson-databind	25.8K
language-tool	171.2K	8.5K	ch.qos.logback:logback-classic	25.5K
Other 83 clients (mean)	371.6K	1.3K	Other 99 libraries (mean)	3.2K
All clients (mean)	358.7K	2.5K	All libraries (mean)	4.8K

research [13, 22, 24, 32]. We include a library in the dataset only if it has more than 100 usages (i.e., clients) on Maven Central. Our selection criteria aimed to ensure the inclusion of popular and widely used client projects and libraries in the dataset, thereby maximizing its relevance and usefulness to the research community.

Among the highly-rated client projects, our selection criteria focused on those that use Maven [16] as their build systems, given its widespread adoption and maturity. Maven provides a standardized approach to managing Java projects and their dependencies, where each library dependency in a Maven client project is represented as an item in a `pom.xml` file, making it easy to identify and edit library versions programmatically. Furthermore, Maven offers built-in functionality for running unit tests and generating test reports, which simplifies the identification and diagnosis of incompatibility issues arising from test executions. Since Maven projects typically rely on Maven Central as their centralized repository for hosting and downloading libraries, the process of obtaining and managing libraries in our dataset is simplified.

Table 1 presents the top 5 client projects and libraries in the COMPSUITE dataset, ranked by popularity. For each client project, we provide information on its lines of code (LoC) and the number of stars it has received on GitHub, while for each library, we include its number of usages by other projects from Maven Central.

In total, COMPSUITE comprises 123 incompatible client-library pairs. These pairs encompass 88 distinct clients and 104 libraries altogether. On average, the affected clients have 2.5K stars on GitHub and 358.7K lines of code, while incompatible libraries have 4.8K usages on Maven Central. Thus, we believe that the incompatibility issues present in the COMPSUITE dataset have a significant impact on a large number of codebases and can affect many users of the libraries, either directly or indirectly.

To ensure that all client projects in the dataset are executable and the runs are reproducible, we performed a series of checks on each project. First, we checked out the project to the version (SHA) at the time of the dataset creation, which we refer to as the *base version*. Next, we ran the standard Maven project compilation command to verify if the project compiles successfully. If the project fails to compile, we excluded it from the dataset. Subsequently, we ran the standard Maven test command to execute all the tests in the project, ensuring that all tests pass on the base version. We excluded any project that fails to pass tests at this stage. Finally, we only included the client projects that successfully compile and pass all tests on the base version, thereby ensuring that the dataset is only consist of projects which can be executed and whose executions can be reproduced.

2.2 Data Collection

We collected the data following the below procedures. Figure 1 visualizes the overall architecture of COMPSUITE. In the upper left portion of Fig. 1, we illustrate the approach taken by COMPSUITE to identify incompatibilities between a client project and its dependent libraries. Specifically, for each client project on its base version, we upgraded each of its dependent libraries and tested if the upgrade caused any test failures. Our intuition behind this approach is that since all the tests in the client passed on the base version, if upgrading any library causes a test failure, that library upgrade must have introduced incompatibility issues. We refer to the test that flips from passing to failing as an *incompatibility-revealing test*.

To automatically upgrade the libraries and run the tests, we utilized the Maven Versions Plugin [18]. For a given client project, we scanned its dependency list using this plugin to identify all the libraries that had newer versions available on Maven Central. If a library had a newer version, we marked it as upgradable. Next, for each upgradable library, we used the plugin to upgrade it by updating the `pom.xml` file to the most recent version on Maven Central. We then re-executed the test suite of the client. If any tests failed during this run, we marked the client-library pair as having an incompatibility issue and marked the test as an incompatibility-revealing test of this issue. It is crucial to note that we only upgraded one library at a time to isolate failures caused by different libraries. To ensure the accuracy and dependability of the dataset, we carried out a manual verification process for each identified incompatibility issue. In particular, we carefully examined the test failure messages and reports to confirm that they were indeed caused by the upgraded library. For each incompatible client-library pair, we selected a single incompatibility-revealing test to be included in the final dataset. In cases where a client-library pair had multiple incompatibility issues, we chose the one that we deemed most representative and easy to comprehend.

