

Are Code Examples on an Online Q&A Forum Reliable?

A study of API Misuse on Stack Overflow

Tianyi Zhang, Ganesha Upadhyaya, Anastasia Reinhardt, Hridesh Rajan, Miryung Kim
ICSE 2018

Presented by: Igor Muzetti Pereira
Software reuse - 18/2 - PPGCC/DCC/UFMG

Index

- Introduction
- API Usage Mining from GitHub
- An Empirical Study of API Misuse on Stack Overflow
- Related Work
- Discussion

Introduction

- Programmers often resort to online code examples during software development.
- Using APIs properly is a key challenge in Programming.
- None studies have investigated the reliability of online code examples in the sense that a code example may lead to API usage violations.
- Developers often search online for code examples to learn APIs.



GitLab



Introduction

- Limitations
 - Programmers can only inspect a handful of search results.
 - Individual code examples may suffer from
 - Insecure coding practices
 - Unchecked obsolete usage
 - Low readability

Introduction

- This paper presents an empirical study on the prevalence and severity of API misuse on Stack Overflow.
 - Aims to assess the reliability of Stack Overflow posts by contrasting code examples against desirable API usage patterns mined from GitHub.
- Was designed API usage mining technology, the EXAMPLECHECK.
- Was targeted one hundred Java and Android APIs that are popular and have corresponding posts in Stack Overflow.
- Was inspected all patterns inferred by EXAMPLECHECK, create a data set of 180 desirable API usage patterns for the 100 APIs, and study the extent of API misuse in Stack Overflow posts.

Introduction

- Out of 217,818 SO posts relevant to our data set, 31% contain potential API misuse that could produce symptoms such as program crashes, resource leaks, and incomplete actions.
- Such API misuse is caused by three main reasons:
 - Missing control constructs, missing or incorrect order of API calls, and incorrect guard conditions.
- 29% of recognized posts—those accepted as correct answers or endorsed by other programmers—have potential API usage violations, while 34% unrecognized posts have API usage violations.

Introduction

How do I write data to a file using FileChannel?

▲ You're probably interested in a `FileChannel`. Channels were designed to perform bulk IO operations to and from `Buffers`.

2

▼ Ex:

```
FileChannel fileOut = new FileOutputStream(file).getChannel();
fileOut.write(ByteBuffer.wrap("whatever you want to write".getBytes()));
```

share edit

answered Apr 9 '12 at 19:28

 Jeffrey
32.2k • 7 • 52 • 10k

(a) An example that does not close FileChannel properly

▲ Somewhat like this:

6

▼

```
short[] payload = {1,2,3,4,5,6,7,8,9,0};
ByteBuffer myByteBuffer = ByteBuffer.allocate(20);
myByteBuffer.order(ByteOrder.LITTLE_ENDIAN);

ShortBuffer myShortBuffer = myByteBuffer.asShortBuffer();
myShortBuffer.put(payload);

FileChannel out = new FileOutputStream("sample.bin").getChannel();
out.write(myByteBuffer);
out.close();
```

(b) An example that misses exception handling

Figure 1: Two code examples about how to write data to a file using FileChannel on Stack Overflow

Introduction

- RQ1. Is API misuse prevalent on Stack Overflow?
- RQ2. Are highly voted posts more reliable?
- RQ3. What are the characteristics of API misuse?

