



An empirical study of question discussions on Stack Overflow

Wenhan Zhu¹ · Haoxiang Zhang² · Ahmed E. Hassan³ · Michael W. Godfrey¹

Accepted: 20 May 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Stack Overflow provides a means for developers to exchange knowledge. While much previous research on Stack Overflow has focused on questions and answers (Q&A), recent work has shown that *discussions* in comments also contain rich information. On Stack Overflow, discussions through comments and chat rooms can be tied to questions or answers. In this paper, we conduct an empirical study that focuses on the nature of question discussions. We observe that: (1) Question discussions occur at all phases of the Q&A process, with most beginning before the first answer is received. (2) Both askers and answerers actively participate in question discussions; the likelihood of their participation increases as the number of comments increases. (3) There is a strong correlation between the number of question comments and the question answering time (i.e., more discussed questions receive answers more slowly). Our findings suggest that question discussions contain a rich trove of data that is integral to the Q&A processes on Stack Overflow. We further suggest how future research can leverage the information in question discussions, along with the commonly studied Q&A information.

Keywords Empirical software engineering · Stack Overflow · Q&A website · Communication channel · Commenting · Real-time chatting · Crowdsourced knowledge sharing

Communicated by: Nicole Novielli

✉ Haoxiang Zhang
haoxiang.zhang@acm.org

Wenhan Zhu
w65zhu@uwaterloo.ca

Ahmed E. Hassan
ahmed@cs.queensu.ca

Michael W. Godfrey
migod@uwaterloo.ca

¹ Software Analytics Group (SWAG), University of Waterloo, Waterloo, ON, Canada

² Centre for Software Excellence at Huawei Canada, Kingston, Canada

³ Software Analysis and Intelligence Lab (SAIL), Queen's University, Kingston, ON, Canada

1 Introduction

Stack Overflow is a technical question answering (Q&A) website widely used by developers to exchange programming-related knowledge through asking, discussing, and answering questions. The Q&A process on Stack Overflow creates a crowdsourced knowledge base that provides a means for developers across the globe to collectively build and improve their knowledge on programming and its related technologies. Stack Overflow has become one of the largest public knowledge bases for developers with more than 21.9 million questions as of December 2021 (Stack Exchange Data Dump 2021). A survey shows that retrieving information from Stack Overflow is an essential daily activity for many software developers (Storey et al. 2017).

On Stack Overflow, users can ask, answer, and discuss questions, and each question can receive multiple proposed answers. The user who asked the question (i.e., the “asker”) can decide to mark one answer as *accepted*, indicating that it resolves their question authoritatively. While ultimately Q&A is the most important activity on Stack Overflow, users can also post comments and/or start chat rooms that are tied to a specific *post* (i.e., question or answer). In this paper, we refer to comments and chat rooms messages on Stack Overflow as *discussions*; each discussion is associated with a single question (a *question discussion*) or proposed answer (an *answer discussion*).

Researchers have extensively studied the questions and answers on Stack Overflow. These studies ranged from finding out common question types (Allamanis and Sutton 2013) to predicting the best answers (Tian et al. 2013). The Q&A processes on Stack Overflow are commonly viewed as two independent events by the studies. The first event is asking the question; this occurs when a user posts a question on the platform. The second event is answering the question; this normally occurs when another user posts an answer to a question. However, commenting as a communication channel allows for user interactions beyond simple asking and answering. A recent study has shown that comments can be helpful in the Q&A process by providing support knowledge, such as code examples, references, and alternative suggestions (Sengupta and Haythornthwaite 2020), whereas previous research has focused primarily on answer comments. Some studies leverage answer comments to study the quality of answers on Stack Overflow. For example, Zhang et al. (2021a) leveraged comments highlighting obsolete information regarding their associated answers. As a Q&A platform, most content on the platform is consumed by viewers long after the question is answered. If misleading information exists on the platform, it can convey false information within the community. Another study (Chen et al. 2019) used comments as a sign of whether the community is aware of the security vulnerabilities contained in the answer. Meanwhile, some studies have also focused on the presentation of knowledge on Stack Overflow. These studies also approach the issue from the answer perspective. One study (Zhang et al. 2019) highlights that while users are reading answers on Stack Overflow, they should not ignore the information contained in their associated comments. In their next study (Zhang et al. 2021b), they showed that the current mechanisms on Stack Overflow to display comments is not ideal and can hurt the users when they are reading answers.

