



Research Note

RN/17/xx

**Awareness of Stack Overflow Users to
Outdated and License-violating Code:
The Online Survey**

August 21, 2017

Chaiyong Ragkhitwetsagul, Jens Krinke, Mathues Paixao, Giuseppe Bianco, Rocco Oliveto

Abstract

ABSTRACT GOES HERE

1 Introduction

Outdated third-party code and software licensing conflicts are ramifications by reusing source code by copying and pasting (i.e. code cloning). Xia et al (2014) show that a large number of open source systems reuse outdated third-party libraries from popular open source projects. Using the outdated code give detrimental effects to the software since they may introduce vulnerabilities. On the other hand, German et al (2009) found conflicts in software license between the clones among different systems that come from the same source, i.e. clone siblings.

The Internet encourages fast and easy way of sharing and copying code. Developers nowadays do not only copy code from local software projects, but also from online sources such as programming Q&A website and online code repositories. Stack Overflow is one of the most popular programming Q&A websites in the world. It has 7.6 million users, 14 million questions, 23 million answers¹, and more than 50 million developers visiting each month². Nevertheless, several code snippets on Stack Overflow are found to be problematic. Many code snippets provided as solutions are workarounds and occasionally contains defects or vulnerabilities. Acar et al (2016) performed a user study and found that Stack Overflow helps developers to solve Android programming problems quicker than other resources while, at the same time, offers less secure code than books or the official Android documentation. Only 17% of the Stack Overflow discussion threads the participant visited during the study contain secure code snippets. They also found that similar piece of API call copied from Stack Overflow by participants in their study existed in a random sample of 200,000 Android apps from Google Play. An et al (2017) showed that 1,279 cloned snippets between 399 Android apps and Stack Overflow potentially violate software licenses.

We are interested in studying the issues of outdated code and software license violation cause by code snippets on Stack Overflow. Asking and answering questions on Stack Overflow usually involve copying and pasting source code snippets, either in a question or an answer. However, the copied snippets are not regularly checked and occasionally not up-to-date with their originals. Some snippets are copied from an open source system with stricter software license than Stack Overflow's Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0). Before this study, we performed an preliminary study using code clone detection tools to locate cloned code snippets in accepted answers. We found that many snippets were copied from open source projects. Some cloned snippets are outdated and possibly violating the original software license. In either case, these snippets are potentially harmful for reuse. To accompany the findings and to gain insights into the cause of this problem, we resort to a qualitative study using an online survey of Stack Overflow developers who 1) answer programming questions with code snippets and 2) reuse code snippets from Stack Overflow answers.

In this paper, we use the term “answerers” to refer to Stack Overflow users who have high reputation, specifically the users with reputation from 963,731 (highest) to 6,999. The answerers gain their reputation from answering questions and receiving votes from other users. The reputation reflects trust they gain from other users and their activeness in answering questions. We ask the answerers the origins of source code snippets in their answers and assess their awareness of outdated and licensed code. We use the term “visitors” the refer to developers, who may or may not have a Stack Overflow account but visit Stack Overflow when facing programming problems. They copy example code snippets that are relevant to their problems and reused them in their software. We ask Stack Overflow visitors, who copied code snippets from Stack Overflow, about their experience of outdated or license-violating code.

2 Research Questions

In this study, we would like to answer the following research questions:

1. **RQ1 (Sources of Stack Overflow Answer Snippets):** *where are the code snippets in Stack Overflow answers from?*
2. **RQ2 (Awareness of Stack Overflow answerers to outdated code):** *are Stack Overflow answerers aware of outdated code in their answers?*

¹ Data as of 21 August 2017 from <https://stackexchange.com/sites>

² Data as of 21 August 2017 from: <https://stackoverflow.com>

Table 1: The Stack Overflow answerer taken the surveys

Target group	Reputation	Sent emails	Answers	Rate
Answerer Group 1	963,731–7,674	300	113	38%
Answerer Group 2	7,636–6,999	307	85	28%
Total	–	607	198	33%

3. **RQ3 (Awareness of Stack Overflow answerers to software licensing conflicts):** *are Stack Overflow answerers aware of software licensing conflicts caused by code snippets in their answers?*
4. **RQ4 (Awareness of Stack Overflow visitors to outdated code):** *are Stack Overflow visitors aware of outdated code on Stack Overflow?*
5. **RQ5 (Awareness of Stack Overflow visitors to software licensing conflicts):** *are Stack Overflow visitors aware of software licensing conflicts caused by reusing code snippets from Stack Overflow?*

3 Research Methodology

The study is performed using online surveys. We designed the surveys using Google Forms by following the principles of survey research by [Pfleeger and Kitchenham \(2001\)](#) and [Kitchenham and Pfleeger \(2002\)](#). We created two versions of the survey: **the answerer survey**, and, **the visitor survey**. The surveys are completely anonymous and the participants could decide to leave at any time. They do not collect any sensitive personal information from the participants and were approved for an ethical waiver by the designated ethics officer in the Computer Science Department at UCL. The complete versions of both surveys can be found in the Appendix A and B.

