

Measuring the Effects of Stack Overflow Code Snippet Evolution on Open-Source Software Security

Alfusainey Jallow, Michael Schilling, Michael Backes, Sven Bugiel

45th IEEE Symposium on Security and Privacy | May 2024



Motivating Example



**Lucy, a software developer
in the year 2011**



Motivating Example

JUnit 5

JWT
JSON WEB Token

MySQL



**Lucy, a software developer
in the year 2011**



Motivating Example

JUnit 5

JWT
JSON WEB Token

MySQL



Lucy, a software developer
in the year 2011

“How can I convert XML to a String object?”

```
String unformattedXml = "<order><number>1</number>  
<customer_code>C7854</customer_code><product_code>CDR1080</  
product_code><description>CD</description><quantity>10</  
quantity></order>
```



```
<?xml version="1.0" encoding="UTF-8"?>  
<order>  
  <number>1</number>  
  <customer_code>C7854</customer_code>  
  <product_code>CDR1080</product_code>  
  <description>CD</description>  
  <quantity>10</quantity>  
</order>
```



Motivating Example

JUnit 5

JWT
JSON WEB Token

MySQL



Lucy, a software developer
in the year 2011

“How can I convert XML to a String object?”

```
String unformattedXml = "<order><number>1</number>  
<customer_code>C7854</customer_code><product_code>CDR1080</  
product_code><description>CD</description><quantity>10</  
quantity></order>
```



```
<?xml version="1.0" encoding="UTF-8"?>  
<order>  
  <number>1</number>  
  <customer_code>C7854</customer_code>  
  <product_code>CDR1080</product_code>  
  <description>CD</description>  
  <quantity>10</quantity>  
</order>
```

“I sure do not need a whole library for that!”



Motivating Example

How to pretty XML from Java



All

Videos

Images

News

Books

: More

Tools



Stack Overflow

<https://stackoverflow.com> › questions › how-to-pretty-p... ⋮

How to pretty print XML from Java?

Insert a newline and spaces after every characters, keep and **indent** counter (to determine the number of spaces) that you increment for every <..

[34 answers](#) · Top answer: `Transformer transformer = TransformerFactory.newInstance().newTr...`

[Pretty print XML in java 8 - Stack Overflow](#)

16 Sept 2014

[JAVA pretty print XML with properly formatted comments](#)

9 Oct 2018

[How to prettify XML String in Java - Stack Overflow](#)

23 Feb 2016

[how to generate formatted .xml file? - java - Stack Overflow](#)

1 Dec 2015

[More results from stackoverflow.com](#)



Motivating Example



[a simpler solution based on this answer:](#)

145



```
public static String prettyFormat(String input, int indent) {
    try {
        Source xmlInput = new StreamSource(new StringReader(input));
        StringWriter stringWriter = new StringWriter();
        StreamResult xmlOutput = new StreamResult(stringWriter);
        TransformerTransformerFactory transformerFactory = TransformerFactory
            transformerFactory.newTransformersetAttribute("indent-number", indent);
        Transformer transformer = transformerFactory.setOutputPropertynewTransformer(Out
            transformer.setOutputProperty("{http://xml.apache.org/xslt}indent-amou
            transformer.transform(xmlInput, xmlOutput);
        return xmlOutput.getWriter().toString();
    } catch (Exception e) {
        throw new RuntimeException(e); // simple exception handling, please review it
    }
}
```




StackOverflow snippets might contain vulnerabilities

Stack Overflow Considered Harmful? The Impact of Copy&Paste on Android Application Security

Felix Fischer, Konstantin Böttinger, Huang Xiao, Christian Stransky*, Yasemin Acar*, Michael Backes*, Sascha Fahl*
Fraunhofer Institute for Applied and Integrated Security; *CISPA, Saarland University

The most copied StackOverflow snippet of all time is flawed!

by Andreas Lundblad, 2019-12-02

In a recent study titled *Usage and Attribution of Stack Overflow Code Snippets in GitHub Projects*, an [answer](#) I wrote almost a decade ago was found to be the most copied snippet on Stack Overflow. Ironically it happens to be buggy.

