

# Identifying Bugs Related to Non-Functional Requirements in Java and Python

---

Aida Radu  
Sarah Nadi, Supervisor  
Dept. of Computing Science, University of Alberta



# Background: Non-Functional Bugs

- ❖ Non-Functional Bug (NFB) = a bug that does not affect what the program does, but **how** it does it
- ❖ Fixes include making programs more time efficient, memory efficient, secure, etc.

# OBJECTIVE

To create a dataset of NFBs in open-source software projects



# Related Work

1. **MuBench**: dataset of 89 API misuses from 33 open source projects and survey
2. **Defects4J**: 357 real Java bugs from 5 open source projects
3. **Bugs.jar**: 1,158 Java bugs from 8 open source Apache projects
4. **Ohira et al**: dataset of 4000 functional and non-functional bugs from 4 open source Apache projects
5. **iBugs**: dataset of 369 realistic Java bugs from AspectJ

## Our Dataset:

- ❖ 138 bugs from 67 open-source projects
- ❖ exclusive to non-functional bugs
- ❖ Projects written in Java or Python

<sup>1</sup>S. Amann, S. Nadi, H.A. Nguyen, T.N. Nguyen, and M. Mezini. MUBench: a benchmark for API-misuse detectors. MSR '16, pages 464-467. ACM, 2016.

<sup>2</sup>R. Just, D. Jalali, and M. D. Ernst. Defects4J: A Database of Existing Faults to Enable Controlled Testing Studies for Java Programs. ISSTA'14, pages 437-440. ACM, 2014

<sup>3</sup>RK. Saha, Y. Lyu, W. Lam, H. Yoshida, and MR. Prasad. Bugs.jar: a large-scale, diverse dataset of real-world Java bugs. MSR '18, pages 10-13. ACM, 2018.

<sup>4</sup>M. Ohira, Y. Kashiwa, Y. Yamatani, H. Yoshiyuki, Y. Maeda, N. Limsettho, K. Fujino, H. Hata, A. Ihara, and K. Matsumoto. A dataset of high impact bugs: manually-classified issue reports. MSR '15, pages 518-521. IEEE Press, 2015.

<sup>5</sup>V. Dallmeier and T. Zimmermann. Extraction of Bug Localization Benchmarks from History. ASE'07, pages 433-436. ACM, 2007.

# Methodology



```
source:  
  name: github-search  
project:  
  name: Fowler  
  url: https://github.com/TheFreshMango/Fowler  
fix:  
  tag: performance  
  description: Replacing "+" with StringBuilder improves performance and memory space  
  commit message: >  
    refactoring: Customer now uses StringBuilder for statement()  
  commit: https://github.com/TheFreshMango/Fowler/commit/d5f5298  
location:  
  file:  
    Fowler/Fowler Refactor/src/de/dhbw/fowler/Customer.java  
  method:  
    public String statement()  
api:  
  java.lang.String  
api change:  
  java.lang.String -> java.lang.StringBuilder  
rules:  
  use a StringBuilder instead of adding strings to concatenate efficiently
```

Identify Candidate  
Repositories

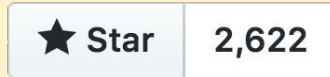
Extract commits  
using PyDriller

Manual Review

Documentation

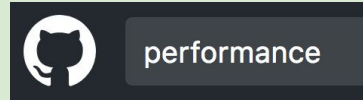
# Identifying Candidate Repositories

## Method One: Star Rating



- ❖ Github users “star” repositories they are interested in
- ❖ We chose the repositories with the highest number of stars
- ❖ 91 Java and 15 Python projects reviewed

## Method Two: Github Search



- ❖ Github allows users to search for keywords in different domains (projects, commits, issues, etc.)
- ❖ 23 Java and 11 Python projects reviewed

## Method Three: RepoReapers dataset



- ❖ Munaiah et al: set of “well-engineered” Github projects<sup>1</sup>
- ❖ we selected those with the most stars
- ❖ 17 Java and 10 Python projects reviewed

<sup>1</sup>Munaiah N, Kroh S, Cabrey C, and Nagappan M. Curating GitHub for engineered software projects. Empirical Software Engineering. 22(6): pages 3219-3253. Springer US. 2017.