In our study, we focus on question comments. More specifically, we theorize that the commenting activities forms a discussion and our focus is to understand how the discussions affects the Q&A process on Stack Overflow. Unlike previous studies that mostly focus on answer comments which occur after a question has been answered, our study focuses on question comments which can occur before the question is answered.

To help understand why it is important to study how question discussions integrate with the Q&A process, we now consider a motivating example. Figure 1 shows a question titled

The screenshot displays a Stack Overflow thread. At the top, a question is posted: "Unable to set the NumberFormat property of the Range class". Below the question, there are three comments. The first comment, marked with a green checkmark, provides a solution: "The strange thing is that I haven't changed anything. I was considering including more cells with number formatting, but I never implemented it :s". This comment is followed by a link to a Microsoft support page. The second comment, marked with a green checkmark, summarizes the chat room discussion: "The problem as discovered in Chat was the workbook had more than 64,000 formats because of which the user was getting the 'Too many different cell formats' error message in Excel". The third comment, marked with a green checkmark, provides the final answer: "The problem as discovered in Chat was the workbook had more than 64,000 formats because of which the user was getting the 'Too many different cell formats' error message in Excel". The thread is labeled with 'A', 'B', 'C', and 'D' in red boxes, corresponding to the steps described in the caption.

Fig. 1 An example of the Q&A process involving discussions: (A) a user (the “asker”) asked a question; (B) another user (the “answerer”) started discussing with the asker in the comment thread; (C) the question was further clarified then resolved in the chat room; (D) the content of the comments and chat messages that led to the resolution of the question were summarized as an answer, which was marked as the accepted answer by the asker

“Unable to set the *NumberFormat* property of the *Range* class.”¹ Four minutes after the question was asked, another user posted a comment—attached to the question—asking for clarification on the problematic code snippet. A chat room was then created for the asker and the user to continue the discussion in real-time. A consensus was reached in the chat, and the results were summarized and posted as a proposed answer by the user, which the asker designated as *accepted*. This example highlights how the process of asking and answering questions is enabled by the discussion mechanisms of commenting and chatting, allowing a resolution to be reached quickly. That is, the question discussion can serve as a simple and effective socio-technical means to achieving closure on the question.

In this work, we use the Stack Overflow data dump from December 2021 (Stack Exchange Data Dump 2021) as our dataset; this dataset contains 43.6 million comments and 1.5 million chat messages. We use this data to explore the nature of question discussions and how they integrate with the crowdsourced Q&A process on Stack Overflow. To make our study easy to follow, we use the following notations to refer to different groups of questions observed within the dataset:

Symbol	Meaning	# in dataset
Q_{disc}	Questions with comments	13.0 M
Q_{chat}	Questions with chat rooms (and comments)	27,146
Q_{nd}	Questions with no discussions	9.0 M
Q_a	Questions with answers	18.8 M
$Q_{d/a}$	Questions with both discussions and answers	10.5 M
$Q_{d/aa}$	Questions with both discussions and accepted answers	6.1 M
$Q_{hd/a}$	Questions with both discussions with “hidden comments” ² and answers	1.6 M

¹<https://stackoverflow.com/questions/10801537/>

Specifically, we investigate and answer three research questions (RQs):

RQ1: How prevalent are question discussions on Stack Overflow?

We found that question discussions occur in 59.2% of the questions on Stack Overflow. More specifically, 13.0 million questions have comments (i.e., Q_{disc}) with a median of 3 comments, and 27,146 questions have chat rooms (i.e., Q_{chat}). The popularity of question discussions is also increasing, with the proportion of questions with discussions nearly doubling from 32.3% in 2008 to 59.3% in 2018. Question discussions exist in all phases of the Q&A process on Stack Overflow. In questions that are both discussed and have an accepted answer (i.e., $Q_{d/aa}$), discussions in 80.6% of the questions begin before the accepted answer was posted. We found that the duration of question discussions can extend beyond the Q&A process: In 28.5% of $Q_{d/aa}$, question discussions begin before the first answer and continue after the accepted answer is posted; and in 19.4% of $Q_{d/aa}$, question discussions begin after the question receives its accepted answer.

