

Profiling Developers Through the Lens of Technical Debt

Zadia Codabux*
University of Saskatchewan
zadiacodabux@ieee.org

Christopher Dutchyn
University of Saskatchewan
dutchyn@cs.usask.ca

ABSTRACT

Context: Technical Debt needs to be managed to avoid disastrous consequences, and investigating developers' habits concerning technical debt management is invaluable information in software development. **Objective:** This study aims to characterize how developers manage technical debt based on the code smells they induce and the refactorings they apply. **Method:** We mined a publicly-available Technical Debt dataset for Git commit information, code smells, coding violations, and refactoring activities for each developer of a selected project. **Results:** By combining this information, we profile developers to recognize prolific coders, highlight activities that discriminate among developer roles (reviewer, lead, architect), and estimate coding maturity and technical debt tolerance.

CCS CONCEPTS

• Information systems → Data mining.

KEYWORDS

Developer Characterization; Open Source Software; Mining Software Repositories; Technical Debt; Code Smell; Refactoring

ACM Reference Format:

Zadia Codabux and Christopher Dutchyn. 2020. Profiling Developers Through the Lens of Technical Debt. In *ACM / IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM) (ESEM '20)*, October 8–9, 2020, Bari, Italy. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3382494.3422172>

1 INTRODUCTION

Open Source Software (OSS) platforms have thrived over the last decade. The latest "State of the Octoverse" report¹ by GitHub², one of the world's most popular and leading software development platforms for both open source and private project repositories, reports that 44+ million projects were developed by 40+ million developers from all over the world. OSS development is typically a collaborative process. Most OSS developers are skilled and experienced professionals with an average of 11.8 years of software development experience. About 50% of the developers have acquired formal

¹<https://octoverse.github.com/>

²<https://github.com/>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ESEM '20, October 8–9, 2020, Bari, Italy

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-7580-1/20/10...\$15.00

<https://doi.org/10.1145/3382494.3422172>

training in computer science or related domain, and the rest are self-taught or learned the skills on the job [8].

Different developers have different weaknesses and tolerances for Technical Debt (TD). TD is a metaphor introduced by Ward Cunningham [4] to help reason about the long-term costs incurred by taking shortcuts in software development. However, when TD is not managed, it can impede software development in multiple ways ranging from quality issues to loss of a system. One of the indicators of TD is *code smells* [15], a term coined by Kent Beck, which refer to surface indication of potential design weaknesses in source code [5]. One of the ways to manage TD is refactoring [3]. Refactoring is a technique for restructuring code by improving its internal structure without affecting its external behavior [5]. Therefore, establishing the developers' shortcomings with respect to TD can help focus scrutiny on portions of source code that may be more likely to contain flaws or debt, both by the coder during software development and by reviewers when they check the code during code reviews.

This study aims to profile developers with respect to technical debt by mining developers' commit actions, their code smells, and refactoring activities. This could help in team selection and organization, task assignments such as complex refactoring or code reviews, and pair programming sessions. We mined the publicly available TD dataset [9] to obtain the commit information from GitHub repositories, code smells, and coding violations from SonarQube³ and Ptidej [6], and refactoring information from Refactoring Miner [11]. We elaborate on the TD dataset and tools in Section 2.

The contributions of this study are:

- (1) We profile developers with respect to technical debt based on the code smells they induce and close and the type of refactoring they perform.
- (2) An updated TD dataset (in the .db format) with views for extracting the raw data for profiling developers, constructed using SQL queries.
- (3) The mapping of SonarQube coding violations to the most common code smells by Fowler [5] and the translation of these rules violations to bad practices, in the format of a CSV file (codified.csv in the replication package (Section 4.4)).

The rest of the paper is organized as follows. Section 2 describes the terminologies associated with the tools in the TD dataset used throughout this study. Section 3 highlights the relevant related work about developers' characterization. Section 4 focuses on the methodology of the study, including the research goal, questions, study design, and the data analysis process. Section 5 provides insights into the results. Section 6 presents the discussion. Section 7 lists threats to the validity of this study. Section 8 concludes and outlines future work.