Snakes in Paradise?: Insecure Python-related Coding Practices in Stack Overflow

Akond Rahman, Effat Farhana, and Nasif Imtiaz
North Carolina State University, Raleigh, North Carolina
Email: aarahman@ncsu.edu, efarhan@ncsu.edu, simtiaz@ncsu.edu

Abstract—Despite being the most popular question and answer website for software developers, answers posted on Stack Overflow (SO) are susceptible to contain Python-related insecure coding practices. A systematic analysis on how frequently insecure coding practices appear in SO answers can help the SO community assess the prevalence of insecure Python code blocks

▲ This [Recipe](#) provides a nice function to do what you are asking. I've modified it to use the MD5 hash, instead of the SHA1, as your original question asks
6 ▼

```
def GetHashofDirs(directory, verbose=0):  
    import hashlib, os
```

You Get Where You're Looking For The Impact of Information Sources on Code Security

Yasemin Acar, Michael Backes, Sascha Fahl, Doowon Kim[†], Michelle L. Mazurek[†], Christian Stransky
CISPA, Saarland University; [†]University of Maryland, College Park



Motivating Example (cont.)

```
public static String prettyFormat(String input, int indent) {  
    try {  
        Source xmlInput = new StreamSource(new StringReader(input));  
        StringWriter stringWriter = new StringWriter();  
        StreamResult xmlOutput = new StreamResult(stringWriter);  
        TransformerFactory transformerFactory = TransformerFactory.  
            newInstance("indent-number", indent);  
        Transformer transformer = transformerFactory.newTransformer(Out  
            transformer.setOutputProperty("{http://xml.apache.org/xslt}indent-amou  
            transformer.transform(xmlInput, xmlOutput);  
        return xmlOutput.getWriter().toString();  
    } catch (Exception e) {  
        throw new RuntimeException(e); // simple exception handling, please review it  
    }  
}
```



Motivating Example (cont.)

```
public static String prettyFormat(String input, int indent) {  
    try {  
        Source xmlInput = new StreamSource(new StringReader(input));  
        StringWriter stringWriter = new StringWriter();  
        StreamResult xmlOutput = new StreamResult(stringWriter);  
        TransformerFactory transformerFactory = TransformerFactory.  
            newInstance("indent-number", indent);  
        Transformer transformer = transformerFactory.newTransformer(Out  
            transformer.setOutputProperty("{http://xml.apache.org/xslt}indent-amou  
            transformer.transform(xmlInput, xmlOutput);  
        return xmlOutput.getWriter().toString();  
    } catch (Exception e) {  
        throw new RuntimeException(e); // simple exception handling, please review it  
    }  
}
```



This code snippet is vulnerable to XML eXternal Entity Injection (XXE).

See: cheatsheetseries.owasp.org/cheatsheets/... – fanbondi Feb 9, 2021 at

14:54



Code Fixes on StackOverflow

✓ 4

Fix XXE vulnerability as mentioned in the comments

Source Link

edit approved Mar 19, 2021 at 22:14



fanbondi

1k ● 5 ● 19 ● 37

Inline

Side-by-side

Side-by-side Markdown

[a simpler solution based on this answer:](#)

```
public static String prettyFormat(String input, int indent) {
    try {
        Source xmlInput = new StreamSource(new StringReader(input));
        StringWriter stringWriter = new StringWriter();
        StreamResult xmlOutput = new StreamResult(stringWriter);
        TransformerFactory transformerFactory = TransformerFactory.newInstance();
        transformerFactory.setAttribute("indent-number", indent);
        Transformer transformer = transformerFactory.newTransformer();
        transformer.setOutputProperty(OutputKeys.INDENT, "yes");
        transformer.transform(xmlInput, xmlOutput);
        return xmlOutput.getWriter().toString();
    } catch (Exception e) {
        throw new RuntimeException(e); // simple exception handling
    }
}

public static String prettyFormat(String input) {
    return prettyFormat(input, 2);
}
```

testcase:

