

Stack Overflow: A Code Laundering Platform?

Le An, Ons Mlouki, Foutse Khomh, and Giuliano Antoniol
Polytechnique Montréal

SANER, 24th February 2017, Klagenfurt, Austria



POLYTECHNIQUE
MONTREAL

WORLD-CLASS
ENGINEERING



Papers on Code Reuse & Recommendation From SANER '17

- *Detecting Similar Repositories on GitHub*, Yun Zhang, David Lo, Pavneet Singh Kochhar, Xin Xia, Quanlai Li, and Jianling Sun
- *StiCProb: A Novel Feature Mining Approach using Conditional Probability*, Yutian Tang and Hareton Leung
- *A Framework for Classifying and Comparing Source Code Recommendation Systems*, Mohammad Ghafari and Hamidreza Moradi

Software Licenses ...

“govern the use or redistribution of software”

```
/*
 * Copyright (C) 2007 The Android Open Source Project
 *
 * Licensed under the Apache License, Version 2.0 (the "License");
 * you may not use this file except in compliance with the License.
 * You may obtain a copy of the License at
 *
 *     http://www.apache.org/licenses/LICENSE-2.0
 *
 * Unless required by applicable law or agreed to in writing, software
 * distributed under the License is distributed on an "AS IS" BASIS,
 * WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
 * See the License for the specific language governing permissions and
 * limitations under the License.
 */
```

Software Licenses ...

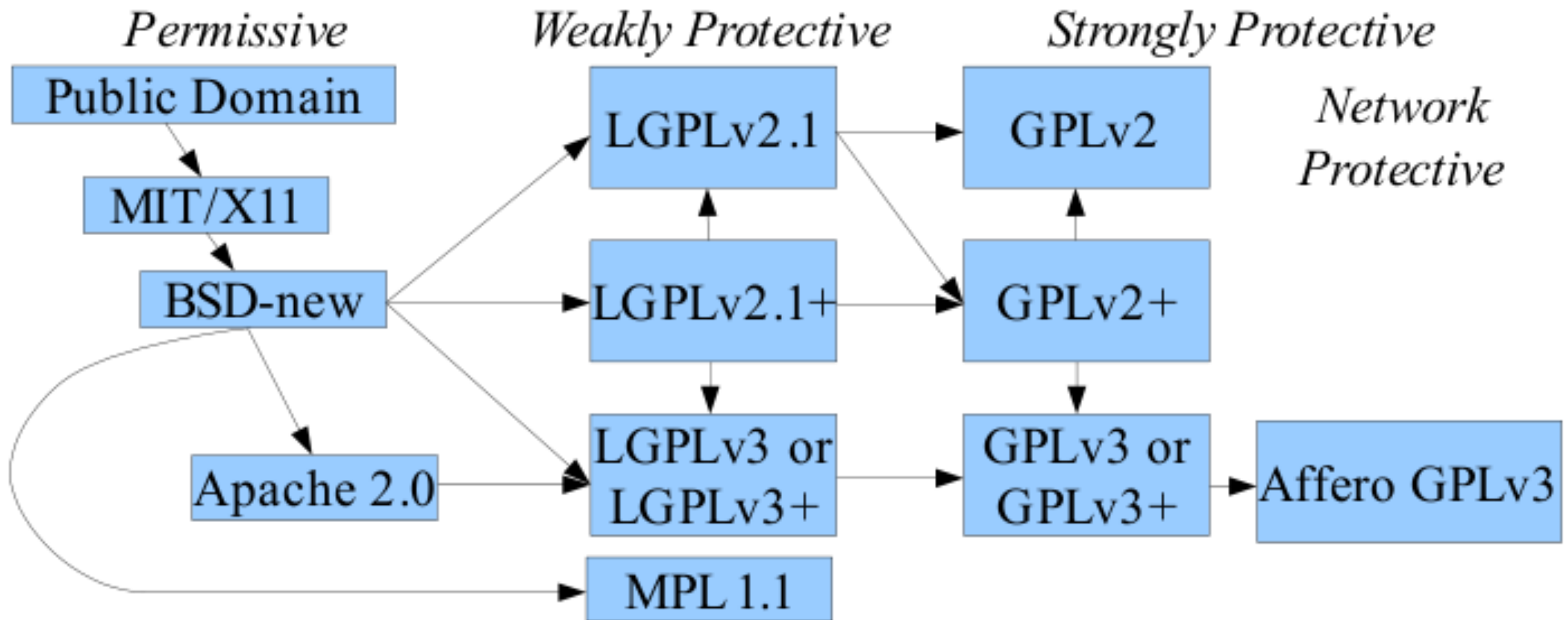
“govern the use or redistribution of software”

```
/*
 * Copyright (C) 2007 The Android Open Source Project
 *
 * Licensed under the Apache License, Version 2.0 (the "License");
 * you may not use this file except in compliance with the License.
 * You may obtain a copy of the License at
 *
 *     http://www.apache.org/licenses/LICENSE-2.0
 *
 * Unless required by applicable law or agreed to in writing, software
 * distributed under the License is distributed on an "AS IS" BASIS,
 * WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
 * See the License for the specific language governing permissions and
 * limitations under the License.
 */
```