3.1 The Answerer Survey

The Stack Overflow answerer survey contains 11 questions: 7 Likert's scale questions, 3 yes/no questions, and one open-ended question for additional comments. The first two questions were mandatory while the other 9 questions were shown to the participants based on their previous answers. The survey collects information of the participant's software development experience, experience of answering Stack Overflow questions, sources of the Stack Overflow snippets, concerns of copying code snippets to Stack Overflow, problems from code snippets on Stack Overflow, fixing outdated code, and additional feedbacks.

We selected the participants for our answerer survey based on their Stack Overflow reputation. On Stack Overflow, a user's reputation reflects how much the community trust the user. A user earns reputation when he or she receives up votes for good questions and useful answers. Accepted answers receive more reputation score than questions, and regular answers³. Thus, Stack Overflow reputation is an indicator of user's skills and their involvement in asking and answering questions on the site.

The participants were invited to take the survey via email addresses publicly available on their public Stack Overflow and GitHub profiles. We selected the answerers based on all-time reputation ranking⁴ and separated them into two groups (see Table 1) so that we can compare the results or observe similar patterns between them. The first group had reputation from 963,731 (the highest) to 7,674, and the second group had reputation from 7,636 to 6,999 which we sent out 300 and 307 emails (excluding undelivered ones) respectively. The survey was open for participation for a month, from 25 July 2017 to 25 August 2017, before we collected and analysed the responses.

³ Stack Overflow Reputation: <https://stackoverflow.com/help/whats-reputation>

⁴ Stack Overflow Users (data as of 25 July 2017): <https://stackoverflow.com/users?tab=Reputation&filter=all>

Table 2: The Stack Overflow visitors taken the survey

Invitation channels	Answers
Chaiyong's Facebook post	47
Blognone.com	32
University of Molise	6
comp.lang.java.programmer	3
Software Engineering Facebook page	1
Total	89

3.2 The Visitor Survey

The Stack Overflow visitor survey is consisted of 16 questions: 9 Likert's scale questions, 3 yes/no questions, 2 multiple-choice questions, and 2 open-ended questions. The first four questions were mandatory while the other 12 questions were shown to the participants based on their previous answers. The survey collects information of the participant's software development experience, importance of Stack Overflow, reasons for reusing Stack Overflow snippets, problems from Stack Overflow snippets, licensing of code on Stack Overflow, and additional feedbacks.

The participants for this survey were invited to take the survey via five channels. The first channel is via the first author's Facebook post where the first author invited friends who are software developers and have experience of copying code snippets from Stack Overflow to take the survey. The second channel is a popular Thai technology news and media community called blognone.com which attracts a high number of software developers. The first author posted an invitation to the visitor survey in a discussion forum. The third channel collects answers from University of Molise which is the third author's workplace. The last two channels are from a comp.lang.java.programmer Google groups and Software Engineering Facebook page. The survey was open for participation for a month, from 25 July 2017 to 25 August 2017, before we collected and analysed the responses. The number of participants taken the survey are shown in Table 2.

4 Results and Discussions

4.1 Answerer Survey

We received 113 answers (38% response rate) from the first group and 85 answers (28% response rate) from the second group of Stack Overflow answerers. The response rate from both groups were high considered other online survey in software engineering that received approximately 14–20% **FIXME: double-check the citation again for the magic numbers.** (Punter et al, 2003).

How long have you been working on developing software?

The majority of users in both groups are experienced developers with more than 10 years of experience or between 5 to 10 years as depicted in Figure 1. They are active users and regularly answer questions on Stack Overflow. Eighty one and sixty percent of users from Group 1 and Group 2 answer questions at least once a week.

How frequently do or did you answer questions on Stack Overflow?

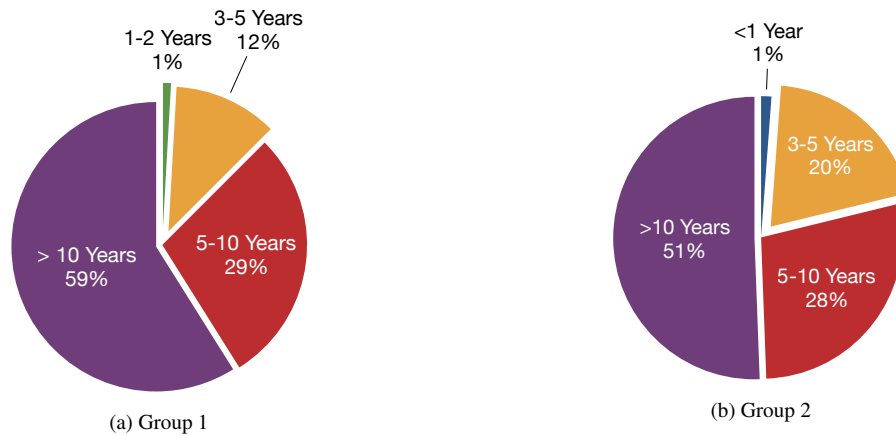


Fig. 1: Experience of the Stack Overflow answerers

How frequently do or did you include code snippets in your answers on Stack Overflow?