RQ2: How do users participate in question discussions?

We found that 16.0% (i.e., 2.6 million) of the registered users on Stack Overflow have participated in question discussions, which is comparable to the number of users who have answered questions (i.e., 16.7%). Question discussions allow askers and answerers to communicate with each other directly, enabling fast exchanges on the issues of concern. For questions that have both discussions and answers (i.e., $Q_{d/a}$), we found that as the number of comments increases, both askers and answerers were more likely to participate in the question discussions. Also, we found that when there are six or more comments present (i.e., $Q_{hd/a}$), then there is a high likelihood of both askers (90.3%) and answerers (51.9%) participating in the discussions.

RQ3: How do question discussions affect the question answering process on Stack Overflow?

Question discussions tend to lead to more substantial updates to the body of the original question. For example, a median of 114 characters are added to the question body when the question discussion has a chat room instance (i.e., Q_{chat}). While most other questions have no change in their question body length, a larger proportion of questions with comments are revised, with an increase in the question body length compared to questions with no discussion. Questions with more comments receive answers more slowly, with a Spearman correlation of $\rho = 0.709$ between the number of comments and the *answer-receiving-time* for the first answer.

The main contribution of our study is to highlight that discussions are an integral part of the Q&A process on Stack Overflow. Compared to the common assumptions that asking and answering questions are separate events in many studies, our work suggests that a large proportion of questions on Stack Overflow are answered after interactions between askers and answerers in question discussions. Our study suggests that question discussions is a common activity comparable to answering questions on Stack Overflow. Question discussions have a high active user base (i.e., 16.0% of active users), and are also comparable to answering (i.e., 16.7% of active users). We also observed a strong correlation between the number of comments and the question answering speed, suggesting that question discussions have an impact on creating answers. Our findings suggest that question discussions can facilitate the Q&A process since they provide a means for askers and potential answerers to communicate throughout the Q&A process. We encourage future research on Stack Overflow to consider question discussions in addition to leveraging the information in the questions and answers of Stack Overflow.

Paper Organization The rest of this paper is organized as follows. Section 2 introduces Q&A on Stack Overflow and commenting/chatting on Stack Overflow. Section 3 describes how we collect data for our analysis. Section 4 details the results of our empirical study. Section 5 discusses our findings and their implications. Section 6 describes threats to the validity of our study. Section 7 surveys related research. Finally, Section 8 summarizes the findings of our study.

2 Background

2.1 The Q&A Process on Stack Overflow

Stack Overflow is a technical Q&A website where users ask, answer, and discuss questions related to programming and software development. Stack Overflow has been widely embraced by the software engineering community, and has become the largest public knowledge base for programming-related questions. There are 21.9 million questions together with 32.7 million answers on Stack Overflow as of December 2021.

The Stack Overflow Q&A process begins with a user posting a *question* that relates to programming or a similar technical topic. At that point, other users can start to engage either by proposing an *answer*, or by taking part in a *discussion* in the form of a *comment* or a *chat room*. Discussions can be attached to either the original question (i.e., a *question discussion*) or one of the proposed answers (i.e., an *answer discussion*). If a proposed answer successfully resolves the question, the user who asked the original question (i.e., the *asker*) may at their discretion choose to designate that answer as the *accepted answer*. Once an accepted answer has been selected, users may continue to contribute to the question thread by adding new answers or editing existing content; in practice, however, user activity related to that question and its answers tends to diminish sharply at that point (Baltes et al. 2018). We note that the Stack Overflow uses the term *post* internally to refer to either a question or answer, but not a discussion.

2.2 Discussions on Stack Overflow

In this work, we focus on question discussions to better understand how discussions affect the crowdsourced knowledge sharing activities once a question is posted, especially those that occur early in the Q&A process.

Stack Overflow offers two different forms of communication channels for users to discuss on questions and answers, that is, commenting as an asynchronous communication channel and chatting as a synchronous communication channel. When users are commenting, they may not expect an immediate reply. Meanwhile, when users are chatting, a live session is formed where information flows freely within the group in real-time (Storey et al. 2017). On Stack Overflow, users begin discussions in comments. When extended discussions occur in comments, users are proposed with continuing the discussions in dedicated chat rooms. While commenting is the dominating communication channel on the Stack Overflow for question discussions, whenever possible, we take special notice of the existence of chat rooms since they represent a different form of communication channel compared to comments.