³<https://www.sonarqube.org/>

2 BACKGROUND

In this section, we elaborate on the tools and associated terminologies that we use in this paper.

Git distinguishes between these different developers' roles⁴:

- *Author* is the person who originally wrote the work
- *Committer* is the person who last applied the work

Refactoring Miner is an open-source tool that mines refactoring operations from the git commits of Java projects. Some examples of refactoring types extracted are: *Extract Interface*, *Move Method*, and *Extract Method* [5].

SonarQube is a web-based open-source platform to analyze source code to detect violations to its pre-defined rules. SonarQube defines three types of issues⁵:

- **Bug**: an error that breaks the code; fix immediately
- **Security Vulnerability**: a point in code that's open to attack
- **Code Smell**: a maintainability issue that makes the code confusing and difficult to maintain

SonarQube has more than 500 rules⁶ for Java issues relating to different aspects of source code. SonarQube also classifies the rules into five severity levels: Blocker, Critical, Major, Minor, and Info. In this study, we only considered the SonarQube rules violations (which SonarQube refers to as "code smells"), but we mapped the rules violations to Fowler's widely-known code-smell classification [5] and bad coding practices. Henceforth, when we mention code smells in this study, we are referring to Fowler's classification. Many bad practices are not categorized in the SonarQube data. In fact, when delving into the TD dataset, we discover that more than 50% of the SonarQube rules violations end up uncategorized.

3 RELATED WORK

Here, we explore literature for studies on profiling OSS developers.

Yang et al. [14] proposed a developer portrait model based on three characteristics: personal information, programming skills, and contributions from the Open Source Platforms GitHub, GitLab, Gitee, and Bitbucket. Then, they successfully applied the model to code recommendation and programming tasks assignments. Augustinov et al. [2] analyzed the static analysis violations over the revision histories of the open-source projects: Hadoop Common, MongoDB, Spark, and Gaia. They tracked down which developers were responsible for introducing and fixing these violations to build developers' profiles about their coding habits. Leveraging the information on how defect is prioritized by software developers in defect repositories for the Open Source Systems Eclipse and Mozilla is the essence of the study conducted by Xuan et al. [13]. The authors modeled the developer's bug prioritization to assist in tasks such as bug triage, severity identification, and reopened bug prediction. Along the same lines, Xia et al. [12] analyzed bug reports from GCC, OpenOffice, Mozilla, Netbeans, and Eclipse to recommend developers for bug fixing. Amor et al. [1] proposed to use information from source code management systems, mailing lists archives, and bug tracking systems to characterize developer activities for more accurate effort and cost estimations in OSS. A

⁴<https://bit.ly/3ezc5GH>

⁵<https://bit.ly/2Cjp1Dz>

⁶<https://rules.sonarsource.com/java>

study by Hauff et al. [7] extracted GitHub user-profiles, specifically, their README.md files, to match developers' information with descriptions from job advertisements. Li [10] conducted a study focused on the analysis of developers' behaviors related to technical debts. Using clustering on Git commits, faults and SonarQube issues, including vulnerabilities, the author came up with the following categories of developers: Veterans, Vulnerability Creators, and Fault Inducers/ Commoners.

Although there are a few studies that explore the characteristics of developers in OSS based on artifacts (e.g., as defect information and source code) for various purposes such recommending specific tasks (e.g., bug fixing) or for mapping developer profiles to jobs, to the best of our knowledge, profiling developers from a technical debt perspective is yet to be explored from a combined code smell and refactoring perspective.

4 METHODOLOGY

4.1 Goal

The goal of this study is to mine developers' actions from code smells and refactoring information to profile developers from a technical debt perspective. In so doing, we will identify the developers who are superior at managing tasks, fixing bugs, and refactoring code. This developer characterization will help inform decisions regarding team selection and organization, task assignments, and pair programming sessions. We propose the following Research Question (RQ) to gather information to profile developers.

What do the developers' activities regarding inducing and refactoring code smells and bad practices tell us about their technical debt management?