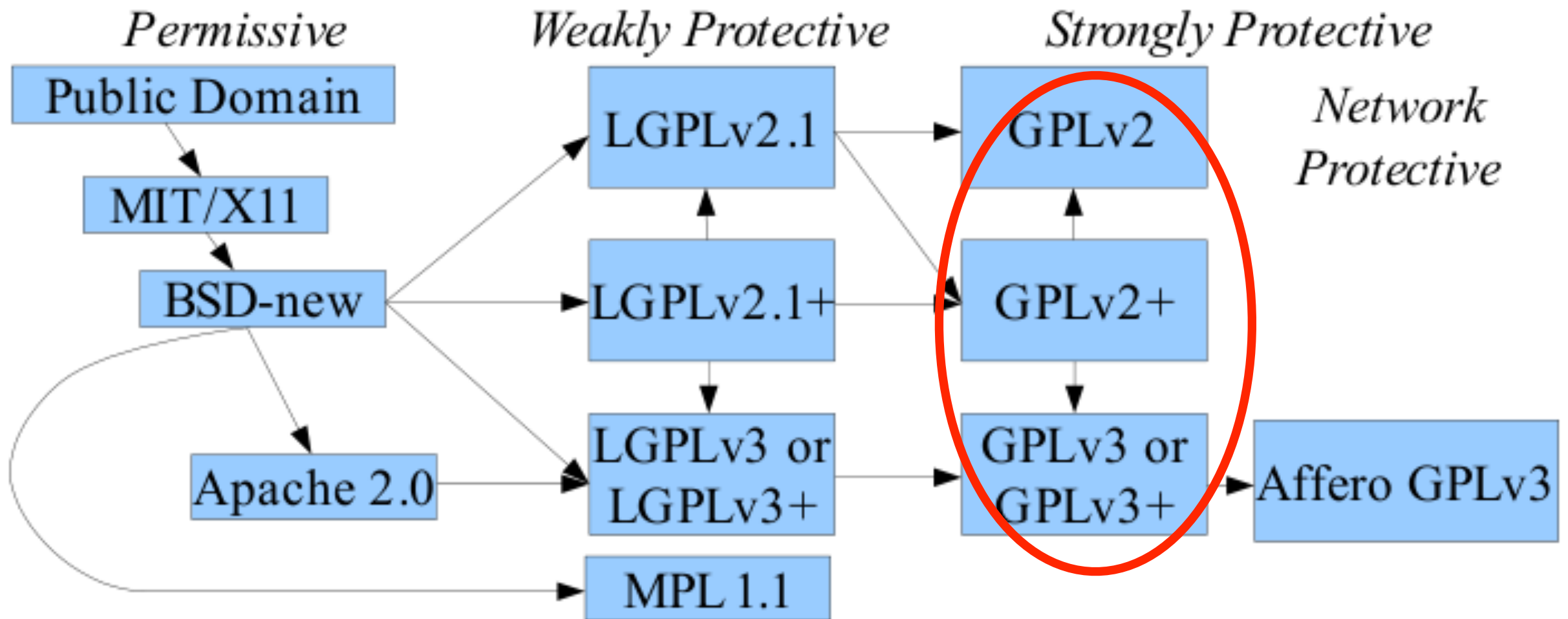


license violation

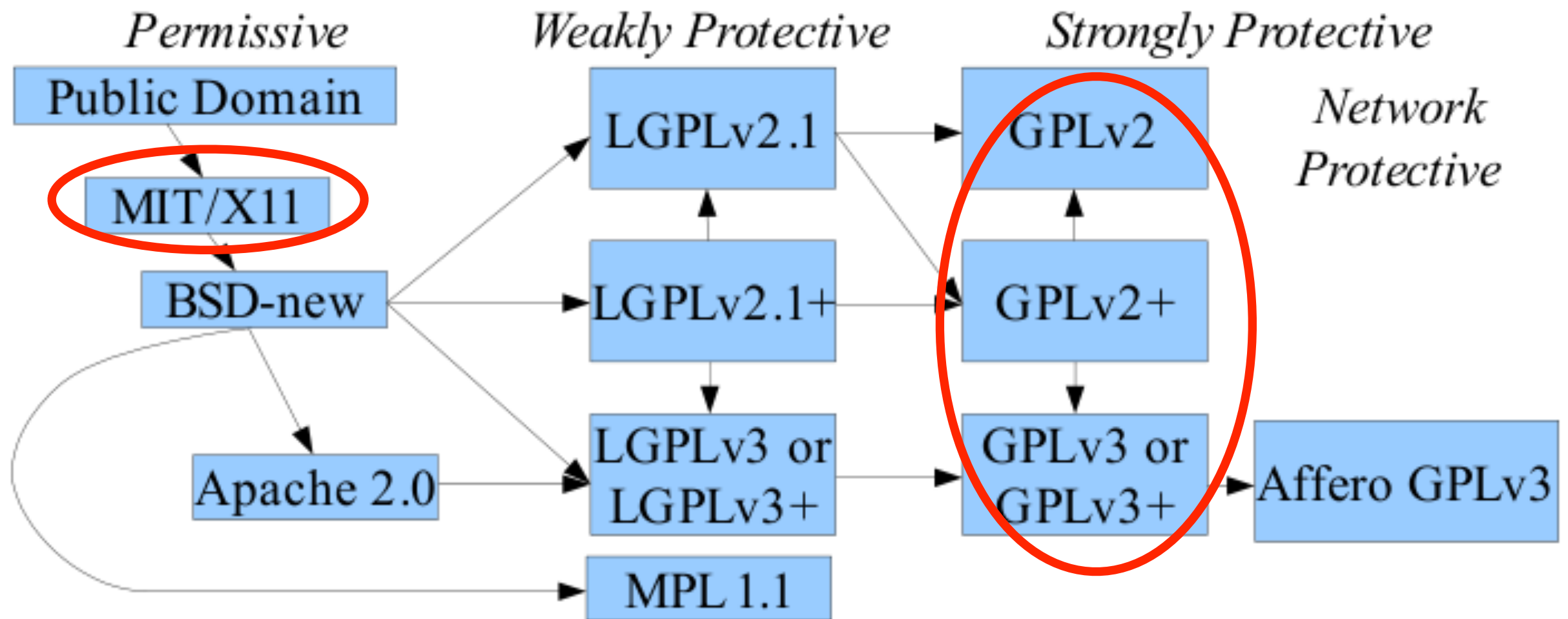
Software License Compatibility

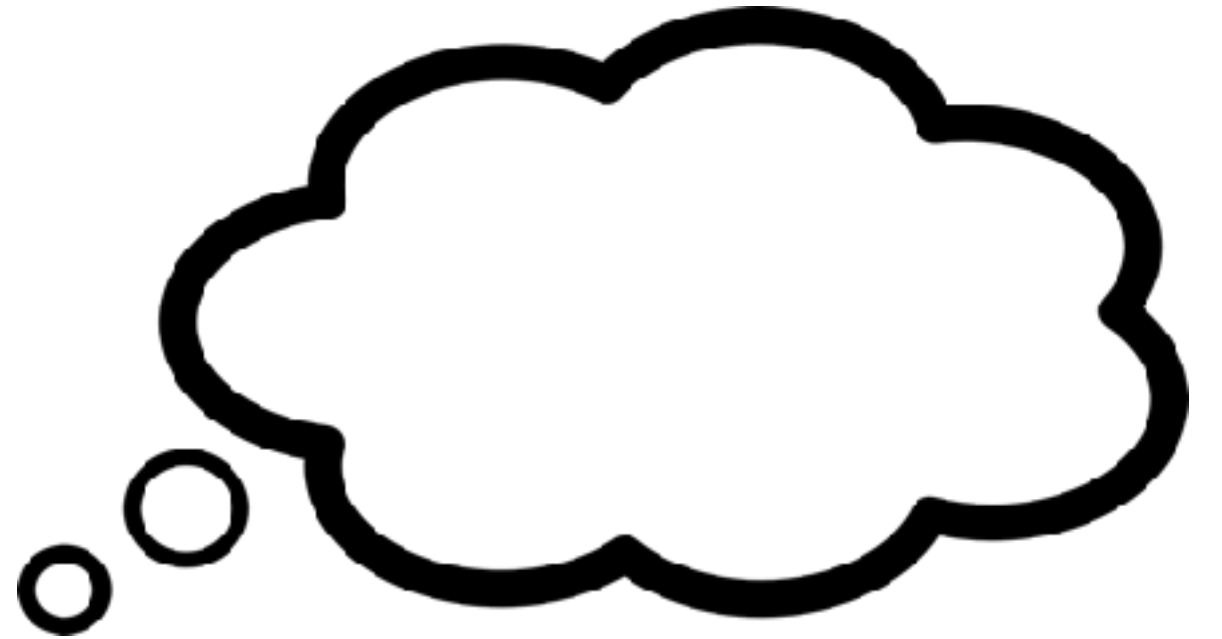


Software License Compatibility

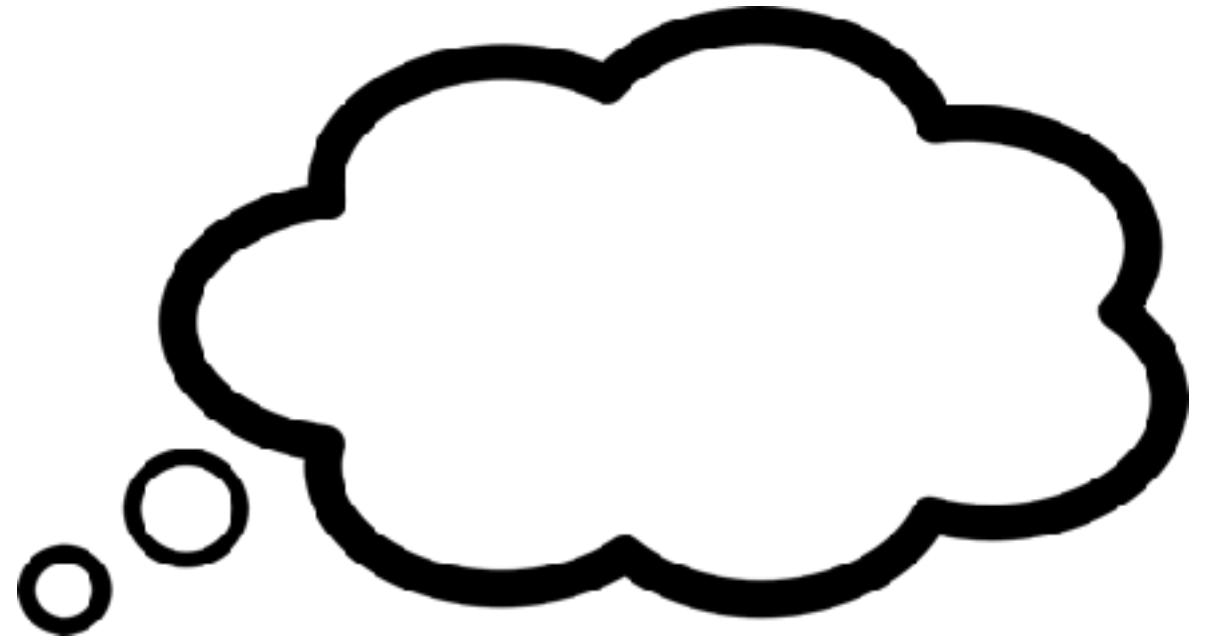


Software License Compatibility



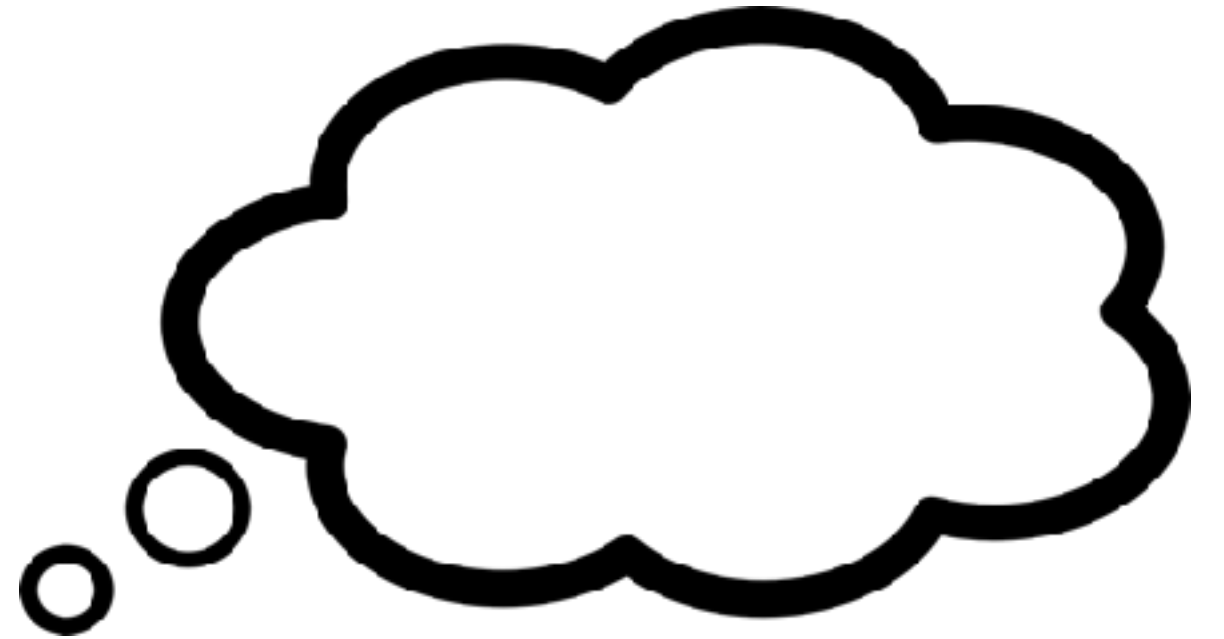


What if we reuse code
from a Q&A website?



What if we reuse code
from a Q&A website?

Quora



What if we reuse code
from a Q&A website?

Quora





What if we reuse code
from a Q&A website?

Quora



**What are the restrictions of reusing
the content of Stack Overflow?**



Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0)

This is a human-readable summary of (and not a substitute for) the [license](#).

[Disclaimer](#)



You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material

for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:



Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.



ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.



Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0)

This is a human-readable summary of (and not a substitute for) the [license](#).

[Disclaimer](#)



You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material

for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:



Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.



ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

- **Attribution:** You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
- **ShareAlike:** If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