As previously mentioned, users can attach comments to a post (i.e., a question or answer). Stack Overflow considers comments as “temporary ‘Post-It’ notes left on a question or

answer.”³ Stack Overflow users are encouraged to post comments “to request clarification from the author; leave constructive criticism to guide the author in improving the post, and add relevant but minor or transient information to a post.” When multiple comments are present in the same post, they form a *comment thread*.

Stack Overflow offers *real-time, persistent collaborative chat for the community*⁴ with chat rooms. Stack Overflow promotes users to continue the discussions in a chat room when there are more than three back-and-forth comments between two users (i.e., at least 6 in total). Users are prompted with a message before a chat room can be created: “Please avoid extended discussions in comments. Would you like to automatically move this discussion to chat?” When the user agrees to create the chat room, an automated comment is posted and contains a link to the newly created chat room. In the newly created chat room, automated messages are posted pointing to the associated question and repeating the comments leading to the chat room. Users can also create chat rooms directly that are not associated with questions or answers.

3 Data Collection

In our study, we use the Stack Overflow data dump from December 2021 (Stack Exchange Data Dump 2021). The data dump is a snapshot of the underlying database used by Stack Overflow; it contains all meta-data for each comment, such as which user posted the comment and which question the comment is associated with. We mainly used the `Posts` and `Comments` table from the dataset to extract the required information. The data dump also contains the history of each question, via the `PostHistory` table. We analyze the history of each question to reconstruct the timeline of when the question was created, edited, commented, and answered.

Data about chat rooms is not contained in the Stack Overflow data dump; instead, we collected it manually by crawling the Stack Overflow website itself.⁵ We also labelled the chat room instances based on whether they are *general*,⁶ attached to a *question*, or attached to an *answer*. After cross-referencing their associated question IDs with the Stack Overflow data dump, we removed chat room discussions that are unrelated to programming, such as those on Meta Stack Overflow which focuses on the operation of Stack Overflow itself. This left us with a total of 27,312 chat rooms comprising 1.5 million messages that are associated with 27,146 questions as of December 2021. Figure 2 shows the detailed extraction process of chat rooms from Stack Overflow.

4 Case Study Results

In this section, we explore the underlying motivation, the approach taken, and the results of our three research questions (RQs) concerning question discussions on Stack Overflow.

³<https://stackoverflow.com/help/privileges/comment>

⁴<https://chat.stackoverflow.com/faq>

⁵We’ve made our dataset open access on Zenodo: <https://zenodo.org/record/5516190>

⁶General chat rooms are standard chat rooms on Stack Overflow that are not associated with a question or an answer.

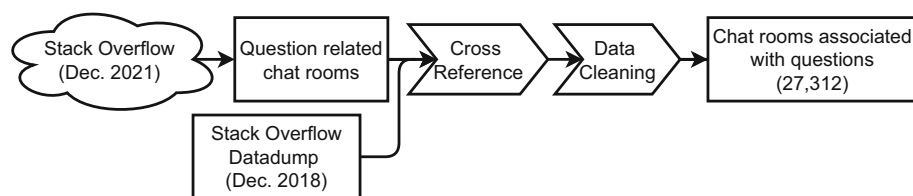


Fig. 2 An overview for the creation of Q_{chat} (questions with chat rooms)

4.1 RQ1: How Prevalent are Question Discussions on Stack Overflow?

Motivation As a technical Q&A platform related to programming, Stack Overflow hosts a large number of questions (Treude et al. 2011). From the user's point of view, creating an answer can be challenging since the initial version of a question is often incomplete or ambiguous. For this reason, potential answerers may first wish to engage the asker in a discussion to clarify their intent and possibly seek additional context, which is typically done using comments attached to the question. If the discussion proves to be fruitful, the user may then post an answer based on the discussion; also, the asker may decide to edit the original question to clarify the intent for other readers. For example, Example 1 shows a comment pointing out a confounding issue in the original question. After the discussion, the asker acknowledged the issue and edited the original question for clarity.

A prior study showed that active tutoring through discussions in chat rooms can substantially improve the quality of newly posted questions by novice users (Ford et al. 2018). However, it is labor intensive to provide such tutoring with an average of more than 7,000 new questions posted per day on Stack Overflow in 2019. At the same time, there has been no detailed study of question discussions as yet; in this RQ, we explicitly study question discussions to gain a better understanding of their prevalence in the Q&A process.