Finally, we persisted the metadata of all the selected incompatibility issues in a collection of json files. Figure 2 presents the metadata of an incompatibility issue in the COMPSUITE dataset. The data schema includes the ID of the issue, client project name, SHA of the client base version, URL of the client project, library name, versions of the old and new libraries, the name of the incompatibility-revealing test, the submodule containing the incompatibility-revealing test, and the command to run the test. The majority of the information is self-explanatory. However, it is worth noting that the old version of the library is the one utilized at the base version of the client, while the new version is the most recent version found on Maven Central that triggers the incompatibility when upgrading, as described in Section 2.2.

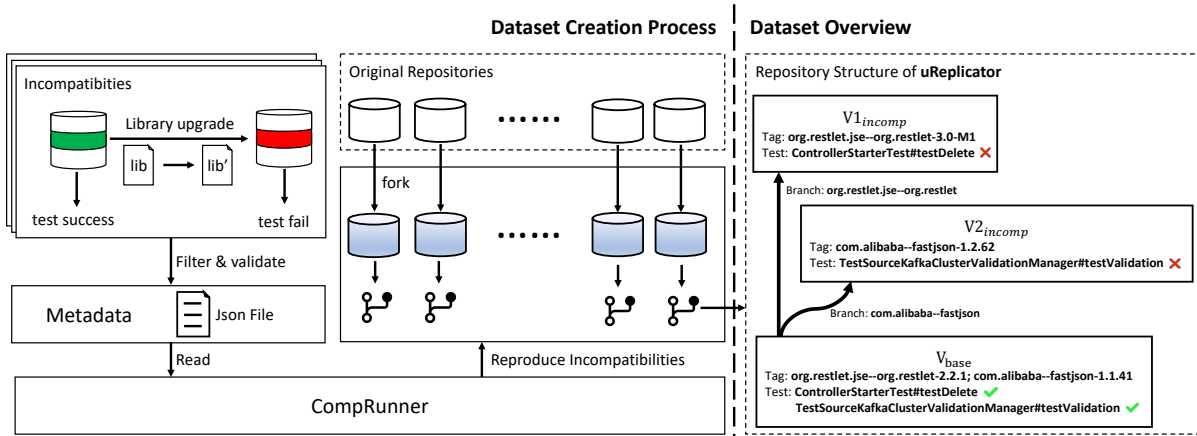


Figure 1: The architecture of COMPSUITE.

```

1 {
2   "id": "i-49",
3   "client": "wasabi",
4   "sha": "9f2aa5f92e49c3844d787320e2d22e15317aa8e2",
5   "url": "https://github.com/intuit/wasabi",
6   "lib": "org.apache.httpcomponents:httpclient",
7   "old": "4.5.1",
8   "new": "4.5.10",
9   "test":
10  ↪ "DefaultRestEndPointTest#testGetRestEndPointURI",
11  ↪ "org.apache.maven.plugins:maven-surefire-plugin:2.20:test",
12  ↪ "-fn -Drat.ignoreErrors=true -DtrimStackTrace=false",
13  ↪ "-Dtest=DefaultRestEndPointTest#testGetRestEndPointURI"
14 }
    
```

Figure 2: The data schema of COMPSUITE

3 DATASET USAGE

In this section, we provide instructions on the usage of our dataset.

3.1 Exploring an Incompatibility Issue

To ensure the reproducibility of incompatibility issues and to facilitate the demonstration of such issues, we have annotated checkpoints in the version histories of the client projects and provided tags that guide users to explore any incompatibility issues present in the COMPSUITE dataset.