- **Attribution:** You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
- **ShareAlike:** If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

Research Objective



Android apps



Stack Overflow

Research objective



Android apps



Stack Overflow

Research Questions

- RQ1: Do developers release apps with code copied from Stack Overflow?
- RQ2: Do developers respect the copyright terms of code reused from Stack Overflow?
- RQ3: Do Stack Overflow users respect copyright terms when publishing code snippets on Stack Overflow?
- RQ4: How long does a Stack Overflow code remain in released versions of an app?

Data Mining



399 Android apps
(79.2k Java files, 1.4GB)



Stack Overflow posts
(2.1M Java & Android
snippets, 8.6GB)

Data Mining

NiCAD
Clone
Detection
Tool



399 Android apps
(79.2k Java files, 1.4GB)

Stack Overflow posts
(2.1M Java & Android
snippets, 8.6GB)

Comparing the Creation Dates of an Android Snippet vs. a Stack Overflow Snippet

```
public class HelloWorld {  
    public static void main(String[] args) {  
        System.out.println("Hello World");  
    }  
}
```

The screenshot shows the SourceTree application interface. The top toolbar includes buttons for View, Commit, Update, Revert, Stash, Add, Remove, Add/Remove, Pull, Push, Branch, Merge, Tag, Show in Finder, Hg Flow, Terminal, and Settings. The left sidebar shows the FILE STATUS (Working Copy) and BRANCHES (1.7.x, 1.8.x, 1.9.x, 2.0.0-beta, 2.0.x, 2.1.x, etc.). The main area displays a commit history table with columns for Graph, Description, Revision, Author, and