[a simpler solution based on this answer:](#)


```
public static String prettyFormat(String input, int indent) {
    try {
        Source xmlInput = new StreamSource(new StringReader(input));
        StringWriter stringWriter = new StringWriter();
        StreamResult xmlOutput = new StreamResult(stringWriter);
        TransformerFactory transformerFactory = TransformerFactory.newInstance();
        transformerFactory.setAttribute("indent-number", indent);
        transformerFactory.setAttribute(XMLConstant.ACCESS_EXTERNAL_ALL, "");
        Transformer transformer = transformerFactory.newTransformer();
        transformer.setOutputProperty(OutputKeys.INDENT, "yes");
        transformer.transform(xmlInput, xmlOutput);
        return xmlOutput.getWriter().toString();
    } catch (Exception e) {
        throw new RuntimeException(e); // simple exception handling
    }
}

public static String prettyFormat(String input) {
    return prettyFormat(input, 2);
}
```



Code Fixes on StackOverflow


4th revision of this post

 **4**

Fix XXE vulnerability as mentioned in the comments

[Source](#) [Link](#)

edit approved Mar 19, 2021 at 22:14

 fanbondi

1k ● 5 ● 19 ● 37

[Inline](#) **[Side-by-side](#)** [Side-by-side Markdown](#)

[a simpler solution based on this answer:](#)

```
public static String prettyFormat(String input, int indent) {
    try {
        Source xmlInput = new StreamSource(new StringReader(input));
        StringWriter stringWriter = new StringWriter();
        StreamResult xmlOutput = new StreamResult(stringWriter);
        TransformerFactory transformerFactory = TransformerFactory.newInstance();
        transformerFactory.setAttribute("indent-number", indent);
        Transformer transformer = transformerFactory.newTransformer();
        transformer.setOutputProperty(OutputKeys.INDENT, "yes");
        transformer.transform(xmlInput, xmlOutput);
        return xmlOutput.getWriter().toString();
    } catch (Exception e) {
        throw new RuntimeException(e); // simple exception handling
    }
}

public static String prettyFormat(String input) {
    return prettyFormat(input, 2);
}
```

testcase:

[a simpler solution based on this answer:](#)

```
public static String prettyFormat(String input, int indent) {
    try {
        Source xmlInput = new StreamSource(new StringReader(input));
        StringWriter stringWriter = new StringWriter();
        StreamResult xmlOutput = new StreamResult(stringWriter);
        TransformerFactory transformerFactory = TransformerFactory.newInstance();
        transformerFactory.setAttribute("indent-number", indent);
        transformerFactory.setAttribute(XMLConstant.ACCESS_EXTERNAL_ALL, "");
        Transformer transformer = transformerFactory.newTransformer();
        transformer.setOutputProperty(OutputKeys.INDENT, "yes");
        transformer.transform(xmlInput, xmlOutput);
        return xmlOutput.getWriter().toString();
    } catch (Exception e) {
        throw new RuntimeException(e); // simple exception handling
    }
}

public static String prettyFormat(String input) {
    return prettyFormat(input, 2);
}
```



Code Fixes on StackOverflow


4th revision of this post contains the fix for a vulnerability

✓ 4

Fix XXE vulnerability as mentioned in the comments

[Source](#) [Link](#)