4.2 RQ1: Where are the code snippets in Stack Overflow answers from?

To answer this research question, we ask the participants this question:

Where did the code snippets in your answers come from?

From the answers, a large amount of 91 (81%) and 58 (68%) answerers in both group include code snippets in more than sixty percent of their answers. Figure 2 explains where are these code snippets from. Stack Overflow answerers mostly write new code from scratch or modify from the code snippets in question for each answer, while fewer numbers are from other sources including their personal projects, their company projects, open source projects. We found 27 answerers in both Group 1 and Group 2 occasionally to very frequently copy code from open source projects.

4.3 RQ2: Are Stack Overflow answerers aware of outdated code in their answers?

Outdated code occurs when code snippets in your answers are no longer up-to-date with the latest version of the software you copied the code from. Have you ever been notified of outdated code in your Stack Overflow answers?

We found that 63.1% of Stack Overflow answerers in Group 1 have been notified of outdated code in their answers (see Figure 4). The ratio drops to 47.6% in Group 2. Thus, at least roughly half of the top answerers on Stack Overflow are aware of their outdated code in some of their answers.

How frequently were you notified of outdated or deprecated code in your Stack Overflow answers?

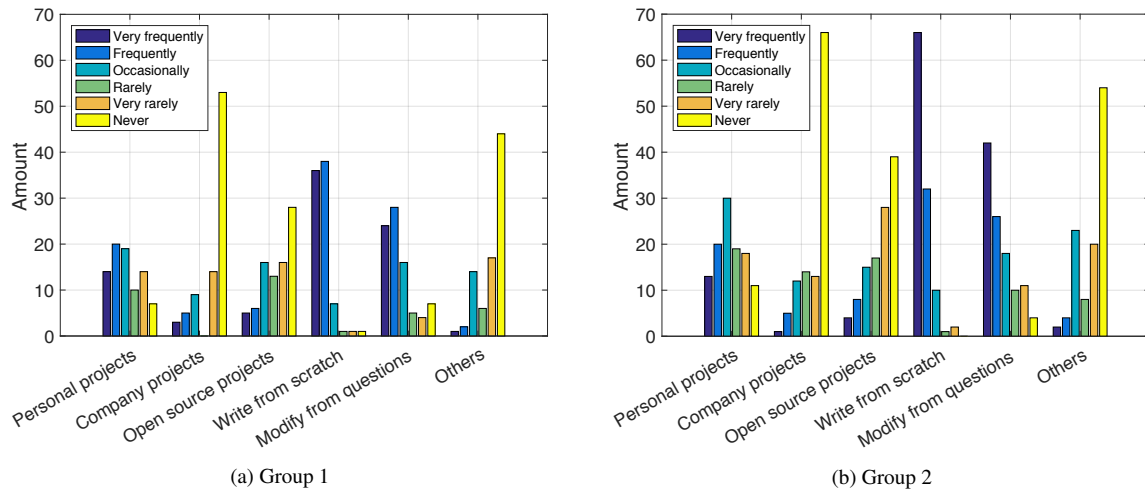


Fig. 2: Sources of code snippets in Stack Overflow answers

How frequently did you fix your outdated code on Stack Overflow?

We then asked the participants to have been notified of their outdated code (80 and 48 participants from Group 1 and 2 respectively) a follow-up questions “how frequently did you fix your outdated code on Stack Overflow?”. The answers, depicted in Figure 5, show that more than half of them frequently fix the outdated code snippets. However, there are 18.8% of the answerers in both groups who rarely or never fix their code.

4.4 RQ2: Are Stack Overflow answerers aware of software licensing conflicts caused by code snippets in their answers?

Were you aware, at the time of copying the code, that Stack Overflow apply Creative Commons Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0) to content in the posts, including code snippets?

More than half of the answerers in both groups (63.1% and 61.9% respectively) are aware that Stack Overflow apply Creative Commons Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0) to content in the posts, including code snippets, while the rest of 36.9% and 38.1% are not. Nevertheless, we found that almost all of them do not include license statement in their code snippets (as depicted in Figure 3) due to several reasons. First, some answerers chose to post only their own code or code that was adapted from the question, hence they are automatically subjected to CC BY-SA 3.0. Second, they copied code from company or open source projects that they knew were permitted to be publicly distributed. Third, some answerers believe that code snippets in their answers are too small to claim any intellectual property on them and fall under fair use (Fair Use, 2017).