Date. The selected commit is Revision 6394 by Manjunath Basara L. on Dec 3, 2015. Below the table, the diff view shows changes to the file SourceTree.xcodeproj/project.pbxproj. The diff is divided into two hunks: Hunk 1 (Lines 1145-1151) and Hunk 2 (Lines 1230-1243). The diff shows changes to the STAccountTaskManager.m file, including the addition of a new class and the removal of an existing class.

Graph	Description	Revision	Author	Date
dp	Head-Identify reflecting changes - installment 1	6394	Manjunath Basara L.	Dec 3, 2015, 3:39...
[S]	Created new branch SRECTREEDEV-269-create-l...	6393	Brian Ganninger c...	Dec 3, 2015, 11:1...
[g]	This LFS check in fetchBinaryPreview is unnecessary, callers check	6392	Stewie Strouting vs...	Dec 3, 2015, 8:13...
[g]	Non-functional attempt at optimizing LFS previews	6391	Stewie Strouting vs...	Dec 3, 2015, 0:0...
[g]	Don't add hunk view components when non-text, can override LFS previews	6390	Stewie Strouting vs...	Dec 3, 2015, 7:52...
[g]	Doh, file type not changeType	6389	Stewie Strouting vs...	Dec 3, 2015, 7:56...
[g]	Support for detecting git-lfs content while parsing the diff	6388	Stewie Strouting vs...	Dec 3, 2015, 3:41...
[g]	Updated colors, table rows, borders and padding.	6387	Joel Unger Kump...	Dec 2, 2015, 5:21...
[g]	Creates 2.1.x branch from default	6386	Brian Ganninger c...	Dec 2, 2015, 3:44...
[g]	Merged in graph-line-tweak (pull request #29)	6385	Brian Ganninger...	Dec 2, 2015, 3:2...
[g]	Close branch graph-line-tweak	6384	Brian Ganninger c...	Dec 2, 2015, 2:21...