edit approved Mar 19, 2021 at 22:14

 fanbondi

1k ● 5 ● 19 ● 37

Inline **Side-by-side** Side-by-side Markdown

[a simpler solution based on this answer:](#)

```
public static String prettyFormat(String input, int indent) {
    try {
        Source xmlInput = new StreamSource(new StringReader(input));
        StringWriter stringWriter = new StringWriter();
        StreamResult xmlOutput = new StreamResult(stringWriter);
        TransformerFactory transformerFactory = TransformerFactory.newInstance();
        transformerFactory.setAttribute("indent-number", indent);
        Transformer transformer = transformerFactory.newTransformer();
        transformer.setOutputProperty(OutputKeys.INDENT, "yes");
        transformer.transform(xmlInput, xmlOutput);
        return xmlOutput.getWriter().toString();
    } catch (Exception e) {
        throw new RuntimeException(e); // simple exception handling
    }
}

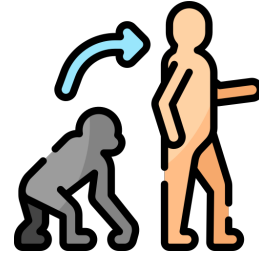
public static String prettyFormat(String input) {
    return prettyFormat(input, 2);
}
```

testcase:

[a simpler solution based on this answer:](#)

```
public static String prettyFormat(String input, int indent) {
    try {
        Source xmlInput = new StreamSource(new StringReader(input));
        StringWriter stringWriter = new StringWriter();
        StreamResult xmlOutput = new StreamResult(stringWriter);
        TransformerFactory transformerFactory = TransformerFactory.newInstance();
        transformerFactory.setAttribute("indent-number", indent);
        transformerFactory.setAttribute(XMLConstant.ACCESS_EXTERNAL_ALL, "");
        Transformer transformer = transformerFactory.newTransformer();
        transformer.setOutputProperty(OutputKeys.INDENT, "yes");
        transformer.transform(xmlInput, xmlOutput);
        return xmlOutput.getWriter().toString();
    } catch (Exception e) {
        throw new RuntimeException(e); // simple exception handling
    }
}

public static String prettyFormat(String input) {
    return prettyFormat(input, 2);
}
```



**Code on Stack Overflow evolves
through revisions**



Motivating Example (cont.)



Lucy, a software developer
in the year 2021

```
public static String prettyFormat(String input, int indent) {  
    try {  
        Source xmlInput = new StreamSource(new StringReader(input));  
        StringWriter stringWriter = new StringWriter();  
        StreamResult xmlOutput = new StreamResult(stringWriter);  
        Transformer transformerFactory = TransformerFactory.  
            transformerFactory.newTransformer(newTransformer(Out  
            transformer.setOutputProperty(newTransformer(Out  
            transformer.setOutputProperty(newTransformer(Out  
            transformer.transform(xmlInput, xmlOutput);  
        return xmlOutput.getWriter().toString();  
    } catch (Exception e) {  
        throw new RuntimeException(e); // simple exception handling, please review it  
    }  
}
```

Lucy is unaware that they missed a fix to a critical vulnerability in a reused snippet!



Two Forms of Software Dependencies



Two Forms of Software Dependencies

Depending on External Libraries

JUnit 5





Two Forms of Software Dependencies

Depending on External Libraries

JUnit 5



Tools to keep **managed**
dependencies up-to-date!



Dependabot



snyk

and more



Two Forms of Software Dependencies

Depending on External Libraries

JUnit 5

JWT
JSON WEB Token

MySQL

Tools to keep **managed**
dependencies up-to-date!



Dependabot



snyk

and more

Depending on Little Copying

 **stackoverflow**



Two Forms of Software Dependencies

Depending on External Libraries

JUnit 5



Tools to keep **managed**
dependencies up-to-date!



Dependabot



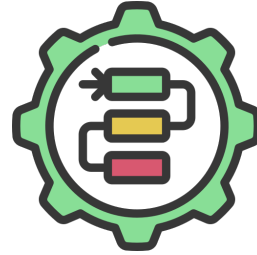
snyk

and more

Depending on Little Copying



No management!



Copying code snippets (from Stack Overflow) creates an **unmanaged dependency**



Are there more Lucy's in the wild?



Are there more Lucy's in the wild?

RQ1: Do Stack Overflow code artifacts reused in developer code bases evolve on Stack Overflow?



Are there more Lucy's in the wild?

RQ1: Do Stack Overflow code artifacts reused in developer code bases evolve on Stack Overflow?

RQ2: Can we find evidence that developers monitor Stack Overflow for code updates?