As illustrated on the right-hand side of Fig. 1, our approach to handling incompatible client-library pairs involved creating a fork [5] of the original client project for each identified pair, while preserving all code and version history information. To mark the base version of the project, we utilized the `git tag` [6] command, designating it as V_{base} . Subsequently, we developed a patch to upgrade the library from its old version to its new version, a simple process that can be accomplished with a single line change in the `pom.xml` file for Maven projects. This patch was then applied to the V_{base} version, resulting in a new version that we identified as V_{incomp} . Notably, the only difference between V_{base} and V_{incomp}

lies in the library version used: the old (compatible) version is utilized on V_{base} while the new (incompatible) version is utilized on V_{incomp} . For instance, in Fig. 1, the client project employs version 2.2.1 of the `org.restlet.jse--org.restlet` library on its V_{base} and version 3.0-M1 on its V_{incomp} . In cases where multiple libraries exhibit incompatibility issues in the client project, we not only create different branches for each library with its name, but also generate a V_{incomp} version tag for each, with accompanying annotations that denote the corresponding library name and version, as depicted in Fig. 1.

The V_{incomp} tag for each client-library pair also specifies the specific test that can reveal the incompatibility issue during its run. Following Maven’s convention, the test name is formatted as `TestClassName#testMethodName`. By simply copying the text from the tag, users can easily run the incompatibility-revealing test on the V_{incomp} version and observe the incompatibility issue. On the V_{base} version, all tests should pass. This design aims to simplify the usage of COMPSUITE and make it more accessible and user-friendly.

Using the forked client repositories and version tags provided in the COMPSUITE dataset, users can easily reproduce any incompatibility issue by checking out to V_{incomp} and running the corresponding incompatibility-revealing test. To compare the behaviors of the client with compatible and incompatible library versions, users can run the incompatibility-revealing test on both V_{base} and V_{incomp} and compare the test outcomes. This allows for a clear understanding of the impact of the library upgrade on the client behaviors.

3.2 COMPRUNNER: An Automated Tool for Reproducing Incompatibility Issues

We further developed an automated tool, named `COMPRUNNER`, which is a part of COMPSUITE. With `COMPRUNNER`, users can easily reproduce and investigate any incompatibility issue in a one-click manner by providing the issue ID as input.

We offer an option which enables users to reproduce an incompatibility issue end-to-end with a single command as is shown

below. The command outputs and saves all intermediate results and logs for future reference.

```
1 python main.py --incompat i-56
```

When `COMPRUNNER` runs, it clones the client project from our forked code repository and saves it in the output directory (which is configurable). Then, it checks out to the base version, compiles the code, and runs the incompatibility-revealing test. Next, it upgrades the library to the new version, reruns the incompatibility-revealing test, and reports any failure information to the user.

We also provide a set of commands that break down the entire cycle of incompatibility exploration into separate steps:

```
1 python main.py --download i-56
2 python main.py --compile i-56
3 python main.py --testold i-56
4 python main.py --testnew i-56
```

We provide several other `COMPRUNNER` commands for users to inspect different aspects of the incompatibility issues from the `COMPSUITE` dataset. A complete list of these commands can be found on `COMPSUITE`'s website at <https://github.com/compsuite-team/compsuite>.

4 APPLICATION SCENARIOS

We anticipate that both researchers and practitioners can benefit from `COMPSUITE` to facilitate their investigations and research on errors and test failures induced by library upgrades. `COMPSUITE` supports the evaluation of various program analysis techniques, such as software upgrade compatibility checking, debugging, and module-level regression test selection techniques.

As an overview, authors of compatibility checkers and detectors may use `COMPSUITE` as a benchmark to evaluate the performance of their techniques against other baseline approaches. Furthermore, authors of debugging techniques can utilize `COMPSUITE` as a dataset of compatibility bugs, where each bug corresponds to a test case that verifies the existence or absence of the bug. Finally, authors of module-level regression test selection techniques can use `COMPSUITE` to assess the safety of their approaches. A safe module-level RTS technique should select all the corresponding incompatibility-revealing test cases when the library changes.

We detail the three usage scenarios as follows.