Sorted by path

SourceTree.xcodeproj/project.pbxproj

Hunk 1: Lines 1145-1151

Reverse hunk

1145 1145 AAC555114704218802F3804 /* security.framework in Frameworks */
1146 1146 AAC55511470625002F3804 /* com.atlassian.SourceTreePrivileged
1147 1147 AAC55511470625002F3804 /* STAccountTaskManager.m in Sources */
1148 1148 AAC55511470625002F3804 /* STAccountTaskManager.m in Sources */
1149 1149 AAC55511470625002F3804 /* STAccountTaskManager.m in Sources */
1150 1150 AAC55511470625002F3804 /* STAccountTaskManager.m in Sources */
1151 1151 AAC55511470625002F3804 /* STAccountTaskManager.m in Sources */

Hunk 2: Lines 1230-1243

Reverse hunk

1230 1230 D2C62201C0A22602D0C0F /* AppDelegate.m in Sources */
1231 1231 D2C62201C0A22602D0C0F /* AppDelegate.strings in Resources
1232 1232 D2C62201C0A22602D0C0F /* AppDelegate.m in Sources */
1233 1233 D2C62201C0A22602D0C0F /* AppDelegate.m in Sources */
1234 1234 D2C62201C0A22602D0C0F /* AppDelegate.m in Sources */

Revision: 6394
Changeset: bb278dbb2dc0f8c4f1b7068477787...
Parents: 6377
Author: Manjunath Basara L. Srinivasa <mba...>
Date: December 3, 2015 at 3:39:27 PM PST

Atlassian

Comparing the Creation Dates of an Android Snippet vs. a Stack Overflow Snippet

```
public class HelloWorld {  
    public static void main(String[] args) {  
        System.out.println("Hello World");  
    }  
}
```

The screenshot shows the SourceTree application interface. The top toolbar includes buttons for View, Commit, Update, Revert, Stash, Add, Remove, Add/Remove, Pull, Push, Branch, Merge, Tag, Show in Finder, Hg Flow, Terminal, and Settings. The left sidebar shows the 'FILE STATUS' with 'Working Copy' and a list of 'BRANCHES' including 1.7.x, 1.8.x, 1.9.x, 2.0.0.0-beta, 2.0.x, 2.1.x, and default. The main area displays a 'Graph' of commit history with a table of revisions. Below the graph, a 'Sorted by path' dropdown is visible. The bottom right pane shows a 'Diff' view for 'SourceTree.xcodeproj/project.pbxproj', highlighting changes in 'src/common/STConstants.m' and 'src/hosting/STAccountTaskManager.m'.

Revision	Author	Date
6394	Manjunath Basara	Dec 3, 2015, 3:39...
6393	Brian Ganninger	Dec 3, 2015, 11:1...
6392	Stewie Strouting	Dec 3, 2015, 8:13...
6391	Stewie Strouting	Dec 3, 2015, 0:0...
6390	Stewie Strouting	Dec 3, 2015, 7:52...
6389	Stewie Strouting	Dec 3, 2015, 7:56...
6388	Stewie Strouting	Dec 3, 2015, 3:41...
6387	Joel Unger	Dec 2, 2015, 3:21...
6386	Brian Ganninger	Dec 2, 2015, 3:44...
6385	Brian Ganninger	Dec 2, 2015, 3:2...
6384	Brian Ganninger	Dec 2, 2015, 3:21...

Revision: 6394
Changeset: bb278dbb2de0b6c4f1b7068477787...
Parents: 6377
Author: Manjunath Basara Srinivasa <mba...>
Date: December 3, 2015 at 3:39:27 PM PST

Manually compare the code again the file's history, and find the earliest introducing date.

RQ1: Do developers release apps with code copied from Stack Overflow?

- 232 Android code snippets are similar to the code in Stack Overflow
- These code snippets are distributed in 135 files from 62 (15.5%) different apps

RQ2: Do developers respect the copyright terms of code reused from Stack Overflow?

(*i.e.*, appropriately using CC-BY-SA license)

- None of the 135 Android files that potentially reused code from Stack Overflow were released under the CC-BY-SA 3.0 or its later versions.
- We found only 2 Stack Overflow posters' names in the list of contributors of the corresponding apps.

Android Developers' Comments

“There is definitely code in our project that is copy-pasted from Stack Overflow, as I have done this several times.”

Android Developers' Comments

“There is definitely code in our project that is copy-pasted from Stack Overflow, as I have done this several times.”

“If I were to never look at code examples, and only write code from reading the APIs, I would probably miss elegant solutions and overlook important pitfalls.”

Android Developers' Comments

“There is definitely code in our project that is copy-pasted from Stack Overflow, as I have done this several times.”

“If I were to never look at code examples, and only write code from reading the APIs, I would probably miss elegant solutions and overlook important pitfalls.”

“We don't have any policy about that. Now might be a good time to have that discussion.”

Android Developers' Comments

“There is definitely code in our project that is copy-pasted from Stack Overflow, as I have done this several times.”

“If I were to never look at code examples, and only write code from reading the APIs, I would probably miss elegant solutions and overlook important pitfalls.”

“We don’t have any policy about that. Now might be a good time to have that discussion.”

“(Stack Overflow’s license) appears to not be compatible with the GNU_GPL ... I hope the staff at Stack Overflow will address the problem.”

RQ3: Do Stack Overflow users respect copyright terms when publishing code fragments on Stack Overflow?

- 1,226 Stack Overflow posts containing code snippets potentially reused from 68 Android apps.
- 5 of the posts provide the apps' license.
- 2 poster names are found in the apps' contributor list.

RQ3: Do Stack Overflow users respect copyright terms when publishing code fragments on Stack Overflow?

- 1,226 Stack Overflow posts containing code snippets potentially reused from 68 Android apps.
- 5 of the posts provide the apps' license.
- 2 poster names are found in the apps' contributor list.