Example 1

In a comment linked to a question titled: “Write to Excel — Reading CSV with Pandas & Openpyxl - Python.⁷”, a user observed that the example CSV file given in the question did not follow the CSV standard, and suggested the asker to double check the input format.

Comment:

The structure of the first three lines doesn't match the structure of lines 5 onwards so you cannot read this file with a CSV library. Please check the provenance of the file and what it should look like. I suspect you probably want to skip the first four lines.

Approach We begin our study of the prevalence of question discussions by investigating the trend in the number and proportion of question discussions over the years. We distinguish

⁷<https://stackoverflow.com/questions/48956597/>

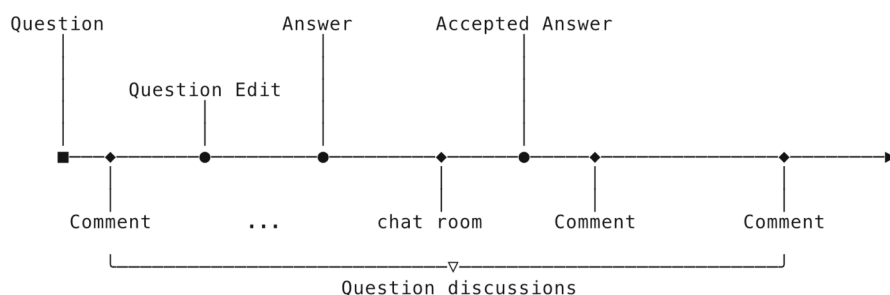


Fig. 3 Timeline of question thread events. Question discussions can occur at any time since the creation of a question

between answered questions with and without an accepted answer to investigate whether there exists a difference between the two groups of questions.

We then study when question discussions occur relative to key events in the Q&A process. After a question is posted on Stack Overflow, several different types of follow-up events may occur, as illustrated by Fig. 3. For example, after a question is posted any of the following can occur:

- other users can propose answers to the question;
- users can post comments to discuss either the question or the associated answers;
- the asker can mark one of the answers as *accepted*; and
- the question (and proposed answers) can be edited for clarity.

For each question, we construct the timeline consisting of each event, and we analyze the prevalence of question discussions with respect to other Q&A activities. Here, we focus mainly on two key events: when the question receives its first answer, and when it receives the accepted answer.

Result Stack Overflow questions are discussed by 43.6 million comments and 1.5 million chat messages, forming a large dataset of community question discussions, in addition to the 22.0 million questions and 32.7 million answers. The proportion of questions with discussions also nearly doubled from 32.3% in 2008 to 59.3% in 2013, and has remained roughly stable since then. Figure 4a shows the number and proportion of questions with discussions per year, and Fig. 4b suggests a similar trend for questions with an accepted answer. Since a question may receive its first comment several years later, it is likely that the proportion of discussed questions in recent years will increase slightly in the future.

Question discussions occur throughout the Q&A process, ranging from before the first answering event to after the accepted answer is posted. Figure 5 shows the proportion of question discussions relative to answering events in the Q&A process. The height of the band across each vertical line indicates the proportion of questions with a specific activity occurring in that phases of a question thread's life cycle. For example, from the left-most bar, all questions can be split into two groups: questions with discussions (Q_{disc}) and questions without discussions (Q_{nd}). The top band (with strata in blue) represents 59.2% of the questions with discussions and the bottom band (with strata in red) represents 40.8% of the questions without any discussions. Flowing from left to right, the strata in blue and red continue to represent the questions with and without discussions until the right most band where it represent the final answering status of the question.