Are there more Lucy's in the wild?

RQ1: Do Stack Overflow code artifacts reused in developer code bases evolve on Stack Overflow?

RQ2: Can we find evidence that developers monitor Stack Overflow for code updates?

RQ3: Do developers miss security fixes on Stack Overflow for code also present in their code bases?



Methodology

StackOverflow





Methodology

StackOverflow



GitHub





Methodology

StackOverflow



3.5 Million Code Snippet Versions
(from 1.5 Million Posts)

GitHub



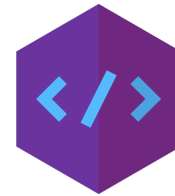
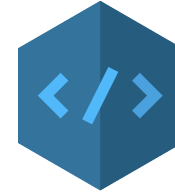


Methodology

StackOverflow



3.5 Million Code Snippet Versions
(from 1.5 Million Posts)



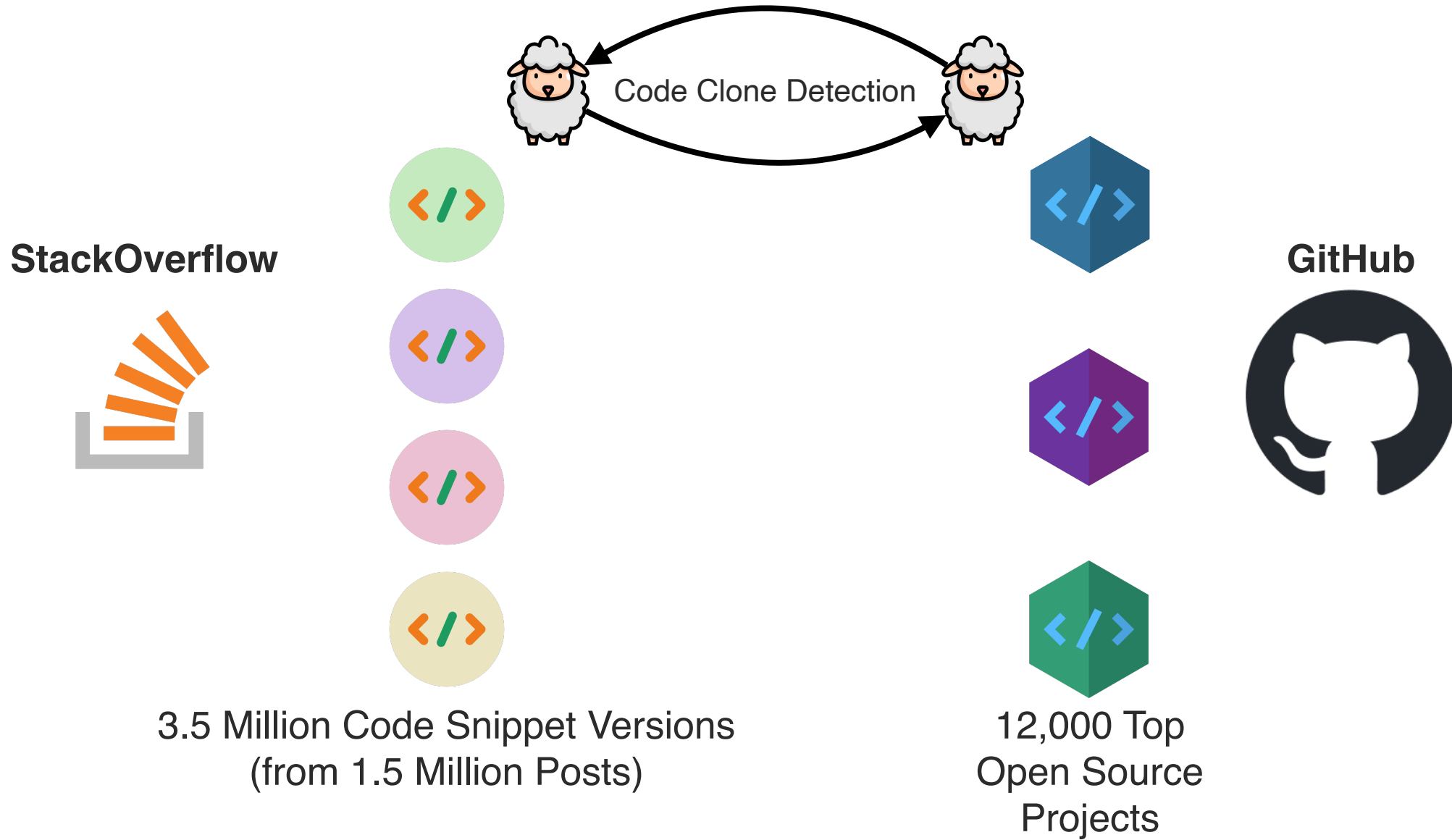
12,000 Top
Open Source
Projects

GitHub



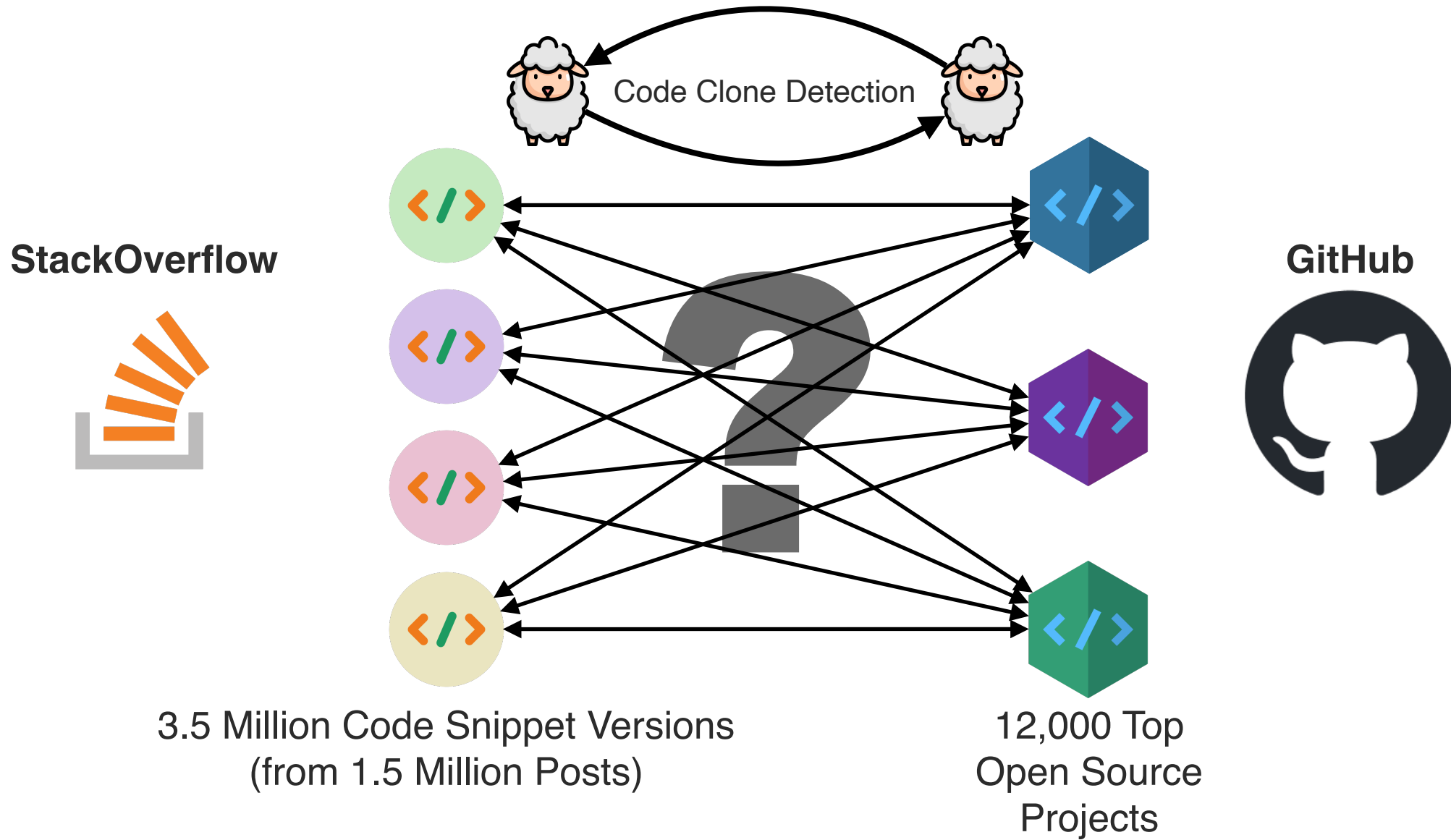


Methodology





Methodology





20% of all GitHub projects contained outdated code from StackOverflow



34 Top-ranked Projects contained **severe security vulnerabilities in outdated code**



Who is Lucy?



[lucene-solr](#) / [solr](#) / [solrj](#) / [src](#) / [java](#) / [org](#) / [apache](#) / [solr](#) / [common](#) / [cloud](#) / [SolrZkClient.java](#)

Code

Blame

909 lines (809 loc) · 32.5 KB

Older  Newer

13 years ago



SOLR-2358: merge in solrclou...