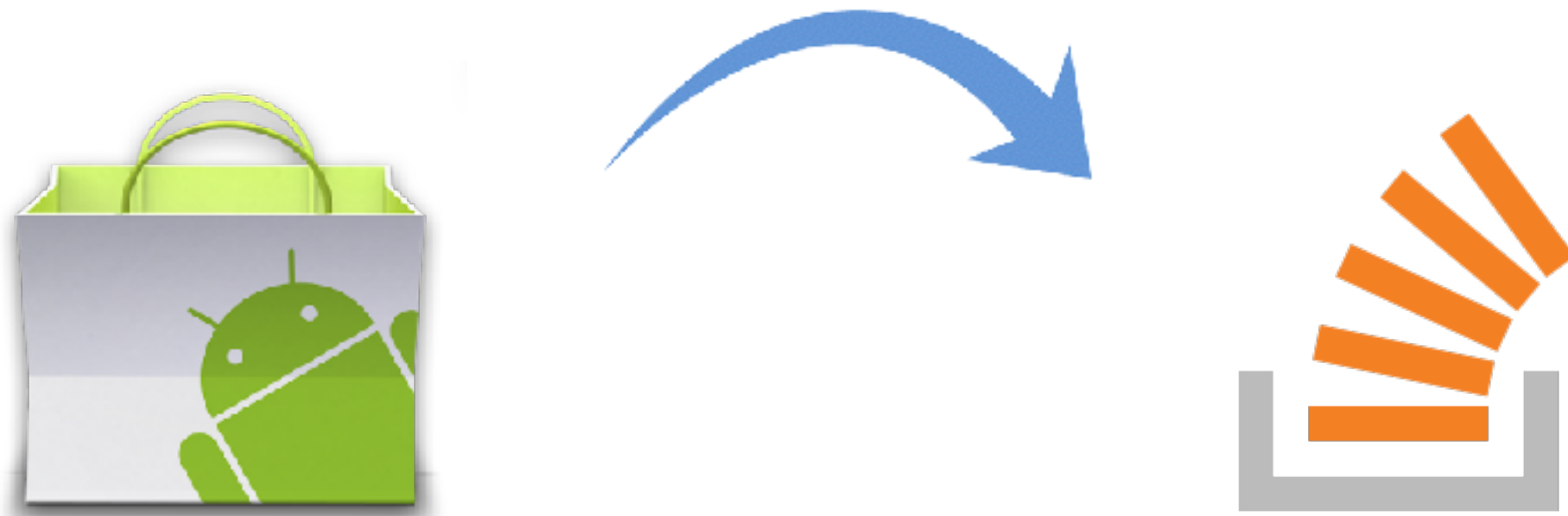


1,219 posts have a risk of license violation.

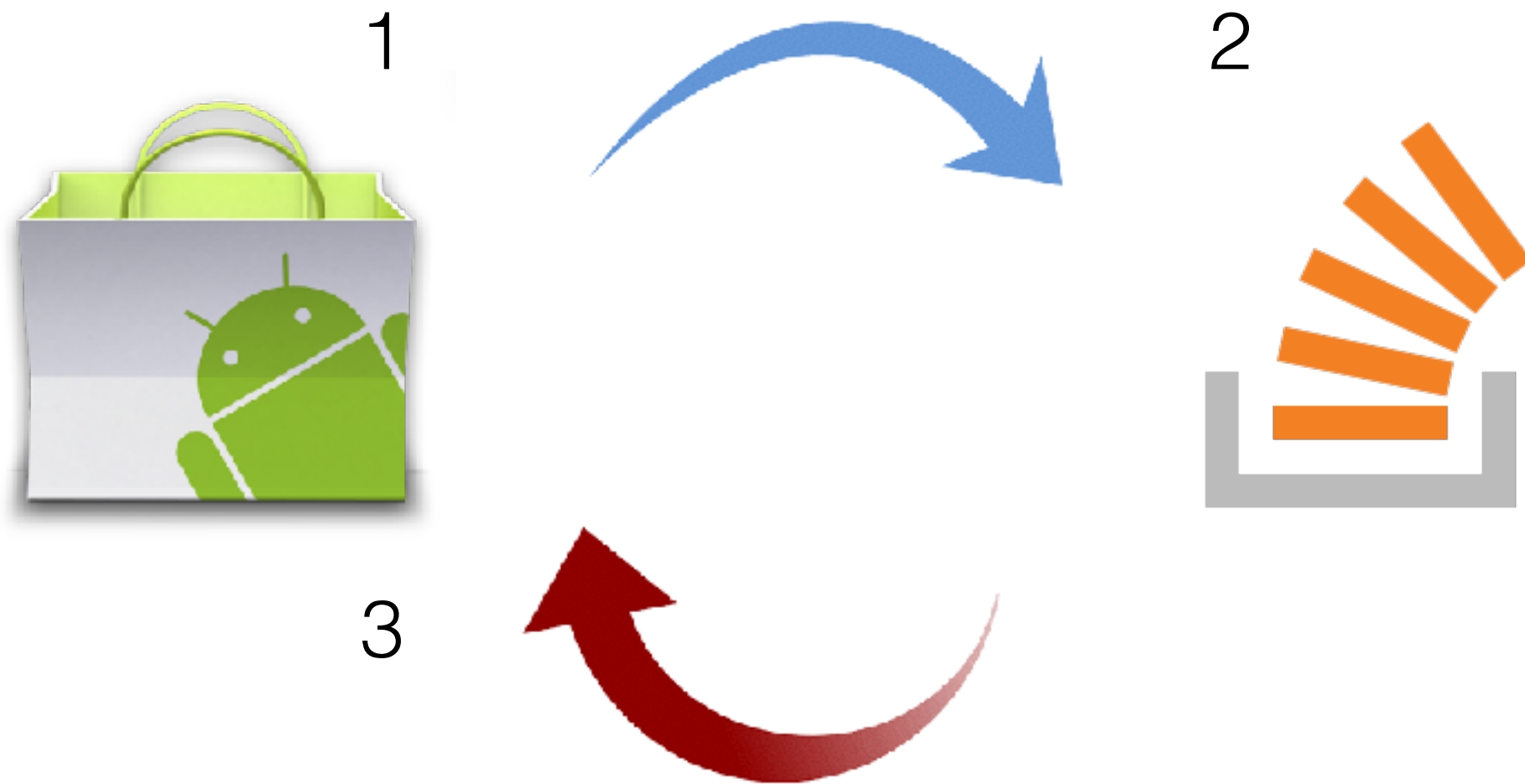
Using Stack Overflow as a “Code Laundering Platform”?