- **Compatibility Checkers and Detectors.** The existing techniques for compatibility checking and detection in Java can be categorized into three groups: i) Techniques for detecting API incompatibility that focus on detecting API-breaking changes, such as renaming of code entities and changes in parameter types [9, 10, 15, 25, 30]. ii) Techniques for detecting behavioral incompatibility that focus on identifying behavioral differences that cause test failures when a library is upgraded in a client, such as changes in program states [21, 37]. iii) Techniques for detecting dependency conflicts [26–28], which aim to identify library APIs that exhibit inconsistent semantics between libraries due to class path shading. We believe that developers of techniques in the first two categories can use `COMPSUITE` as a benchmark to evaluate their tools' performance, such as precision and recall. They can run their tools on the `COMPSUITE` dataset and compare the results with the incompatibility issues present in

the dataset. On the other hand, developers of techniques for detecting dependency conflicts can slightly modify `COMPSUITE`'s dataset by placing both old and new libraries on the class path, running library conflict detection, and checking if the issues can be detected.

- **Module-Level Regression Test Selection.** Regression test selection (RTS) is a technique that aims to reduce the cost of regression testing by selecting a subset of tests that may change the behavior due to code changes on each program version [7, 14, 34, 35]. Module-level RTS focuses on selecting the affected client tests when a dependent library is updated [8]. The developers of module-level RTS techniques can evaluate the safety of their tools using `COMPSUITE`. For each client-library pair in `COMPSUITE`, a module-level RTS tool should select all the corresponding incompatibility-revealing tests when upgrading the library from the old version to the new version.
- **Debugging.** The existing debugging techniques for Java, include symbolic execution [1], delta debugging [33], fault localization [31], etc. These techniques aim to identify the root cause of errors or failures in software. Developers of debugging techniques can use `COMPSUITE` as a dataset of compatibility bugs, where each compatibility bug corresponds to a test case that checks the presence or absence of the bug. They can use `COMPSUITE` to evaluate their techniques' ability to perform root cause analysis by trying to identify the corresponding library change that caused the compatibility issue.

5 RELATED WORK

To cater to the requirements of various research endeavors, numerous outstanding datasets have been made available to date. Just et al. [12] introduced `Defects4J`, a database supplies actual bugs, fixed program versions, and corresponding test suites. Bui et al. [2] introduced `Vul4J` focusing on Java vulnerabilities. Jezek et al. [11] released their dataset of compatibility issues arising from program evolution. There are also many datasets cater for other research domains and ecosystems [23, 29, 36].

Distinct from the previously discussed datasets, `COMPSUITE` is the first dataset emphasizes the incompatibility issues caused by Java library behavior changes. This type of issues are prevalent and difficult to detect. Additionally, Our developed automated tools also have the capability to assist researchers in swiftly reproducing issues. We believe that a dataset targeting the library upgrade incompatibility issue will contribute to the advancement of the associated technologies.

6 CONCLUSION

This paper presents `COMPSUITE`, a dataset containing 123 real-world Java client-library pairs where library upgrades cause compatibility issues in the corresponding clients. On top of it, we also developed a command-line interface, `COMPRUNNER`, which allows users to quickly check incompatibility issues with a single command or reproduce an incompatibility programmatically for in-depth analysis. We believe that various program analysis techniques, such as library compatibility checking, debugging, and regression test selection, may benefit from our dataset.