4.2 Study Design

The data used in this study is from the publicly available Technical Debt Dataset (TD dataset), downloaded as a database, which consists of 33 Java projects from the Apache Software Foundation⁷ repository. The TD dataset contains commit information from the GitHub commit logs, refactoring extracted using Refactoring Miner, code smells, and programming violations from SonarQube and Ptidej. The projects in the TD dataset are more than three years old, contained at least 500+ commits and 100+ classes. *Because the dataset is public domain and contains full developer identities, our anonymization can be trivially reversed.*

Table Name	Description
Git Commits	Commit information retrieved from the commit log
Sonar Measures	SonarQube measures for the commits
Sonar Issues	SonarQube issues, antipatterns and code smells
Refactoring Miner	List of refactoring activities

Table 1: TD dataset SQL Tables

For this study, we selected the project Apache Beam (beam)⁸ which is one of the largest projects in the TD dataset. beam is a unified programming model used to implement batch and streaming data processing pipelines. The data for beam has been collected

⁷<https://www.apache.org/>

⁸<https://beam.apache.org/>

over 5 years (2014-2019) and consist of 22,300+ commits, 10,100+ refactoring, 74,400+ technical debt items. Since the focus of this study is on profiling developers from a refactoring and code smells perspective, we only used the related tables described in Table 1.

4.3 Data Analysis

Beginning with the TD dataset, we construct SQL queries to illuminate our research question, with summary statistics presented graphically. The data slices into three different perspectives:

- (1) The **commits** (GIT_COMMITS) will allow us to characterize the distribution of authorship, identify prolific code authors, and recognize central authorities who commit changes.
- (2) Details of statically-recognized programming concerns using SonarQube rules, which we categorized as **code smells** and **bad practices** along with the author of the defective code, the author and the committer of the repaired code, and the status of the concern.
- (3) **Refactoring** data allows us to trace standard refactorings [5] back to the original author; ranging from simple *RenameVariable* tasks to complex *ExtractSuperclass* actions.

Commits This data, displayed in Figure 2, tells us about authorship and commits for each developer, and we recognize the long-tail distribution and anomalies, which help us recognize specific developers. Also, the data only contains commits to the main branch, so we do not see prototyping or alternate exploration activities. However, the commits are tagged by *author*, so we know who constructed the code. The developer is frequently different from the *committer*, a point for discussion later.

There are several obstacles which introduce noise and complexity to the data supporting the analyses:

- Some individuals use more than one distinct but unambiguous *name* in the database— anonymized e.g.: `jsmith` and `Jon Smith` and `Jonathan Smith`
- Some entries are unambiguous email addresses or user-ids— anonymized e.g.: `nasa` and `nasa@nasa.gov`
- Other short forms, are ambiguous and cannot be resolved to an individual in the project— anonymized e.g.: `jsmith` might be `John Smith`, `Jon Smith`, or `Jane Smith`.
- Some commits are not tagged or use a non-descript e.g. `GitHub` or `MergeBot`

We enumerate these various names into the table `BEAM_HANDLES`, which contains rows, each with an `asEntered` name-string, and a `handle` for each; the handles are then anonymized into strings such as `Dev1`. Each of these fields are marked as case-indifferent⁹, and so collating sequences will put `Bill` near `bill` and `BILL`. By joining the `HANDLE` table on the data-source’s name-string, and extracting the `handle`, we can gather the various different names for a single developer to one canonical `handle`. This is seen in our queries in the extended SQL-Lite database as:

```
select ..., h.handle, ... from ...
  join BEAM_HANDLES h on h.asEntered = ...
```

In addition, our `BEAM_HANDLES` includes counts for the number of `GIT_COMMITS` with the `asEntered` name-string as the *author* and *committer*, as well as counts of SonarQube rule violations closed.

⁹This is the SQL-Lite term for case-insensitive.