Do you usually include a software license in your code snippets on Stack Overflow?

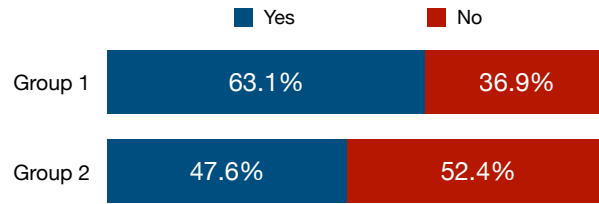


Fig. 4: Percentage of answerers who are notified of outdated code in their Stack Overflow answers.

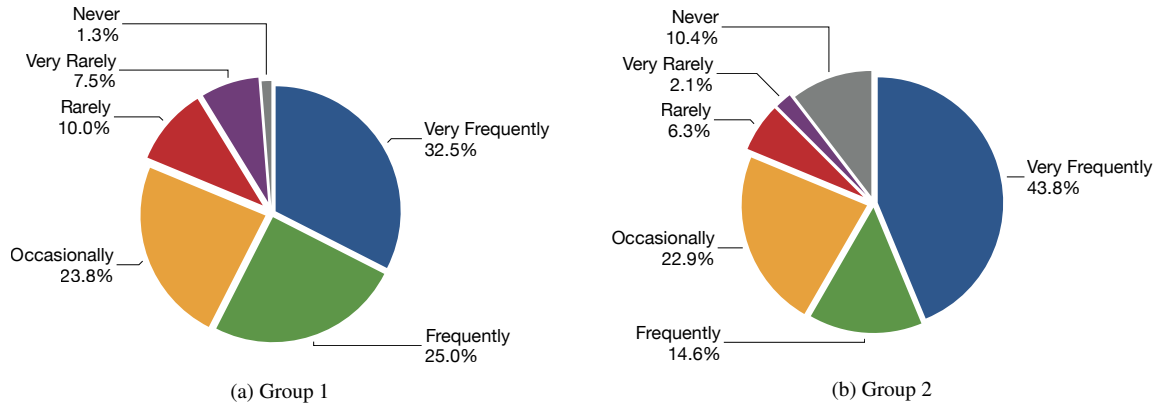


Fig. 5: Frequency of the answerers fixing their outdated code.

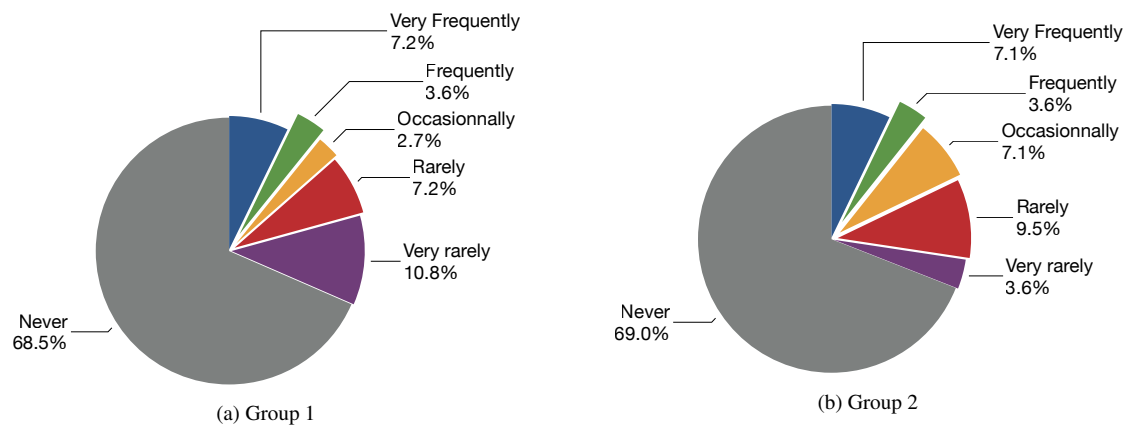


Fig. 6: Frequency of the answerers checking license of their code snippets against Stack Overflow's CC BY-SA 3.0.

While nobody explicitly includes a software license in their snippets, many users include a statement on their profile page that states all their answers are under a certain license. For example, *All code posted by me on Stack Overflow should be considered public domain without copyright. For countries where public domain is not applicable, I hereby grant everyone the right to modify, use and redistribute any code posted by me on Stack Overflow for any purpose. It is provided "as-is" without warranty of any kind.* Many users either declare their snippets to be public domain, or they grant additional licenses, e.g. Apache 2.0 or MIT/Expat.