REFERENCES

- [1] Roberto Baldoni, Emilio Coppa, Daniele Cono D’elia, Camil Demetrescu, and Irene Finocchi. 2018. A survey of symbolic execution techniques. *ACM Computing Surveys (CSUR)* 51, 3 (2018), 1–39.
- [2] Quang-Cuong Bui, Riccardo Scandariato, and Nicolás E Díaz Ferreyra. 2022. Vul4J: a dataset of reproducible Java vulnerabilities geared towards the study of program repair techniques. In *Proceedings of the 19th International Conference on Mining Software Repositories*. 464–468.
- [3] Darius Foo, Hendy Chua, Jason Yeo, Ming Yi Ang, and Asankhaya Sharma. 2018. Efficient static checking of library updates. In *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. 791–796.
- [4] Github. 2023. *GitHub*. <https://github.com/>.
- [5] Github. 2023. *GitHub Fork*. <https://docs.github.com/en/get-started/quickstart/fork-a-repo>.
- [6] Github. 2023. *GitHub Tag*. <https://docs.github.com/en/desktop/contributing-and-collaborating-using-github-desktop/managing-commits/managing-tags>.
- [7] Milos Gligoric, Lamyaa Eloussi, and Darko Marinov. 2015. Practical regression test selection with dynamic file dependencies. In *Proceedings of the 2015 International Symposium on Software Testing and Analysis*. 211–222.
- [8] Alex Gyori, Owolabi Legunsen, Farah Hariri, and Darko Marinov. 2018. Evaluating regression test selection opportunities in a very large open-source ecosystem. In *2018 IEEE 29th International Symposium on Software Reliability Engineering (ISSRE)*. IEEE, 112–122.
- [9] Dongjie He, Lian Li, Lei Wang, Hengjie Zheng, Guangwei Li, and Jingling Xue. 2018. Understanding and detecting evolution-induced compatibility issues in android apps. In *Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering*. 167–177.
- [10] Huaxun Huang, Lili Wei, Yepang Liu, and Shing-Chi Cheung. 2018. Understanding and detecting callback compatibility issues for android applications. In *Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering*. 532–542.
- [11] Kamil Jezek and Jens Dietrich. 2017. API Evolution and Compatibility: A Data Corpus and Tool Evaluation. *J. Object Technol.* 16, 4 (2017), 2–1.
- [12] René Just, Darioush Jalali, and Michael D Ernst. 2014. Defects4J: A database of existing faults to enable controlled testing studies for Java programs. In *Proceedings of the 2014 international symposium on software testing and analysis*. 437–440.
- [13] Raula Gaikovina Kula, Coen De Roover, Daniel German, Takashi Ishio, and Katsuro Inoue. 2014. Visualizing the evolution of systems and their library dependencies. In *2014 Second IEEE Working Conference on Software Visualization*. IEEE, 127–136.
- [14] Owolabi Legunsen, Farah Hariri, August Shi, Yafeng Lu, Lingming Zhang, and Darko Marinov. 2016. An extensive study of static regression test selection in modern software evolution. In *Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering*. 583–594.
- [15] Li Li, Tegawendé F Bissyandé, Haoyu Wang, and Jacques Klein. 2018. Cid: Automating the detection of api-related compatibility issues in android apps. In *Proceedings of the 27th ACM SIGSOFT International Symposium on Software Testing and Analysis*. 153–163.
- [16] Maven. 2023. *Apache Maven*. <https://maven.apache.org/>.
- [17] Maven. 2023. *Maven Central Repositories*. <https://mvnrepository.com/>.
- [18] Maven. 2023. *Maven Versions Plugin*. <https://maven.apache.org/plugins/index.html>.
- [19] Gianluca Mezzetti, Anders Møller, and Martin Toldam Torp. 2018. Type regression testing to detect breaking changes in Node.js libraries. In *32nd european conference on object-oriented programming (ECOOP 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
- [20] Anders Møller and Martin Toldam Torp. 2019. Model-based testing of breaking changes in Node.js libraries. In *Proceedings of the 2019 27th ACM joint meeting on european software engineering conference and symposium on the foundations of software engineering*. 409–419.