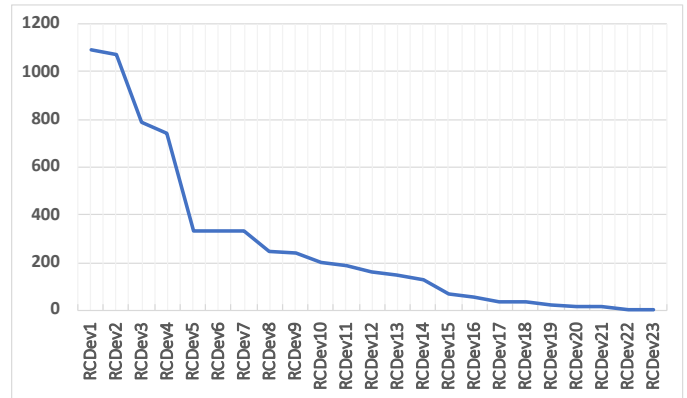


Figure 1: Count of Refactoring Commits

For the remaining three data-sources, just as above, we construct similar queries to select, combine, and summarize the different elements; then create informative charts to illustrate our learnings. **Code Smells and Bad Practices** In the TD dataset, `SONAR_MEASURES` (almost) joins on the `commitHash` with `GIT_COMMITS`, connecting SonarQube data with commits. Unfortunately, the `SONAR` tables truncate the hash by one leading character, but entries still remain in a 1-1 correspondence: every row in `MEASURES` has a unique `COMMIT` with a longer hash. Therefore, we extend `_BEAM_GIT_COMMITS` with an additional computed field, `shortHash`, that contains the truncated `commitHash`, which we use to build `BEAM_SONAR_MEASURES_COMMITS`. Then, we can join that view with the `_BEAM_SONAR_ISSUES` to connect an issue `_DETAIL` to each closing commit and thus to each author and committer for that issue.

Refactoring Activities Many of the Git commits have refactoring descriptions attached, using the `commitHash` as the join field. This allows us to compute summary statistics for the frequency of the various kinds of refactoring as a cross-table by author in Figure 1. But first, we recognize that the most prolific refactorers are also the authors tending to the right-hand side (expert) of the SonarQube data in that figure.

4.4 Replication Package

The full set of queries and views along with the data underlying our study is publicly available¹⁰. Specifically, we provide the SQL-Lite database with executable views for all tables, various spreadsheets formatting the diagrams, and the codified code smells file.

5 RESULTS

This section reports findings for our research question based on the three perspectives in Section 4.3.

Commits From the commit view in Section 4.3, we identify 754 unique name-strings, and this reduces down to 643 unique handles, shown in our `BEAM_HANDLES_SUMMARY` view. By sorting the view by descending author count, Figure 2 displays the long-tail distribution of authors as a blue-line from the highest at 2256 on the left, to the lowest at 1 on the right. Overlaying the committer count for each handle, as an orange-line superimposed in the same figure, we see

¹⁰<https://bit.ly/3h7PGSb>

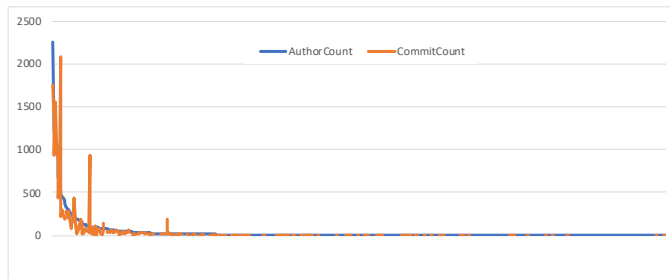


Figure 2: Distribution of Authoring and Committing

that commits tend to follow the same curve, but with a few notable differences, which highlight specific developers.

Most committers are authors, and over-committers stand out as blips where the orange (CommitCount) line jumps up above (or below) the blue-line (AuthorCount). Four authors stand out:

- at rank 9 with 506 authorships, 2085 commits,
- at rank 14 with 377 authorships, but no commits,
- at rank 39 with 98 authorships, and many (929) commits, and
- at rank 119 with 19 authorships, and many (188) commits.

The small blip at rank 23 in Figure 2 is MergeBot with 218 authorships and 429 commits. This is a generic handle for some automated process and does not inform us further. Overall, the top 22 handles represent 63% (14,106 of the 22332) of all authorships; and 70% (13,060 of 18,550) of all commits to the main branch. Therefore, we will tend to focus on the top 22 handles, and the two other over-committers (at ranks 39 and 119); and these are noted in the BEAM_HANDLES table as having an entry in a column, significantContributor which IS NOT NULL.