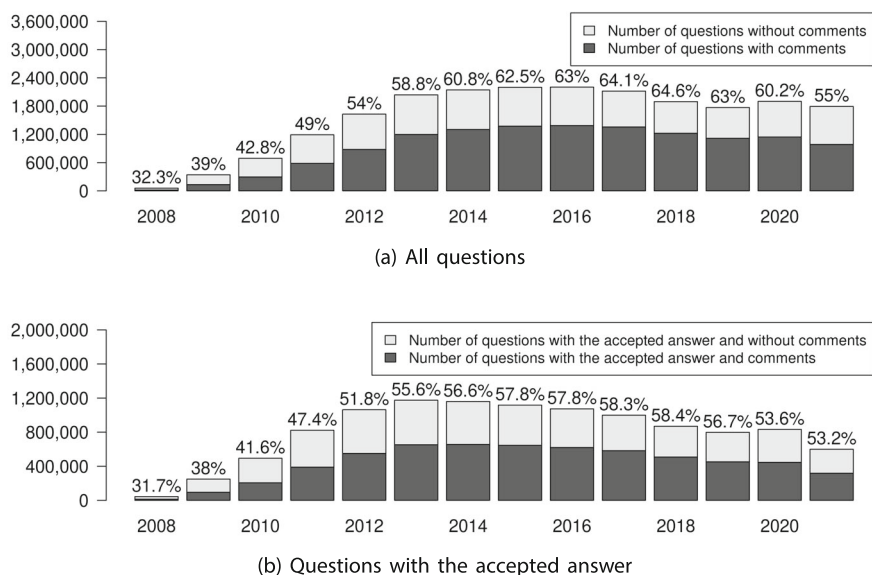


Fig. 4 The number and proportion of questions with comments

In $Q_{d/aa}$, 76.2% (i.e., 8.0 million) of the question discussions begin before the first answer is posted, suggesting an influence of question discussions on answering activities. Furthermore, 80.6% (i.e., 4.9 million) of the question discussions begin before the accepted answer is posted, indicating a slightly more active involvement of question discussions in $Q_{d/aa}$. In answered and solved questions of Q_{chat} , 59.1% (i.e., 12,507) of the chat activities begin before the first answer is received, and 72.9% (i.e., 10,172) of the chat activities begin before the accepted answer is posted.

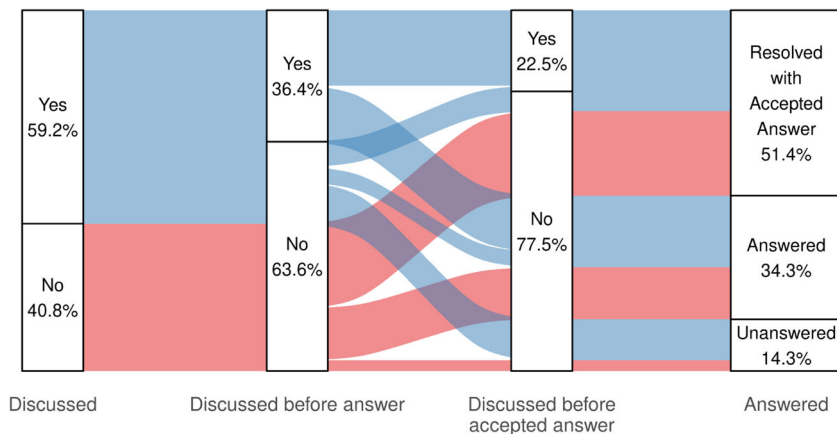


Fig. 5 Question discussion with respect to answering events during the Q&A process. The blue bands represent questions with discussions and the red bands represent questions without discussions

The early occurrence of question discussions in the Q&A process suggests that they enable interested users to engage with the asker informally, to allow for clarification. For example, in Example 2, 13 min after the question was initially posted, a user asked for a concrete example that can demonstrate the problem the asker had. The asker then updated the question with the requested information. The question was answered 15 minutes later, incorporating the newly added information based on the discussions.

Example 2

A user comments to ask for information in a question titled “Can I modify the text within a beautiful soup tag without converting it into a string?⁸”

Comment:

UserB: Please give an example html that demonstrates the problem. Thanks. [2014-09-16 13:15]

UserA (the asker): Just added some example html, sorry about that. [2014-09-16 13:20]

In 28.5% (i.e., 1.7 million) of $Q_{d/aa}$, the discussions begin before the accepted answer has been received, and continue after the accepted answer is posted. Furthermore, 19.4% (i.e., 1.2 million) of the question discussions begin after the accepted answer is posted. These findings indicate that the community may continue to discuss questions even after the asker has designated a “best” answer that solves their problem (Anderson et al. 2012). This may be due to the fact that software development technologies tend to evolve rapidly; old “truths” may need to be updated over time, and additional discussions may provide new insights despite the asker considering the question to be solved. Example 3 shows a comment that pointed out a potential security vulnerability in the code snippet 5 years after the initial question is posted.