- [21] Federico Mora, Yi Li, Julia Rubin, and Marsha Chechik. 2018. Client-specific equivalence checking. In *Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering*. 441–451.
- [22] Shaikh Mostafa, Rodney Rodriguez, and Xiaoyin Wang. 2017. A study on behavioral backward incompatibility bugs in Java software libraries. In *2017 IEEE/ACM 39th International Conference on Software Engineering Companion (ICSE-C)*. IEEE, 127–129.
- [23] Sebastian Nielebock, Paul Blockhaus, Jacob Krüger, and Frank Ortmeier. 2021. AndroidCompass: A Dataset of Android Compatibility Checks in Code Repositories. In *2021 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR)*. IEEE, 535–539.
- [24] Dong Qiu, Bixin Li, and Hareton Leung. 2016. Understanding the API usage in Java. *Information and software technology* 73 (2016), 81–100.
- [25] Simone Scalabrino, Gabriele Bavota, Mario Linares-Vásquez, Michele Lanza, and Rocco Oliveto. 2019. Data-driven solutions to detect api compatibility issues in android: an empirical study. In *2019 IEEE/ACM 16th International Conference on Mining Software Repositories (MSR)*. IEEE, 288–298.
- [26] Ying Wang, Ming Wen, Zhenwei Liu, Rongxin Wu, Rui Wang, Bo Yang, Hai Yu, Zhiliang Zhu, and Shing-Chi Cheung. 2018. Do the dependency conflicts in my project matter?. In *Proceedings of the 2018 26th ACM joint meeting on european software engineering conference and symposium on the foundations of software engineering*. 319–330.
- [27] Ying Wang, Ming Wen, Rongxin Wu, Zhenwei Liu, Shin Hwei Tan, Zhiliang Zhu, Hai Yu, and Shing-Chi Cheung. 2019. Could i have a stack trace to examine the dependency conflict issue?. In *2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE)*. IEEE, 572–583.
- [28] Ying Wang, Rongxin Wu, Chao Wang, Ming Wen, Yepang Liu, Shing-Chi Cheung, Hai Yu, Chang Xu, and Zhiliang Zhu. 2021. Will Dependency Conflicts Affect My Program’s Semantics? *IEEE Transactions on Software Engineering* 48, 7 (2021), 2295–2316.
- [29] Lili Wei, Yepang Liu, and Shing-Chi Cheung. 2016. Taming android fragmentation: Characterizing and detecting compatibility issues for android apps. In *Proceedings of the 31st IEEE/ACM International Conference on Automated Software Engineering*. 226–237.
- [30] Lili Wei, Yepang Liu, and Shing-Chi Cheung. 2019. Pivot: learning api-device correlations to facilitate android compatibility issue detection. In *2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE)*. IEEE, 878–888.
- [31] W Eric Wong, Ruizhi Gao, Yihao Li, Rui Abreu, and Franz Wotawa. 2016. A survey on software fault localization. *IEEE Transactions on Software Engineering* 42, 8 (2016), 707–740.
- [32] Wei Wu, Foutse Khomh, Bram Adams, Yann-Gaël Guéhéneuc, and Giuliano Antoniol. 2016. An exploratory study of api changes and usages based on apache and eclipse ecosystems. *Empirical Software Engineering* 21 (2016), 2366–2412.
- [33] Andreas Zeller. 1999. Yesterday, my program worked. Today, it does not. Why? *ACM SIGSOFT Software engineering notes* 24, 6 (1999), 253–267.
- [34] Lingming Zhang. 2018. Hybrid regression test selection. In *Proceedings of the 40th International Conference on Software Engineering*. 199–209.
- [35] Chenguang Zhu, Owolabi Legunsen, August Shi, and Milos Gligoric. 2019. A framework for checking regression test selection tools. In *2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE)*. IEEE, 430–441.
- [36] Chenguang Zhu, Yi Li, Julia Rubin, and Marsha Chechik. 2017. A dataset for dynamic discovery of semantic changes in version controlled software histories. In *2017 IEEE/ACM 14th International Conference on Mining Software Repositories (MSR)*. IEEE, 523–526.
- [37] Chenguang Zhu, Mengshi Zhang, Xiuheng Wu, Xiufeng Xu, and Yi Li. 2023. Client-Specific Upgrade Compatibility Checking via Knowledge-Guided Discovery. *ACM Trans. Softw. Eng. Methodol.* (feb 2023). <https://doi.org/10.1145/3582569> Just Accepted.