Code Smells and Bad Practices Summarizing our statistics for SonarQube rule violations (which include Fowler code-smells and other bad practices), we find 78,642 instances of an author making a change that closes a recognized SonarQube rule violation, with 2602 distinct messages. We extracted all 2602 unique Detail messages from this table, and manually marked them with with a Type and SubType, as described in codified.csv in our replication package. Our summary view, BEAM_SONAR_ISSUES_SUMMARY shows the details; briefly, there are three Types:

- Anti-Pattern (with one SubType, *Anti-Singleton*)
- Bad Practice (with numerous SubTypes, including our own)
- Code Smell (with 12 standard SubTypes).

We summarized SonarQube messages against significant contributors by handle, as our BEAM_SONAR_MESSAGES_BY_AUTHOR view.

Seventeen of our significant contributors authored changes that closed SonarQube rule violations. The most prolific contributor removed 12,500 SonarQube violations of all types. All told, seventeen authors repaired 51,187 instances. The BEAM_SONAR_MESSAGES_SUMMARY computes these values; and Figure 3 displays the curve.

We extracted those records to a csv file for import into a presentation tool. Using simple text-processing tools, we converted the row representation into a cross-table in graphical format, c.f. Figure 4. The graph has rows and columns reordered to emphasize interesting details, deferred to the Discussion section. In this figure, the total number of commits is normalized to 100%, and the variation among the types and subtypes is emphasized. Note that the order of authors is anonymized separately and in a different order

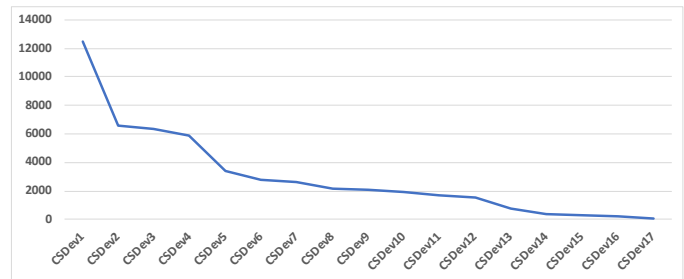


Figure 3: Count of Sonar Violations Resolved by Author

from previous figures. Each vertical bar corresponds to a single author (captioned along the bottom) and is divided into distinguishing color bands based on type and subtype. The order of the bands is identical across all authors.

The sole *AntiPattern* type (with unique *Anti-Singleton* subtype) is so small as to be indistinguishable in black at the top of each bar: indeed, there were only 17 instances of this among 51,187 SonarQube violations. For the remaining SonarQube rule violations, we stacked code smells above bad practices. Twelve of the recognized code smells [5] were identified in the data, but only three showed substantial difference across the developers: *Long-Method*, *DeadCode*, and *ComplexMethod*. Therefore, we color-coded the three interestingly-varied code-smells in different bright-orange and merged the other nine into dim-orange.

In the same way, we examined the bad practices data, and there were 34 distinct sub-types; but only eight showed up more than 1% of the time, and all of those had noticeable variation (as measured by standard deviation); they are bolded in Figure 5.

The other 26 subtypes of bad practices were present in 1% of the violations, and standard deviation of at most 1%, and merged together into the very dim blue at the bottom. We note a few of these:

- Unlike most practices, only three developers engaged in *MoveFile* (the yellow band), and one of them performed almost all of the file moves, and almost half of his work was resolving those rule violations.
- Abstraction violations were almost one-quarter of all rule violations but tended to dominate the others who also had higher levels of style violations, obsolete code fixes, and magic numbers.

We defer the interpretation of these two points to a later section.

Refactoring Activities With the cross-table query, we get 23 distinct significant contributors and a categorization of their commits into 16 distinct types. Some of them are small contributions (less than 10% of the instances, and single-digit variances), but six have some presence, and two more are close. Figure 5 shows the distribution of activity, much as we did previously.