# Extracting Commits Using PyDriller



- ❖ PyDriller<sup>1</sup> is an open-source software-mining tool for Git
- ❖ Filtered the commit history of projects to messages containing **keywords** related to NFBs
- ❖ Restricted search to commits that changed .java or .py files

"fix"	"bug"	"error"
"refactor"	"secur[ity]"	"maint[enance]"
"stab[ility]"	"portab[ility]"	"efficien[cy]"
"usab[ility]"	"reliab[ility]"	"testab[ility]"
"changeab[ility]"	"replac[e]"	"memory"
"resource"	"runtime"	"crash"
"leak"	"attack"	"authentikat[ion]"
"authoriz[ation]"	"cipher"	"crack"
"decrypt"	"encrypt"	"vulnerab[ility]"
"minimiz[e]"	"optimiz[e]"	"slow"
"#"	"fast"	"perform[ance]"

terms filtered out		
"typo"	"npe"	"spell"

These keywords include original words relevant to the study, as well as subsets from Hindle et al.<sup>2</sup> and de la Mora and Nadi<sup>3</sup>.

<sup>1</sup>D. Spadini, M. Aniche, and A. Bacchelli. PyDriller: Python Framework for Mining Software Repositories. ESEC/FSE. 2018. <https://github.com/ishepard/pydriller>

<sup>2</sup>A. Hindle, NA. Ernst, MW. Godfrey, and J. Mylopoulos. Automated topic naming to support cross-project analysis of software maintenance activities. MSR '11, pages 163-172. ACM, 2011

<sup>3</sup>FL. de la Mora and S. Nadi. Which library should I use? A metric-based comparison of software libraries. ICSE NIER '18, pages 37-40. ACM, 2018

# Manual Review



- ❖ Weeding out false positives
- ❖ Example:

```
2 █ █ █ █ speed_bomb.py

@@ -245,7 +245,7 @@ def reset_bomb_length(self):
245 245         self.bomb_length = 5.0
246 246
247 247     def get_bomb_length(self):
248 + -         self.bomb_length -= 0.2
248 +         self.bomb_length -= 0.3
249 249         if self.bomb_length < 1:
250 250             self.bomb_length = 1
251 251         return self.bomb_length


```

**Commit Message:** "Update speed\_bomb.py made speed change faster"

<https://github.com/adangert/JoustMania/commit/a9711c2>



# Documenting Bugs

```
source:
  name: github-search
project:
  name: Fowler
  url: https://github.com/TheFreshMango/Fowler
fix:
  tag: performance
  description: Replacing "+" with StringBuilder improves performance and memory space
  commit message: >
    refactoring: Customer now uses StringBuilder for statement()
  commit: https://github.com/TheFreshMango/Fowler/commit/d5f5290
location:
  file:
    Fowler/Fowler Refactor/src/de/dhbw/fowler/Customer.java
  method:
    public String statement()
api:
  java.lang.String
api change:
  java.lang.String -> java.lang.StringBuilder
rule:
  use a StringBuilder instead of adding strings to concatenate efficiently
```

# Documenting Bugs

```
source:
  name: github-search
project:
  name: Fowler
  url: https://github.com/TheFreshMango/Fowler
fix:
  tag: performance
  description: Replacing "+" with StringBuilder improves performance and memory space
  commit message: >
    refactoring: Customer now uses StringBuilder for statement()
  commit: https://github.com/TheFreshMango/Fowler/commit/d5f5290
location:
  file:
    Fowler/Fowler Refactor/src/de/dhbw/fowler/Customer.java
  method:
    public String statement()
api:
  java.lang.String
api change:
  java.lang.String -> java.lang.StringBuilder
rule:
  use a StringBuilder instead of adding strings to concatenate efficiently
```



bug category

# Documenting Bugs

```
source:
  name: github-search
project:
  name: Fowler
  url: https://github.com/TheFreshMango/Fowler
fix:
  tag: performance
  description: Replacing "+" with StringBuilder improves performance and memory space
  commit message: >
    refactoring: Customer now uses StringBuilder for statement()
  commit: https://github.com/TheFreshMango/Fowler/commit/d5f5290
location:
  file:
    Fowler/Fowler Refactor/src/de/dhbw/fowler/Customer.java
  method:
    public String statement()
api:
  java.lang.String
api change:
  java.lang.String -> java.lang.StringBuilder
rule:
  use a StringBuilder instead of adding strings to concatenate efficiently
```