```
650 public static String prettyPrint(String input, int indent) {
651     try {
652         Source xmlInput = new StreamSource(new StringReader(input));
653         StringWriter stringWriter = new StringWriter();
654         StreamResult xmlOutput = new StreamResult(stringWriter);
655         TransformerFactory transformerFactory = TransformerFactory.newInstance();
656         transformerFactory.setAttribute("indent-number", indent);
657         Transformer transformer = transformerFactory.newTransformer();
658         transformer.setOutputProperty(OutputKeys.INDENT, "yes");
659         transformer.transform(xmlInput, xmlOutput);
660         return xmlOutput.getWriter().toString();
661     } catch (Exception e) {
662         throw new RuntimeException("Problem pretty printing XML", e);
663     }
664 }
```



Top projects with vulnerabilities

Project	Weakness	Language	Watch	Fork	Star	Contributors	Last Commit (year)
Odoo	Undefined Behaviour	JavaScript	1,494	16,121	24,881	1,385	2022
Wikimedia Commons Android app	CWE-404	Java	60	968	731	270	2022
The Fuck	Others	Python	848	3,189	70,614	176	2022
Apache NetBeans	CWE-404	Java	157	689	1,836	175	2022
Amazon S3cmd	Others	Python	103	850	3,962	175	2022
Open Event Frontend	CWE-1339	Python	22	1,694	2,188	151	2022