Of note in the table,

- Two authors performed one refactoring exclusively: *Inline Method* and *Push Down Method* respectively
- And except for those two, everyone else employed *Extract Method*, a relatively sophisticated refactoring type
- *Move Class* was a common refactoring for most other authors, and in some cases, a substantial majority of their actions

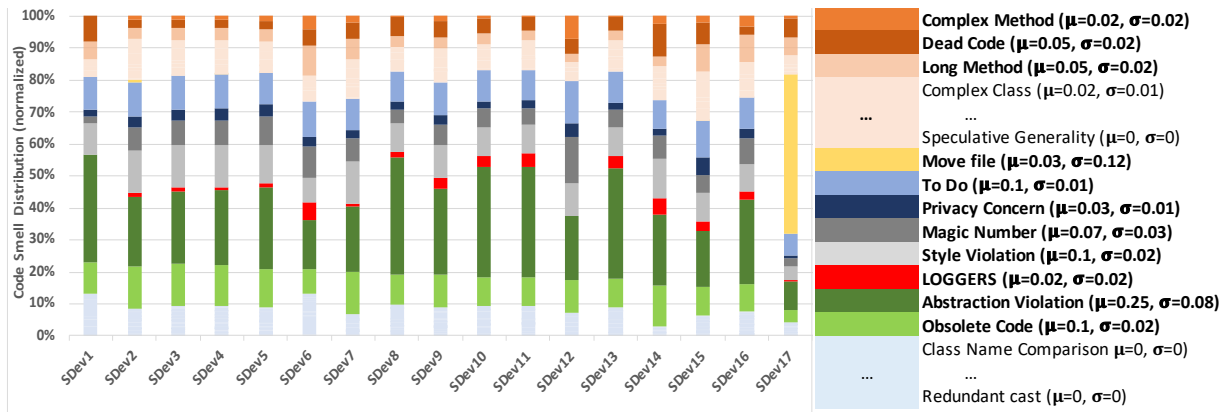


Figure 4: Sonar Violations Distribution by Author and Type

- Only five authors employed *Move Method*, but they used it extensively (25% or more).

6 DISCUSSION

Our results offer numerous opportunities for discussion. We have confidence that our analysis offers a fair assessment of the available data, regardless of our choice to focus on a small subset of contributors. Overall, the top 22 handles represent 63% (14,106 of the 22,332) of all authorships; and 70% (13,060 of 18,550) of all commits to the main branch. This gives us a good understanding of the various developers involved in the beam project.

Profiling Developers We do not produce a profile document for each developer; instead, we perform the act of profiling, recognizing the similarities and differences among developers and relating those to a level of coding maturity and coding specialization. We profile developers by author and commit activity, recognizing those with unusual commit counts compared to authoring—these individuals appear to be distinguishable code-reviewers, technical leads, or project managers. It also allows us to discriminate different kinds of code authors: some perform global changes such as moving code among packages—these might be architects. Others create or close large numbers of *ToDo* items—perhaps quality assurance personnel. One focused exclusively on pushing methods down: dealing with *Generalization* problems. He would need to carefully consider the risks of running general code (a super method) over methods specialized to different classes—i.e., a developer with in-depth knowledge of the package. Another simply inlined methods: a method-composition act yielding performance optimization. This is a task for a developer who is confident that the original method is superfluous: one who understands when a method abstraction is unnecessary and will not add excessive complexity—i.e., someone with experience, knowledge of the code, and good intuitions about code complexity. Some others stand out for large numbers of Abstraction Violations, Complex code, obsolete code (perhaps cribbed from StackOverflow), or trivial Style-Guide or Magic-Number violations—often a hallmark of immature developers. Developers with less than 80% of their SonarQube rule violations identified as bad practices tended to introduce more code smells (except for the file-mover)—developers with fewer bad practices, especially of the more trivial

sort, seem to tolerate larger amounts of code smell. This profiling exercise suggests that mature developers produce fewer bad practices, but admit larger amounts of technical debt.

7 THREATS TO VALIDITY

This is a preliminary study based on one system, beam, specific to the Apache foundation and written in Java. Although it is a large project over multiple years, we cannot generalize our results to other projects, both OSS and industrial. However, this study serves as a basis to expand this research to other systems. The TD dataset provides snapshots of projects over a specific timeframe, and there might be some missing data. However, the data for beam spans since 2014, which coincides with the first release of an open SDK implementation by Google. So we can assume that we have a fairly complete picture of the project over the TD dataset timeline.