bug category



our interpretation  
of the fix

# Documenting Bugs

```
source:
  name: github-search
project:
  name: Fowler
  url: https://github.com/TheFreshMango/Fowler
fix:
  tag: performance
  description: Replacing "+" with StringBuilder improves performance and memory space
  commit message: >
    refactoring: Customer now uses StringBuilder for statement()
  commit: https://github.com/TheFreshMango/Fowler/commit/d5f5290
location:
  file:
    Fowler/Fowler Refactor/src/de/dhbw/fowler/Customer.java
  method:
    public String statement()
api:
  java.lang.String
api change:
  java.lang.String -> java.lang.StringBuilder
rule:
  use a StringBuilder instead of adding strings to concatenate efficiently
```



bug category



our interpretation  
of the fix



related api  
(if applicable)

# Documenting Bugs

```
source:
  name: github-search
project:
  name: Fowler
  url: https://github.com/TheFreshMango/Fowler
fix:
  tag: performance
  description: Replacing "+" with StringBuilder improves performance and memory space
  commit message: >
    refactoring: Customer now uses StringBuilder for statement()
  commit: https://github.com/TheFreshMango/Fowler/commit/d5f5290
location:
  file:
    Fowler/Fowler Refactor/src/de/dhbw/fowler/Customer.java
  method:
    public String statement()
api:
  java.lang.String
api change:
  java.lang.String -> java.lang.StringBuilder
rule:
  use a StringBuilder instead of adding strings to concatenate efficiently
```