How frequently did you check the software license of the code snippets you copy to Stack Overflow if they conflict with Stack Overflow's CC BY-SA 3.0 license?

We asked the answerers a follow-up question of how frequently they checked for a conflict between software license of the code snippets they copied to their answers and Stack Overflow's CC BY-SA 3.0. As shown in Figure 6, approximately 69% of answerers from both groups did not perform the checking. Nonetheless, they are about 10% of the answerers who frequently checks for licensing conflicts when they copy code snippets to Stack Overflow.

Do you have any other concerns regarding answering Stack Overflow with code snippets?

We also invited the participants to give any comments regarding answering Stack Overflow with code snippets. These are some interesting comments we received.

Comment 1: "The real issue is less about the amount the code snippets on SO than it is about the staggeringly high number of software "professionals" that mindlessly use them without understanding what they're copying, and the only slightly less high number of would-be professionals that post snippets with built-in security issues. A related topic is beginners who post (at times dangerously) misleading tutorials online on topics they actually know very little about. Think php/mysql tutorials written 10+ years after mysql_* functions were obsolete, or the recent regex tutorial that got posted the other day on HackerNew (<https://news.ycombinator.com/item?id=14846506>). They're also full of toxic code snippets."

Comment 2: "When I copy code it's usually short enough to be considered "fair use" but I am not a lawyer or copyright expert so some guidance from SO would be helpful. I'd also like the ability to flag/review questions that violate these guidelines."

Comment 3: "My only concern, albeit minor, is that I know people blindly copy my code without even understanding what the code does."

Comment 4: "The main problem for me/us is outdated code, esp. as old answers have high google rank so that is what people see first, then try and fail. That's why we're moving more and more of those examples to knowledge base and docs and rather link to those."

Comment 5: "Lot of the answers are from hobbyist so the quality is poor. Usually they are hacks or workarounds (even MY best answer on SO is a workaround)."

4.4.1 Visitors' Experience

5 Related Work

5.1 Stack Overflow

Stack Overflow is a gold mine for software engineering research and has been put to use in several previous studies. In terms of developer-assisting tools, Seahawk is an Eclipse plug-in that searches and recommends relevant code snippets from Stack Overflow (Ponzanelli et al, 2013). A follow up work, Prompter, by Ponzanelli et al. (Ponzanelli et al, 2014) achieves the same goal but with improved algorithms. The code snippets on Stack Overflow are mostly examples or solutions to programming problems. Hence, several code search systems use whole or partial data from Stack Overflow as their code search databases (Keivanloo et al, 2014; Park et al, 2014; Stolee et al, 2014; Subramanian and Holmes, 2013; Diamantopoulos and Symeonidis, 2015). Furthermore, Treude et al. Treude and Robillard (2016) use machine learning techniques to extract insight sentences from Stack Overflow and use them to improve API documentation.

Another research area is knowledge extraction from Stack Overflow. Nasehi et al (2012) studied what makes a good code example by analysing answers from Stack Overflow. Similarly, Yang et al (2016) report the number of reusable code

snippets on Stack Overflow across various programming languages. Wang et al (2013) use Latent Dirichlet Allocation (LDA) topic modelling to analyse questions and answers from Stack Overflow so that they can automatically categorise new questions. There are also studies trying to understand developers' behaviours on Stack Overflow, e.g. A study by Movshovitz-Attias et al (2013); Rosen and Shihab (2016); Choetkiertikul et al (2015); Bosu et al (2013). Yang et al (2017) analysed 909k non-fork Python projects on Github and 1.9 million python code snippets on Stack Overflow and found thousands of code blocks that are copied from Stack Overflow to GitHub.

5.2 Software Licensing

Software licensing is crucial for open source and industrial software development. Di Penta et al. Di Penta et al (2010) studied the evolution of software licensing in open source software and found that licensing statements change over time. German et al (2009) found that licensing conflicts occur between the clone siblings, i.e. clones among different systems that come from the same source. Later, German et al (2010) created an automated tool for software license identification, Ninka. An et al (2017) detected clones between 399 Android apps and Stack Overflow Java code snippets. They found that that 1,279 cloned snippets potentially violate software licenses. Our study gives insights into the software licensing of Stack Overflow code snippets. We study the licensing awareness of developers when they copy and paste code to/from Stack Overflow.

5.3 Reusing of outdated third-party source code

Outdated code occurs in software development. Xia et al. Xia et al (2014) discover that there are code reused from popular open source projects in a large number of open source systems. More importantly, some of the reused code are outdated which affect quality and security of the software. Our study focuses on a similar problem but on the context of Stack Overflow.