Using Stack Overflow as a “Code Laundering Platform”?



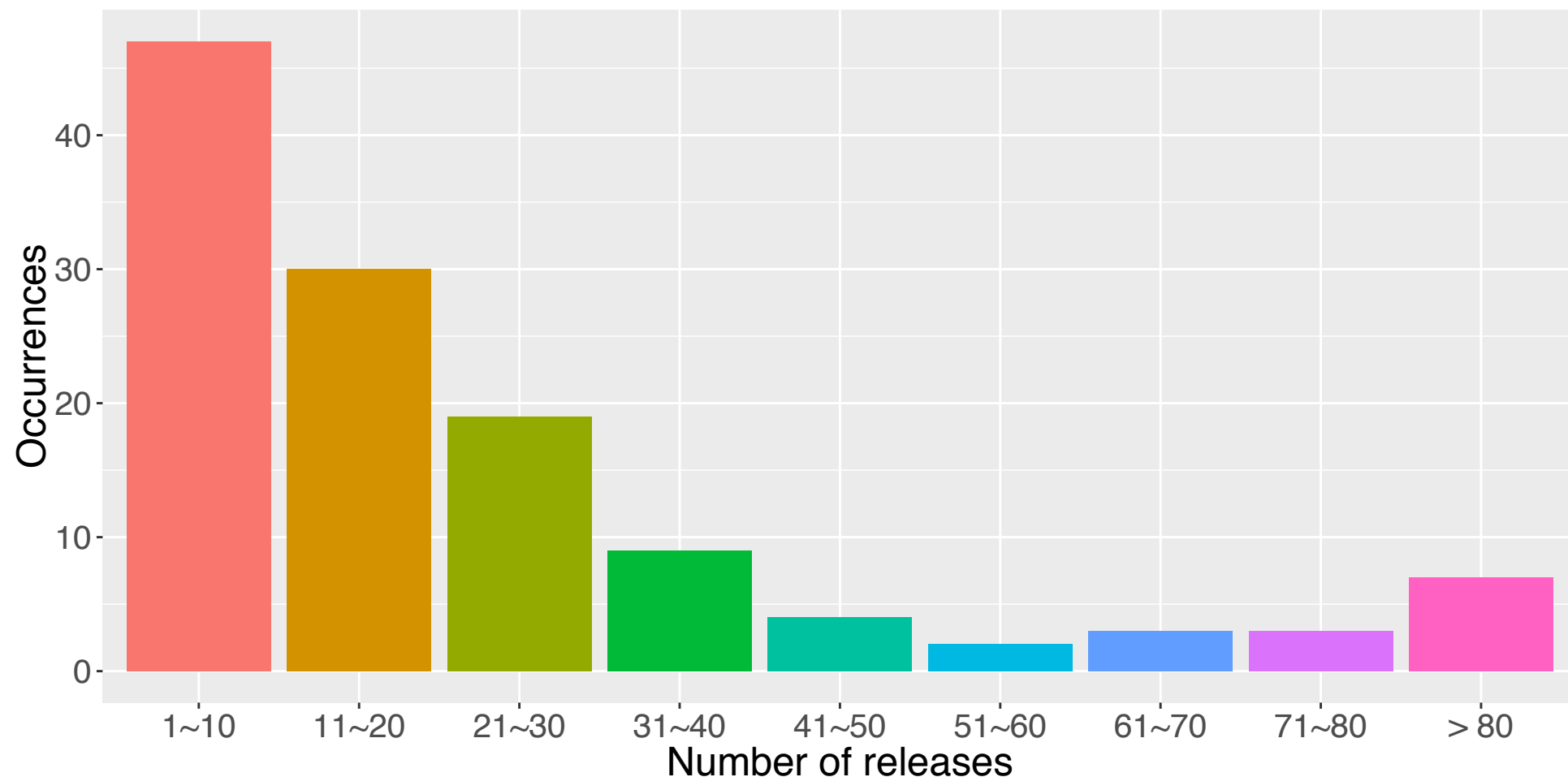
Using Stack Overflow as a “Code Laundering Platform”?



Using Stack Overflow as a “Code Laundering Platform”?

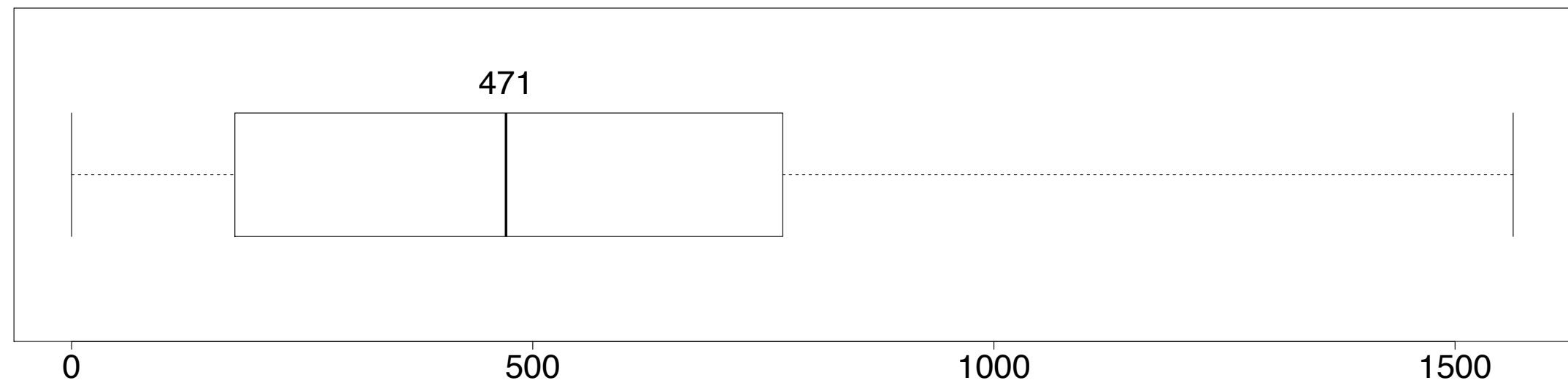
- 126 code snippets that seems to have migrated from one app to Stack Overflow, then to another app.
- In 12 of these snippets, the software licenses in the 1st app and in the 2nd app are different.