bug category



our interpretation  
of the fix

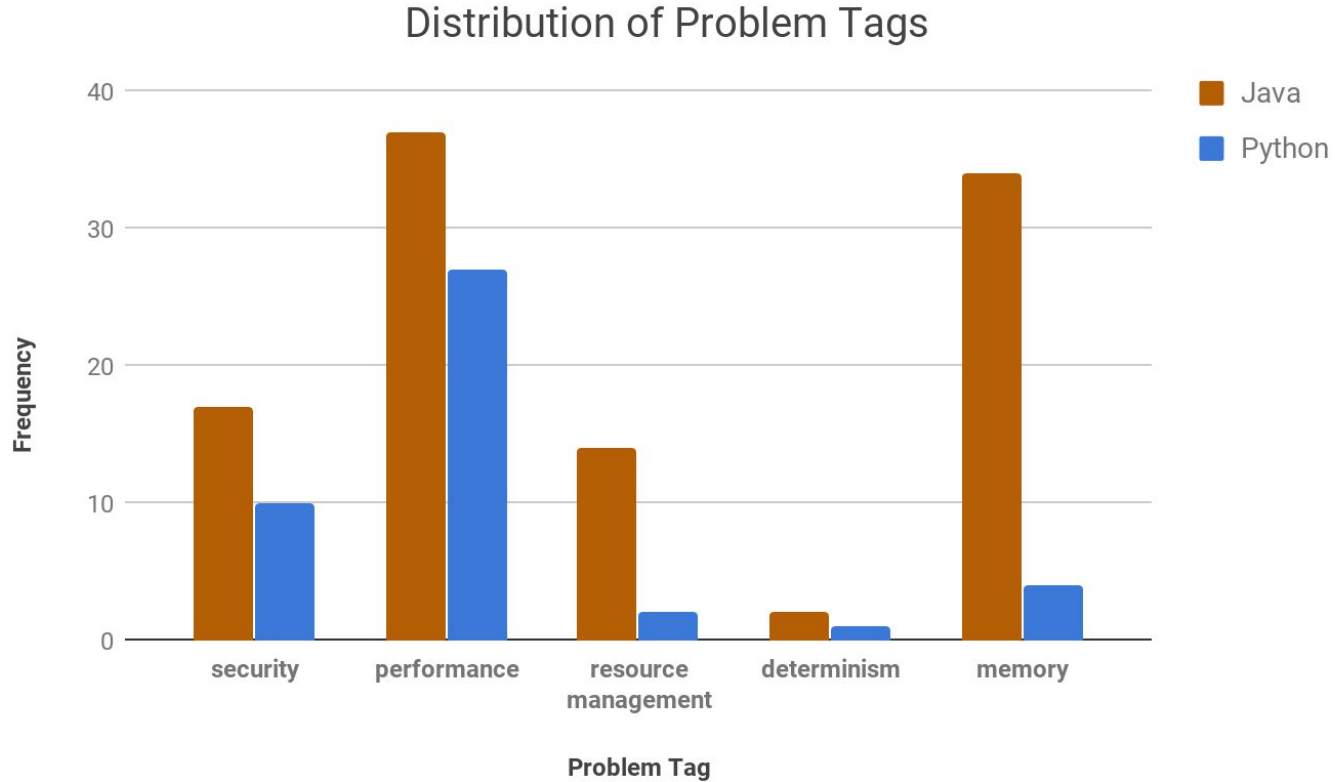


related api  
(if applicable)