API Usage Mining from GitHub

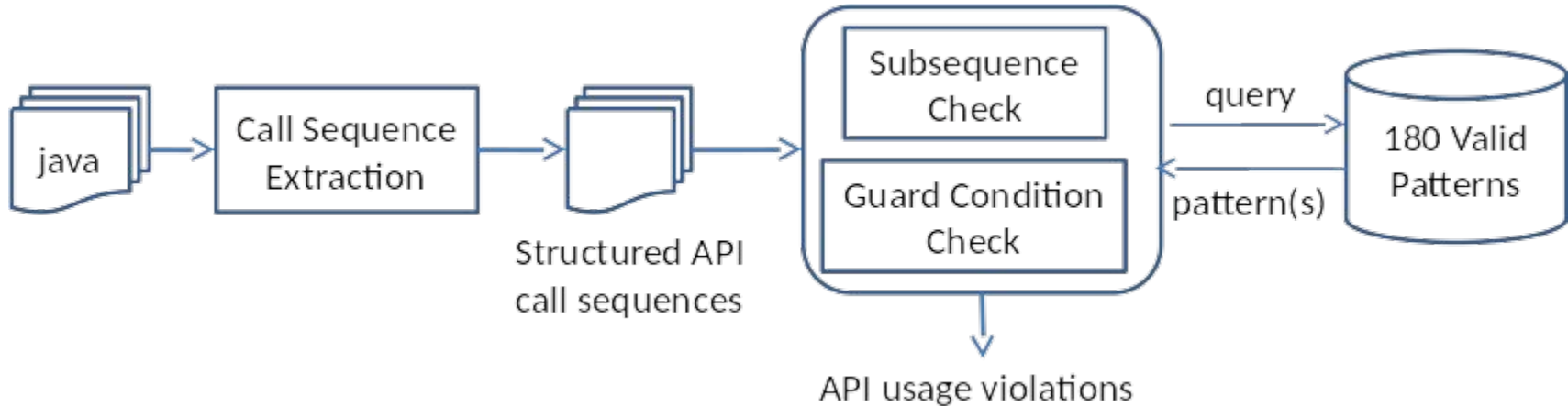
- **Insight 1: Mining a Large Code Corpus**
 - Contrast SO snippets with API usage patterns mined from 380K GitHub projects.
 - Code corpus includes 380K GitHub projects with at least 100 revisions and 2 contributors.
- **Insight 2: Removing Irrelevant Statements via Program Slicing**
 - Perform backward and forward slicing to identify data- and control-dependent statements to an API method of interest.
- **Insight 3: Capture the Semantics of API Usage**
 - It is important to capture the temporal ordering, enclosing control structures, and appropriate guard conditions of API calls.
- **Insight 4: Variations in Guard Conditions**
 - GitHub developers may write the same predicate in different ways.

An Empirical Study of API Misuse on Stack Overflow

- Data Set: API Usage Patterns
 - Learn 245 API usage patterns of 100 Java API methods.
 - Manually inspect these patterns.
 - 180 patterns can be confirmed by online documentation.
- Data Set: SO Posts
 - Extract code snippets in the markdown `<code>` from the SO posts with the Java tag.
 - Parse and analyze these snippets using JavaBaker.
 - Find 217,818 SO posts with code snippets that use the 100 Java API methods in our study scope.

An Empirical Study of API Misuse on Stack Overflow

- Developed an automated API misuse detection technique and examine 220K SO posts with 180 confirmed patterns.



RQ1. Is API Misuse Prevalent on Stack Overflow?

- 31% of SO posts contain API usage violations.
- Two authors independently inspected 400 SO posts with reported API usage violations.
- 289 posts can negatively impact a target program if reused as-is.

RQ3. What are characteristics of API misuse?

- There are three types of API misuse--missing control structures, missing or incorrect order of API calls, and incorrect guard conditions.

```
Cursor emails = db.query(Email.CONTENT_URI,...);  
while (emails.moveToNext()) {  
    String email = emails.getString(..);  
}  
emails.close();
```

```
ByteBuffer bb = ByteBuffer.allocate(4);  
bb.put(newArgb);  
int i = bb.getInt();
```

```
TreeMap map = new TreeMap();  
//OR SortedMap map = new TreeMap();  
map.firstKey();
```

ExampleCheck: Checking API Misuse in Stack Overflow using Patterns Mined from GitHub

The screenshot shows a Stack Overflow question page. A red-bordered pop-up window titled "Potential API Misuse" is overlaid on the right side of the question. The pop-up contains a description of the issue, a code snippet with a try-catch block, and a pagination bar. Red annotations with circled numbers 1 through 6 highlight specific features of the pop-up.