CARTO	CWE-690	JavaScript	207	672	2,582	134	2022
logback	CWE-404	Java	167	1,115	2,404	104	2022
Cider	CWE-754 (2)	JavaScript	24	174	348	65	2022
Apache Nutch	Undefined Behaviour	Java	240	1,199	2,356	46	2022
JPEXS Free Flash Decompiler	Others	Java	187	534	3,128	33	2022
Python Audio Analysis Library	CWE-754	Python	207	1,083	4,742	22	2022
WordOps	CWE-772	Python	58	167	866	21	2022
Eclipse N4JS	CWE-772	Java	11	25	26	18	2022
Weevely	Undefined Behaviour	Python	132	548	2,596	18	2022
Gmail Backup Software (Gmvault)	CWE-172	Python	80	272	3,429	16	2022
RomRaider	CWE-172	Java	80	272	3,429	15	2022
Kubebox	CWE-690	JavaScript	47	139	1,912	13	2022
* AndroidTreeView	CWE-475	JavaScript	84	614	2,884	5	2022
Chrome-Extensions	CWE-1339	JavaScript	80	483	894	5	2022
Apache Lucene-Solr	CWE-611	Java	315	2,723	4,357	234	2021
MoneyManagerEx for Android	CWE-404	Java	49	167	327	17	2021
Apache Fineract Android Client	CWE-404	Java	20	147	31	13	2021
Faster-RCNN in Tensorflow	Others	Python	88	1,149	2,461	7	2021
STATSD-C	CWE-120	C	6	13	75	6	2021
Faster RCNN with PyTorch	Others	Python	52	464	1,609	3	2021
Gnucash for Android	CWE-475	Java	101	525	1,143	46	2020
AppScale GTS	Undefined Behaviour	Python	159	293	2,419	39	2020
Eclipse Ceylon	Undefined Behaviour	Java	41	64	385	34	2020