We used the data from the TD dataset database, which uses specific tools and algorithms. Most of the tools were used with the standard configuration. So we might have missed some of the data that the tools might not have reported. Most of the data in the Git commits are self-reported by the developers and might not be an accurate reflection of the actual data due to a lack of knowledge and experience to identify the correct components or assign the right labels.

The mapping of the SonarQube rule violations to the more commonly known code smells by Fowler, and bad practices (captured in our `codified.csv`) is a manual process by one author and cross-checked by the other. We tried to compare with research papers and the SonarQube set of rules to be as unbiased and accurate as possible, but it is possible that we have misinterpreted or wrongly classified some of the rules.

Lastly, some of the developers' handles were ambiguous, that is, a particular handle could refer to different people as described in Section 4.3. Therefore, we had to exclude the records of those developers. We also had to ignore the commits and violations that were not tagged as reporter, assignee, or author; and those with developers such as GitHub or MergeBot.

8 SUMMARY AND FUTURE WORK

Overall, the TD dataset offers a rich data source for profiling developers, including

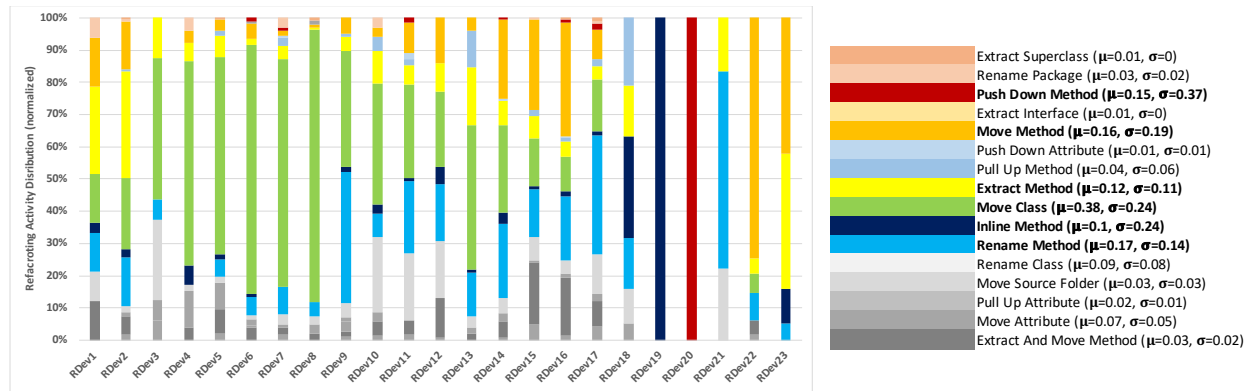


Figure 5: Refactorings by Type and Author

- (1) raw commits to main branches, tagged by author and committer; which allows us to identify and distinguish OSS contributors by the level of contribution
- (2) SonarQube diagnostics and closing commits; which allow us to identify the developers and the kinds of diagnostics they contribute and discharge
- (3) categorization of refactorings, allowing us to examine the range of refactoring techniques that authors employed: some used very few; others used many.

By analyzing these data sources, we characterize developers by author and commit activity, recognizing those with unusual commit counts compared to authoring. The SonarQube data allows us to profile each developer, distinguishing them by code smells and bad practices. The profiling exercise uncovers evidence supporting maturity and experience.

Last, the Refactoring Miner data supports that the mature/experienced developers employ more sophisticated refactorings, involving generalization and in-depth code knowledge, rather than the simpler renamings that the other developers usually perform.