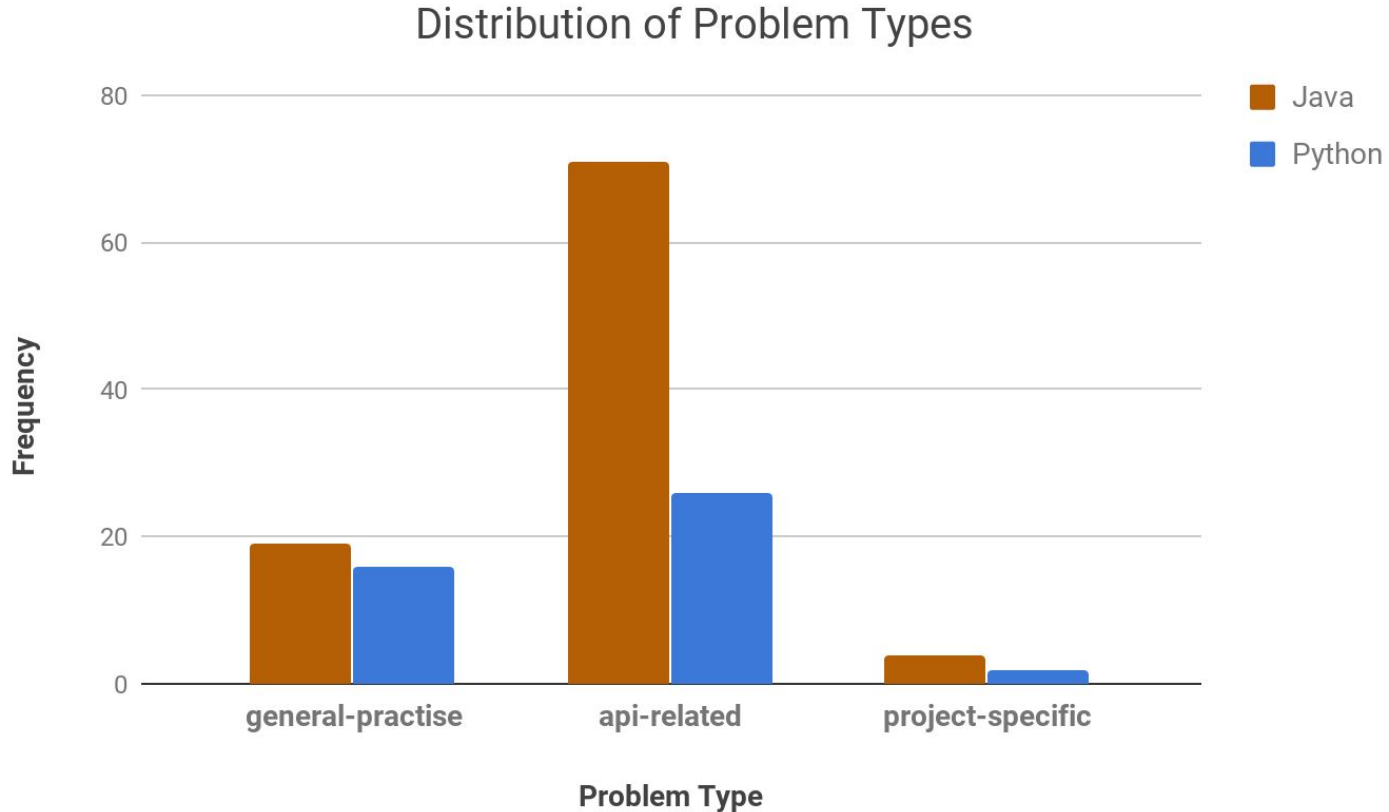


conclusion

# Data Distribution

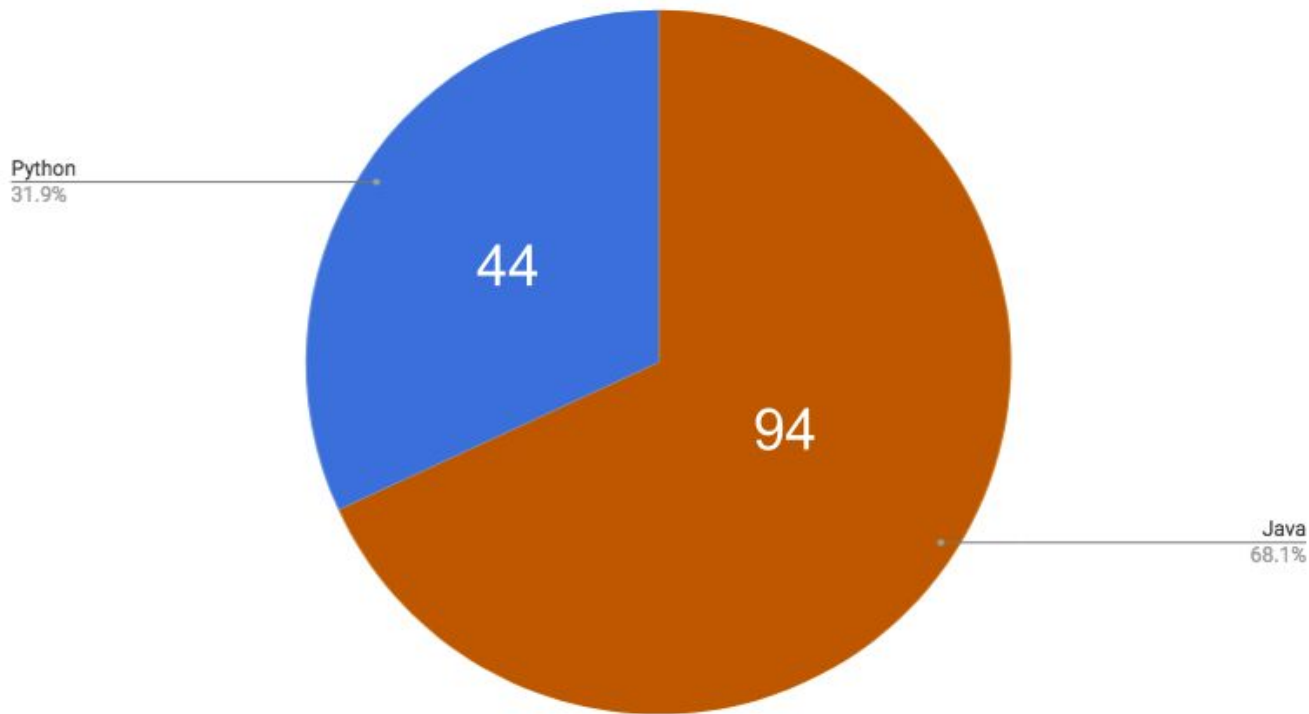


# Data Distribution



# Data Distribution

Number of Problems from Each Language





# Repo Organization

aradu12 added number of repos ...	
📁 GumCode	Create APIExtractor.java
📁 PyDrillerMining	Delete commit_out1.csv
📁 java-data	Merge branch 'master' of https://gith
📁 py-data	Update problem.yml
📁 scripts	add contact info
📄 README.md	added number of repos
📄 ReposMined.yml	added owner names
📄 miscCommits.yml	Create miscCommits.yml

- ❖ Separate folders for data in each language

# Repo Organization

aradu12 added number of repos ...	
📁 GumCode	Create APIExtractor.java
📁 PyDrillerMining	Delete commit_out1.csv
📁 java-data	Merge branch 'master' of https://gith...
📁 py-data	Update problem.yml
📁 scripts	add contact info
📄 README.md	added number of repos
📄 ReposMined.yml	added owner names
📄 miscCommits.yml	Create miscCommits.yml