6 Conclusions

References

- Acar Y, Backes M, Fahl S, Kim D, Mazurek ML, Stransky C (2016) You get where you're looking for: The impact of information sources on code security. In: SP '16, pp 289–305
- An L, Mlouki O, Khomh F, Antoniol G (2017) Stack Overflow: A code laundering platform? In: SANER '17
- Bosu A, Corley CS, Heaton D, Chatterji D, Carver JC, Kraft NA (2013) Building reputation in StackOverflow: An empirical investigation. In: MSR '13, pp 89–92
- Choetkiertikul M, Avery D, Dam HK, Tran T, Ghose A (2015) Who will answer my question on Stack Overflow? In: ASWEC '15, pp 155–164
- Di Penta M, German DM, Guéhéneuc YG, Antoniol G (2010) An exploratory study of the evolution of software licensing. In: ICSE '10, vol 1, p 145
- Diamantopoulos T, Symeonidis AL (2015) Employing source code information to improve question-answering in Stack Overflow. In: MSR '15, pp 454–457
- Fair Use (2017) Intellectual property for cs students: Copyrights - fair use. <https://www.cs.duke.edu/courses/cps182s/fall02/cscopyright/Copyrights/Copyright-Fairuse.htm>
- German DM, Di Penta M, Gueheneuc YG, Antoniol G (2009) Code siblings: Technical and legal implications of copying code between applications. In: MSR '09, pp 81–90
- German DM, Manabe Y, Inoue K (2010) A sentence-matching method for automatic license identification of source code files. In: ASE '10, p 437

- Keivanloo I, Rilling J, Zou Y (2014) Spotting working code examples. In: ICSE '14, pp 664–675
- Kitchenham BA, Pfleeger SL (2002) Principles of Survey Research Part 2 : Designing a Survey Sample size Experimental designs. *Software Engineering Notes* 27(1):18–20
- Movshovitz-Attias D, Movshovitz-Attias Y, Steenkiste P, Faloutsos C (2013) Analysis of the reputation system and user contributions on a question answering website: StackOverflow. In: ASONAM '13, pp 886–893
- Nasehi SM, Sillito J, Maurer F, Burns C (2012) What makes a good code example?: A study of programming Q&A in StackOverflow. In: ICSM '12, pp 25–34
- Park Jw, Lee MW, Roh JW, Hwang Sw, Kim S (2014) Surfacing code in the dark: an instant clone search approach. *Knowledge and Information Systems* 41(3):727–759
- Pfleeger SL, Kitchenham BA (2001) Principles of survey research part 1: Turning lemons into lemonade. *Software Engineering Notes* 26(6):16
- Ponzanelli L, Bacchelli A, Lanza M (2013) Seahawk: Stack Overflow in the IDE. In: ICSE '13, pp 1295–1298
- Ponzanelli L, Bavota G, Di Penta M, Oliveto R, Lanza M (2014) Mining StackOverflow to turn the IDE into a self-confident programming prompter. In: MSR '14, pp 102–111
- Punter T, Ciolkowski M, Freimut B, John I (2003) Conducting on-line surveys in software engineering. In: Proceedings of the International Symposium on Empirical Software Engineering (ISESE'03), pp 80–88
- Rosen C, Shihab E (2016) What are mobile developers asking about? a large scale study using Stack Overflow. *Empirical Software Engineering* 21(3):1192–1223
- Stolee KT, Elbaum S, Dobos D (2014) Solving the search for source code. *Transactions on Software Engineering and Methodology* 23(3):1–45
- Subramanian S, Holmes R (2013) Making sense of online code snippets. In: MSR '13, pp 85–88
- Treude C, Robillard MP (2016) Augmenting API documentation with insights from Stack Overflow. In: ICSE '16, pp 392–403
- Wang S, Lo D, Jiang L (2013) An empirical study on developer interactions in StackOverflow. In: SAC '13, pp 1019–1024
- Xia P, Matsushita M, Yoshida N, Inoue K (2014) Studying reuse of out-dated third-party code. *Information and Media Technologies* 9(2):155–161
- Yang D, Hussain A, Lopes CV (2016) From query to usable code: An analysis of Stack Overflow code snippets. In: MSR '16, pp 391–402
- Yang D, Martins P, Saini V, Lopes C (2017) Stack Overflow in Github: Any snippets there? In: Proceedings of the International Conference on Mining Software Repositories (MSR '17)

7 Appendix

A Stack Overflow Answerer Survey

A survey of developers' experiences on answering Stack Overflow questions with code examples

Dear developer,

Since you are one of the top answerers on Stack Overflow, we hope you can help us with a not-for-profit study.

We are researchers in the Software Systems Engineering Group at University College London, UK. We are studying problems caused by outdated and license-violating code snippets on Stack Overflow. We have designed a survey to understand these problems and would be grateful if you would complete it.