Top projects with vulnerabilities

Project	Weakness	Language	Watch	Fork	Star	Contributors	Last Commit (year)
Odoo	Undefined Behaviour	JavaScript	1,494	16,121	24,881	1,385	2022
Wikimedia Commons Android app	CWE-404	Java	60	968	731	270	2022
The Fuck	Others	Python	848	3,189	70,614	176	2022
Apache NetBeans	CWE-404	Java	157	689	1,836	175	2022
Amazon S3cmd	Others	Python	103	850	3,962	175	2022
Open Event Frontend	CWE-1339	Python	22	1,694	2,188	151	2022
CARTO	CWE-690	JavaScript	207	672	2,582	134	2022
logback	CWE-404	Java	167	1,115	2,404	104	2022
Cider	CWE-754 (2)	JavaScript	24	174	348	65	2022
Apache Nutch	Undefined Behaviour	Java	240	1,199	2,356	46	2022
JPEXS Free Flash Decompiler	Others	Java	187	534	3,128	33	2022
Python Audio Analysis Library	CWE-754	Python	207	1,083	4,742	22	2022
WordOps	CWE-772	Python	58	167	866	21	2022
Eclipse N4JS	CWE-772	Java	11	16	16	18	2022
Weevely	Undefined Behaviour	Python	132	58	2,596	18	2022
Gmail Backup Software (Gmvault)	CWE-172	Python	80	272	3,429	16	2022
RomRaider	CWE-172	Java	80	272	3,429	15	2022
Kubebox	CWE-690	JavaScript	47	139	1,912	13	2022
* AndroidTreeView	CWE-475	JavaScript	84	614	2,884	5	2022
Chrome-Extensions	CWE-1339	JavaScript	80	483	894	5	2022
Apache Lucene-Solr	CWE-611	Java	315	2,723	4,357	234	2021
MoneyManagerEx for Android	CWE-404	Java	49	167	327	17	2021
Apache Fineract Android Client	CWE-404	Java	20	147	31	13	2021
Faster-RCNN in Tensorflow	Others	Python	88	1,149	2,461	7	2021
STATSD-C	CWE-120	C	6	13	75	6	2021
Faster RCNN with PyTorch	Others	Python	52	464	1,609	3	2021
Gnucash for Android	CWE-475	Java	101	525	1,143	46	2020
AppScale GTS	Undefined Behaviour	Python	159	293	2,419	39	2020
Eclipse Ceylon	Undefined Behaviour	Java	41	64	385	34	2020

Lucy is not the only one!



Future work



Future work

- Developer tools to monitor and track updates on “snippet dependencies” and their surrounding contexts



Future work

- Developer tools to monitor and track updates on “snippet dependencies” and their surrounding contexts
- Bridging the gap between various knowledge-sharing ecosystems to transfer insights and evolution for any reused code



Future work

- Developer tools to monitor and track updates on “snippet dependencies” and their surrounding contexts
- Bridging the gap between various knowledge-sharing ecosystems to transfer insights and evolution for any reused code
- Study the impact of StackOverflow code evolution on (prior) developer-centric studies



Future work

- Developer tools to monitor and track updates on “snippet dependencies” and their surrounding contexts
- Bridging the gap between various knowledge-sharing ecosystems to transfer insights and evolution for any reused code
- Study the impact of StackOverflow code evolution on (prior) developer-centric studies

Thank you!