Branch: master	researchwiki-radu / java-data /	Create new file	Upload files	Find file	History
aradu12 Merge branch 'master' of https://github.com/uAlberta-smr/researchwiki... Latest commit c26c018 5 days ago					
..					
📁 ADFGVX	Update problem.yml	2 months ago			
📁 AirMapView	remove @	10 days ago			
📁 AndroidBillingLibrary	minimize (#9) and rename	2 months ago			
📁 Catacomb-Snatch	copy pasted the import by accident	a month ago			
📁 EfficientAdapter	minimize usage (close #5) @snadi Thoughts/feedback?	2 months ago			
📁 ExampleProject	Rename java-data/ExampleProject/general-practise/1/problem.yml to jav...	a month ago			
📁 ForestFireModel	Update problem.yml	2 months ago			
📁 Fowler	Update problem.yml	2 months ago			
📁 Glazed	Rename java-data/Glazed/problems/1/problem.yml to java-data/Glazed/pr...	a month ago			
📁 POO	#8	2 months ago			
📁 PointCloudDatasetReader	Update problem.yml	2 months ago			
📁 PyDev	Rename java-data/PyDev/problems/1/problem.yml to java-data/PyDev/prob...	a month ago			
📁 Raiawali	oops wrong source	7 days ago			

- ❖ Separate folders for data in each language
- ❖ Each project has its own folder

# Repo Organization



these folders contain problems, problem version IDs, and project metadata

# Repo Organization



these folders contain problems, problem version IDs, and project metadata



problems are grouped by type

# Repo Organization



these folders contain problems, problem version IDs, and project metadata



problems are grouped by type



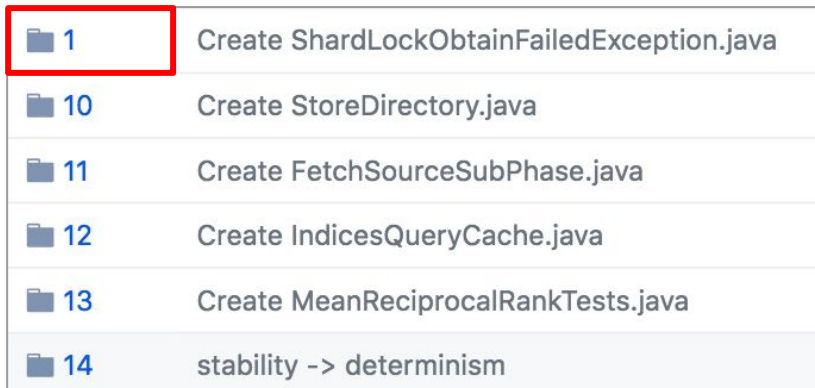
1	Create ShardLockObtainFailedException.java
10	Create StoreDirectory.java
11	Create FetchSourceSubPhase.java
12	Create IndicesQueryCache.java
13	Create MeanReciprocalRankTests.java
14	stability -> determinism

each problem has its own folder

# Repo Organization



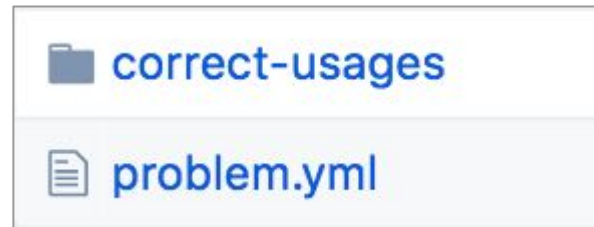
these folders contain problems, problem version IDs, and project metadata



each problem has its own folder



problems are grouped by type



the problem details

# Processing the Data

aradu12 added number of repos ...	
📁 GumCode	Create APIExtractor.java
📁 PyDrillerMining	Delete commit_out1.csv
📁 java-data	Merge branch 'master' of https://github.com
📁 py-data	Update problem.yml
📁 <b>scripts</b>	add contact info
📄 README.md	added number of repos
📄 ReposMined.yml	added owner names
📄 miscCommits.yml	Create miscCommits.yml