① Pop-up window

② API misuse description

③ Fix suggestion

④ Like or dislike

⑤ Supporting GitHub examples

⑥ Pagination for multiple misuses

```
try {
  if (match_number != null) {
    match_number.getAsToString();
  } catch (Exception e) {
  }
}
```

The screenshot shows the Chrome extension interface for "ExampleCheck". The extension is offered by Tianyi Zhang and has 7 users. The interface includes tabs for "OVERVIEW", "REVIEWS", and "RELATED". The "OVERVIEW" tab is active, showing a preview of the Stack Overflow question and the pop-up window. On the right side, there is a section titled "A Chrome extension that augments Stack Overflow with API usage patterns mined from GitHub" with a note about enabling the extension and a "Report Abuse" link. Below this, there is "Additional Information" including the version (2.0), update date (May 30, 2018), size (617KB), and language (English).

ExampleCheck

offered by Tianyi Zhang

★★★★★ (0) | [Developer Tools](#) | 7 users

ADDED TO CHROME

OVERVIEW | REVIEWS | RELATED

Compatible with your device

A Chrome extension that augments Stack Overflow with API usage patterns mined from GitHub

Note: to use ExampleCheck, please enable to load javascript from untrusted sources in your Chrome browser (see the second screenshot).

ExampleCheck proactively detects API usage violations in a Stack Overflow post and alerts users about potential API misuse. Currently, ExampleCheck includes hundreds of API usage patterns of 100 Java API methods learned from 300K Github repositories.

[Report Abuse](#)

Additional Information

Version: 2.0

Updated: May 30, 2018

Size: 617KB

Language: English

Related Work

- **Quality Assessment of Online Code Examples.**
 - Prior work has investigated the quality of online code examples from different perspectives.
 - Security, copyright issues, survey to investigate comprehension difficulty of code examples in SO.

- **API Usage Mining.**
 - There is a large body of literature in mining implicit programming rules, API usage, and temporal properties of API calls.

Discussion

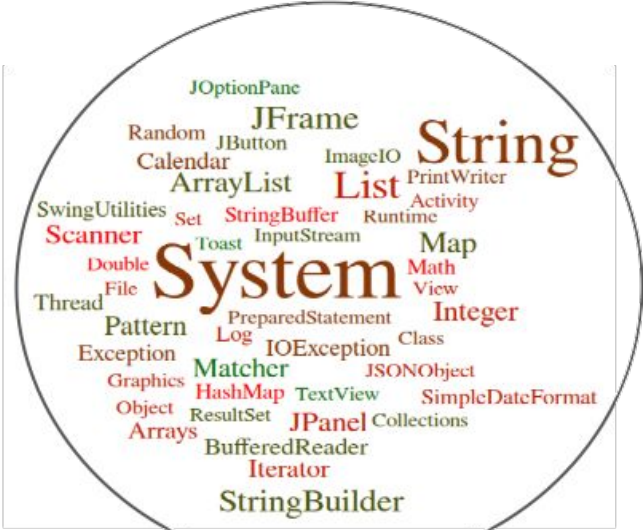
- Presented an API usage mining technique that extracts patterns from over 380K GitHub projects.
- A large-scale empirical study of API misuse in 220K SO posts.
- A Chrome extension that augments Stack Overflow with API usage patterns mined from GitHub.

Discussion

- **Threats to Validity.**

- This study is limited to 100 Java APIs that frequently appear in Stack Overflow and thus may not generalize to other Java APIs or different languages.
- Scope is limited to code snippets found on Stack Overflow.
- Other types of online resources such as programming blogs and other Q&A forums may have better curated examples.
- ExampleCheck may overlook or mis-identify API misuses.

Doubts?



e.g., Java APIs