RQ4: How long (releases) does the Stack Overflow code remain in released versions of an app?

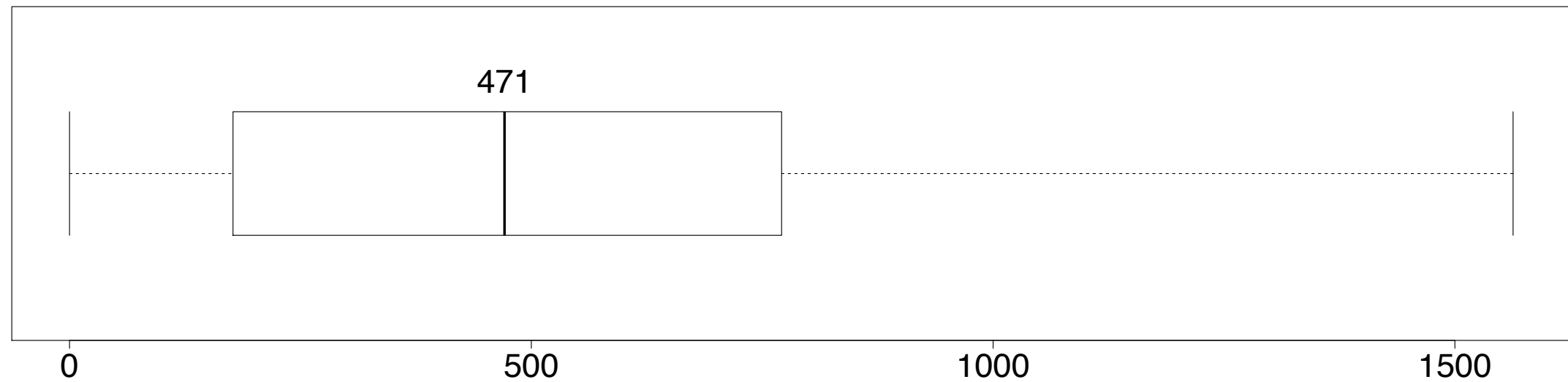


62% Stack Overflow snippets remained in the app for up to 20 releases

RQ4: How long (days) does the Stack Overflow code remain in released versions of an app?

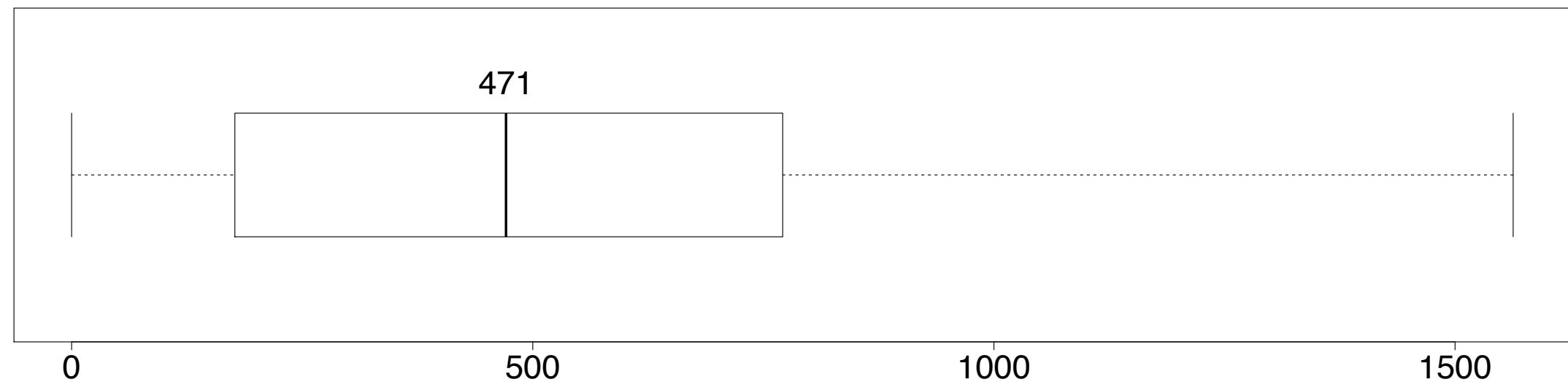


RQ4: How long (days) does the Stack Overflow code remain in released versions of an app?



*Stack Overflow code
tend to stay long time in
the Android apps*

RQ4: How long (days) does the Stack Overflow code remain in released versions of an app?



Stack Overflow code tend to stay long time in the Android apps

Developers may not realize the potential risk of license violations



Android apps



Stack Overflow



Android apps



Stack Overflow





Android apps



Stack Overflow



1



2



3



Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0)

This is a human-readable summary of (and not a substitute for) the [license](#).

[Disclaimer](#)



You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material

for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Not Well Respected



Android apps



Stack Overflow



1



2



3



Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0)

This is a human-readable summary of (and not a substitute for) the [license](#).

[Disclaimer](#)



You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material

for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Not Well Respected



Developers and software organizations
need to raise awareness on the copyright terms
when reusing code from the Internet



Android apps



Stack Overflow



1



2



3



Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0)

This is a human-readable summary of (and not a substitute for) the [license](#).

[Disclaimer](#)



You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material

for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Not Well Respected

Developers and software organizations
need to raise awareness on the copyright terms
when reusing code from the Internet