In profiling developers, we uncover details informing decisions

- for assigning pair-programming teams: by knowing the strengths of various developers, teams can be purpose-built to transfer knowledge, or expose new techniques, or to make speedy progress on paramount tasks
- for inviting individuals to review-panels to ensure diversity of viewpoints, strengths, and techniques

Future Work The TD dataset contains additional data sources, including the full details of each commit in `GIT_COMMITS_CHANGES`, and fault-tracking information in `SZZ_FAULT_INDUCING_COMMITS` tables. The former would give precise commit data, telling us, on a file-by-file basis, the number of lines inserted, deleted, and moved. This could sharpen our profiling focus to include expertise in specific packages or code sections. The latter can be used to identify developers who introduce faults, and the developers who repair them—additional dimensions to profile. The SonarQube data could also be analyzed in another dimension: as a time-series. Analyzing in this way can allow us to see how developers change their behaviors over time. Results could inform questions like "do developers become more tolerant to more or different debt as they become more experienced?" and "do developers change the set of repair

actions as they become more involved in a project?". These exciting questions remain for later work. Our data comes from the TD dataset; this decimates the information down to the main branch. In the future, we can draw on additional data sources, including the entire commit log to date, including branches that were prototypes or explored alternate constructions. Eventually, these profiling activities might form the basis of an actual artifact, a *developer profile*, containing salient details about an individual.

REFERENCES

- [1] Juan Jose Amor, Gregorio Robles, and Jesus M Gonzalez-Barahona. 2006. Effort estimation by characterizing developer activity. In *2006 international workshop on Economics Driven Software Engineering Research*. IEEE, 3–6.
- [2] Pavel Avgustinov, Arthur I Baars, Anders S Henriksen, Greg Lavender, Galen Menzel, Oege de Moor, Max Schäfer, and Julian Tibble. 2015. Tracking static analysis violations over time to capture developer characteristics. In *2015 IEEE/ACM 37th International Conference on Software Engineering*. Vol. 1. IEEE, 437–447.
- [3] Zadia Codabux, Byron J Williams, and Nan Niu. 2014. A quality assurance approach to technical debt. In *International Conference on Software Engineering Research and Practice*. WASET.
- [4] Ward Cunningham. 1992. The WyCash Portfolio Management System, OOPSLA, 1992. *SIGPLAN OOPS Messenger* 4, 2 (December 1992), 29–30.
- [5] Martin Fowler. 2018. *Refactoring: improving the design of existing code*. AW.
- [6] Yann-Gaël Guéhéneuc. 2007. Ptidej: A flexible reverse engineering tool suite. In *2007 International Conference on Software Maintenance*. IEEE, 529–530.
- [7] Claudia Hauff and Georgios Gousios. 2015. Matching GitHub developer profiles to job advertisements. In *2015 IEEE/ACM 12th Working Conference on Mining Software Repositories*. IEEE, 362–366.
- [8] Karim R Lakhani and Robert G Wolf. 2003. Why hackers do what they do: Understanding motivation and effort in free/open source software projects. MIT Sloan working paper.
- [9] Valentina Lenarduzzi, Nyyti Saarimäki, and Davide Taibi. 2019. The technical debt dataset. In *15th International Conference on Predictive Models and Data Analytics in Software Engineering*. 2–11.
- [10] Xiaozhou Li. 2019. Research on Software Project Developer Behaviors with K-means Clustering Analysis. In *SSSME 2019: Summer School on Software Maintenance and Evolution*. CEUR-WS.
- [11] Nikolaos Tsantalos, Matin Mansouri, Laleh Eshkevari, Davood Mazinanian, and Danny Dig. 2018. Accurate and efficient refactoring detection in commit history. In *2018 IEEE/ACM 40th International Conference on Software Engineering*. IEEE, 483–494.
- [12] Xin Xia, David Lo, Xinyu Wang, and Bo Zhou. 2013. Accurate developer recommendation for bug resolution. In *20th Working Conference on Reverse Engineering*. IEEE, 72–81.
- [13] Jifeng Xuan, He Jiang, Zhilei Ren, and Weiqin Zou. 2012. Developer prioritization in bug repositories. In *34th International Conference on Software Engineering*. IEEE, 25–35.
- [14] Wenhua Yang, Minxue Pan, Yu Zhou, and Zhiqiu Huang. 2020. Developer Portraying: A Quick Approach to Understanding Developers on OSS Platforms. *Information and Software Technology* 125 (2020), 106336.
- [15] Nico Zazworka, Clemente Izurieta, Sunny Wong, Yuanfang Cai, Carolyn Seaman, Forrest Shull, et al. 2014. Comparing four approaches for technical debt identification. *Software Quality Journal* 22, 3 (2014), 403–426.