The survey is completely anonymous and has 11 questions and should only take about 3-5 minutes to complete. We hope that you will complete the entire form but if you do not wish to continue, you can just quit the session and your input will be discarded.

The survey results will be used only for academic research purposes and we plan to release the results to Stack Overflow and in the form of academic papers and presentations.

This research project has been approved by the designated ethics officer in the Computer Science Department at UCL.

If you have any question, please feel free to contact me.

Chaiyong Ragkhitwetsagul
chaiyong.ragkhitwetsagul.14@ucl.ac.uk
<http://www.cs.ucl.ac.uk/staff/C.Ragkhitwetsagul>

* Required

1. **How long have you been working on developing software? ***

Mark only one oval.

- ☐ Less than a year
- ☐ 1 - 2 years
- ☐ 3 - 5 years
- ☐ 5 - 10 years
- ☐ More than 10 years

2. **How frequently do or did you answer questions on Stack Overflow? ***

Mark only one oval.

- ☐ Very Frequently (every day)
- ☐ Frequently (roughly 3-6 times a week)
- ☐ Occasionally (roughly once or twice a week)
- ☐ Rarely (roughly once or twice a month)
- ☐ Very Rarely (roughly once or twice a year)
- ☐ Never *Skip to question 11.*

Experience of Answering Stack Overflow Questions

This section aims to gain understanding on how developers choose code snippets to put as a solution on Stack Overflow.

3. **How frequently do or did you include code snippets in your answers on Stack Overflow? ***

Mark only one oval.

- ☐ Very Frequently (81--100% of the time)
- ☐ Frequently (61--80% of the time)
- ☐ Occasionally (41--60% of the time)
- ☐ Rarely (21--40% of the time)
- ☐ Very Rarely (1--20% of the time)
- ☐ Never (0% of the time) *Skip to question 11.*

Sources of the Stack Overflow Snippets

This section will gather information about the origins of the example snippets in Stack Overflow answers.

4. **Where did the code snippets in your answers come from? ***

Mark only one oval per row.

	Very frequently	Frequently	Occasionally	Rarely	Very rarely	Never
I copied them from my own personal projects.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I copied them from my company's projects.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I copied them from open source projects.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I wrote the new code from scratch.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I copied the code from the question and modified it for the answer.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Others	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Concerns of Copying Code Snippets to Stack Overflow

This section aims to study concerns of the answerers regarding software licensing of their code snippets in Stack Overflow answers.

5. **Were you aware, at the time of copying the code, that Stack Overflow apply Creative Commons Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0) to content in the posts, including code snippets? ***

Mark only one oval.

- ☐ Yes
- ☐ No

6. **Do you usually include a software license in your code snippets on Stack Overflow? ***

Mark only one oval.

- ☐ Yes, as a comment in the code.
- ☐ Yes, in a text surrounding the source code.
- ☐ Yes, both in code comments and text.
- ☐ No.

7. **How frequently did you check the software license of the code snippets you copy to Stack Overflow if they conflict with Stack Overflow's CC BY-SA 3.0 license? ***

Mark only one oval.

- ☐ Very Frequently (81--100% of the time)
- ☐ Frequently (61--80% of the time)
- ☐ Occasionally (41--60% of the time)
- ☐ Rarely (21--40% of the time)
- ☐ Very rarely (1--20% of the time)
- ☐ Never (0% of the time)

Problems from Code Snippets on Stack Overflow

This section will gather information of code snippets that were outdated after being copied to Stack Overflow.

8. **Outdated code occurs when code snippets in your answers are no longer up-to-date with the latest version of the software you copied the code from. Have you ever been notified of outdated code in your Stack Overflow answers? ***

Mark only one oval.

- ☐ Yes
- ☐ No

Frequency of Notifications

9. **How frequently were you notified of outdated or deprecated code in your Stack Overflow answers? ***

Mark only one oval.

- ☐ Very frequently (81--100% of my answers)
- ☐ Frequently (61--80% of my answers)
- ☐ Occasionally (41--60% of my answers)
- ☐ Rarely (21--40% of my answers)
- ☐ Very rarely (1--20% of my answers)
- ☐ Never (0% of my answers) *Skip to question 11.*

Fixing Outdated Code

10. How frequently did you fix your outdated code on Stack Overflow? **Mark only one oval.*

- ☐ Very frequently (81--100% of the cases)
- ☐ Frequently (61--80% of the cases)
- ☐ Occasionally (41--60% of the cases)
- ☐ Rarely (21--40% of the cases)
- ☐ Very rarely (1--20% of the cases)
- ☐ Never (0% of the cases)

Additional feedbacks**11. Do you have any other concerns regarding answering Stack Overflow with code snippets?**

Powered by
 Google Forms

B Stack Overflow Visitor Survey**A survey of developers' experiences on reusing code snippets from Stack Overflow**

Dear developer,

We are researchers in the Software Systems Engineering Group at University College London, UK. We are studying problems caused by outdated and license-violating code snippets on Stack Overflow. We have designed a survey to understand these problems and would be grateful if you would complete it.