📄 DataBox.py
📄 DataBoxClient.py
📄 __init__.py

- ❖ We provide scripts to manage the data
- ❖ DataBox class allows users to filter problems by star rating, tag, problem type, etc.
- ❖ Users can write their own client script or use our example: DataBoxClient.py

# Applications of the Dataset


---

- ❖ Code Recommenders
  - ❖ Training set for bug detection tools
  - ❖ Analysis of coding practices between veterans and beginners
-



# OBJECTIVE

To create a dataset of NFBs in open-source software projects



3

# OBJECTIVE

To create a dataset of NFBs in open-source software projects

---

3

## Methodology

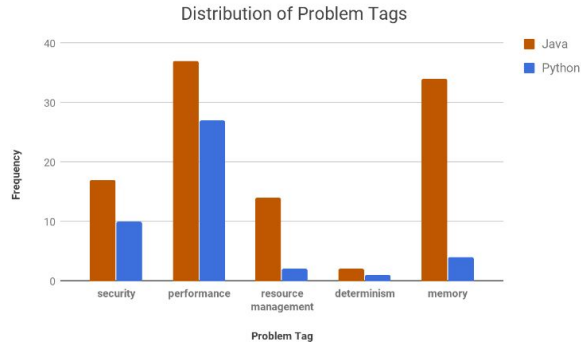


5

# OBJECTIVE

To create a dataset of NFBs in open-source software projects

## Data Distribution



14

## Methodology

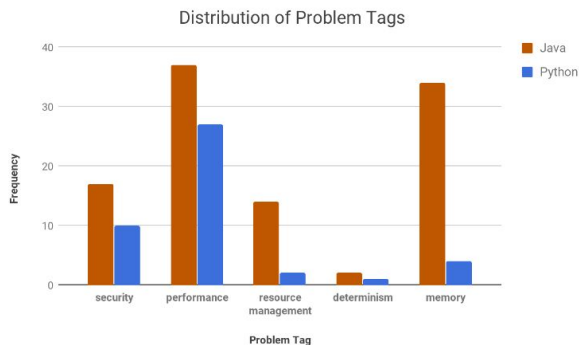


5

# OBJECTIVE

To create a dataset of NFBs in open-source software projects

## Data Distribution



14

## Methodology



5

## Processing the Data

aradur12 added number of repos	
GitCode	Create APIExtractor.java
PyDrillerMining	Delete commit_out1.csv
java-data	Merge branch 'master' of https://github
py-data	Update problem.yml
scripts	add contact info
README.md	added number of repos
ReposMined.yml	added owner names
miscCommits.yml	Create miscCommits.yml



DataBox.py
DataBoxClient.py
__init__.py

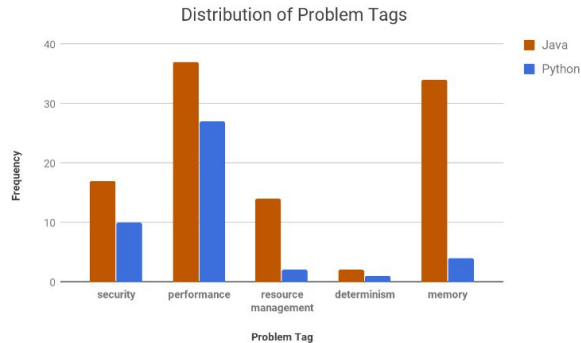
- ❖ We provide scripts to manage the data
- ❖ DataBox class allows users to filter problems by star rating, tag, problem type, etc.
- ❖ Users can write their own client script or use our example: DataBoxClient.py

23

## OBJECTIVE

To create a dataset of NFBs in open-source software projects

## Data Distribution



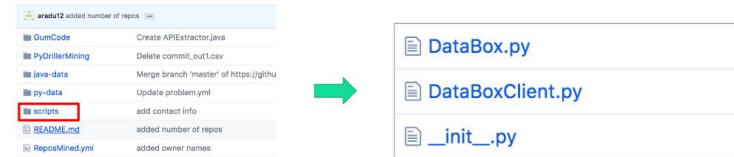
14

## Methodology



5

## Processing the Data



- ❖ We provide scripts to manage the data
- ❖ DataBox class allows users to filter problems by star rating, tag, problem type, etc.
- ❖ Users can write their own client script or use our example: DataBoxClient.py

23