The survey is completely anonymous and has 15 questions and should only take about 5-7 minutes to complete. We hope that you will complete the entire form but if you do not wish to continue, you can just quit the session and your input will be discarded.

The survey results will be used only for academic research purposes and we plan to release the results to Stack Overflow and in the form of academic papers and presentations.

This research project has been approved by the designated ethics officer in the Computer Science Department at UCL.

If you have any question, please feel free to contact me.

Chaiyong Ragkhitwetsagul
chaiyong.ragkhitwetsagul.14@ucl.ac.uk
<http://www.cs.ucl.ac.uk/staff/C.Ragkhitwetsagul>

* Required

1. How long have you been working on developing software?

Mark only one oval.

- ☐ Less than a year
- ☐ 1-2 years
- ☐ 3-5 years
- ☐ 5-10 years
- ☐ More than 10 years

Importance of Stack Overflow to Developers

This section aims to measure importance of Stack Overflow to developers when they solve programming tasks.

2. When you had a problem with your programming tasks, please rank in which order did you search for help (1 for the first, and 5 for the last, without a tie)? *

Mark only one oval per row.

	1st	2nd	3rd	4th	5th
Books	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Official documentations	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Stack Overflow	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Online repositories (e.g. GitHub)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Others	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

3. **How frequently do or did you copy source code snippets from Stack Overflow? ***

Mark only one oval.

- ☐ Very frequently (everyday)
- ☐ Frequently (roughly 3-6 times a week)
- ☐ Occasionally (roughly once or twice per week)
- ☐ Rarely (roughly once or twice a month)
- ☐ Very rarely (roughly once or twice a year)
- ☐ Never *Stop filling out this form.*

Reasons for Reusing Stack Overflow's Snippets

This section aims to understand why developers reuse code from Stack Overflow and their problems.

4. **Why did you copy and reuse code snippets from Stack Overflow?**

Mark only one oval per row.

	Strongly agree	Agree	Undecided	Disagree	Strongly Disagree
They are easy to find by searching the web.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
They solve problems similar to my problems with minimal changes.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The context of questions and answers helped me understand the code snippets better.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The voting mechanism and accepted answers helped filtering good code from bad code.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

5. **Have you ever found any problems from reusing Stack Overflow code snippets? ***

Mark only one oval.

- ☐ Yes
- ☐ No *Skip to question 10.*

Problems from Stack Overflow Code Snippets

This section aims to understand the problems from Stack Overflow snippets and whether the developers notify the answerers of the problems.

6. **How frequently did you find problems from reusing Stack Overflow code snippets? ***

Mark only one oval.

- ☐ Very frequently (81--100% of the reused snippets)
- ☐ Frequently (61--80% of the reused snippets)
- ☐ Occasionally (41--60% of the reused snippets)
- ☐ Rarely (21--40% of the reused snippets)
- ☐ Very rarely (1--20% of the reused snippets)
- ☐ Never

7. What were the problems? **Check all that apply.*

- ☐ Incorrect solution (the code claims to solve the problem in the question while it does not).
- ☐ Outdated solution (the code may work with the some older versions of the library or API, but not the one you are using).
- ☐ Mismatched solution (the code solves the problem in the question but it is not exactly the right solution for your problem).
- ☐ Other: _____

8. How frequently did you report the problems back to the Stack Overflow discussion threads? **Mark only one oval.*

- ☐ Very frequently (81--100% of the time)
- ☐ Frequently (61--80% of the time)
- ☐ Occasionally (41--60% of the time)
- ☐ Rarely (21--40% of the time)
- ☐ Very rarely (1--20% of the time)
- ☐ Never (0% of the time) *Skip to question 10.*

Reporting problems of Code on Stack Overflow**9. How did you report the problems?***Check all that apply.*

- ☐ I down-voted the answer containing the problematic code snippet
- ☐ I wrote a comment saying that the code has problems.
- ☐ I contacted the answerers regarding the problems directly.
- ☐ I never report problems.
- ☐ Other: _____

Licensing of Code on Stack Overflow

This section will study awareness of developers regarding licensing of code snippets on Stack Overflow.

10. Were you aware, at the time of copying the code, that Stack Overflow apply Creative Commons Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0) to content in the posts, including code snippets? **Mark only one oval.*

- ☐ Yes
- ☐ No

15. Could you please briefly explain what are the legal problems that you faced by copying code snippets from Stack Overflow?

Other Comments

16. Did you have any other problems from using code snippets from Stack Overflow?

Powered by
 Google Forms