Example 3

A user posted a comment to warn about a potential security vulnerability 5 years after a question was posted.⁹

Comment:

Beware. If you've configured your Struts application in this particular way (setting 'alwaysSelectFullNamespace' to 'true'), your application is very likely vulnerable to CVE-2018-11776: semml.com/news/apache-struts-CVE-2018-11776

⁸<https://stackoverflow.com/questions/25869533/>

⁹<https://stackoverflow.com/questions/17690956/>

RQ1 Summary:

There are 44.6 million comments and 1.5 million chat room messages in our dataset, which forms a large corpus of question discussion activities on Stack Overflow. Since the introduction of comments, the popularity of question discussions has nearly doubled from 32.3% in 2008 to 59.3% in 2013 and has remained stable since. The occurrence of question discussions is prevalent throughout the Q&A process. While question discussions in most questions (76.2% in $Q_{d/a}$ and 80.6% in $Q_{d/aa}$) begin before the answering activities, question discussions can continue or even begin after the accepted answer is posted.

4.2 RQ2: How do Users Participate in Question Discussions?

Motivation The crowdsourced Q&A process on Stack Overflow is driven by user participation. In addition to the questions and answers, question discussions are also part of the user-contributed content on Stack Overflow. In this RQ, we explore how different users participate in question discussions, to better understand how question discussions facilitate the Q&A process.

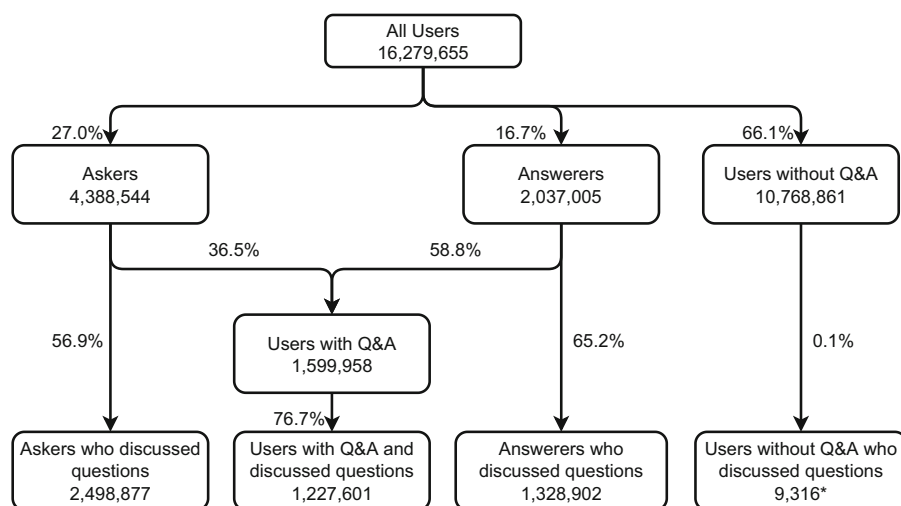
We focus on two aspects of user participation. First, we investigate the overall user participation in question discussions on Stack Overflow. We note that in RQ1, we observed a high proportion of questions with discussions; here, we focus on the users who participate in question discussions. Second, we change the scope to focus on the question-level discussion participation. We are interested in what other activities that the participating users join in on. For example, did the user participating in comments also proposed an answer for the question later.

Approach To study user participation in question discussions and to gain an overall idea of the popularity of discussion activities compared to other activities on Stack Overflow, we extract from the data dump the list of all users who contributed content to Stack Overflow. In particular, we sought users who asked, answered, or discussed questions; we note that while other activities, such as voting, may help the community, we do not consider these activities in our study as they do not directly contribute content. We also ignored activity related to answer discussions, as it was outside of the scope of our investigations.

We extracted the unique *UserIDs* from all questions, answers, and question comments to build the groups of users who participated in each of those activities. We then compared the intersection between the different sets of users to determine which of them participated in multiple types of activities on Stack Overflow.

Results 2.6 million (i.e., 16.0%) users on Stack Overflow have participated in question discussions. Figure 6 shows the overlap of the number of users participating in different activities on Stack Overflow. We observe that 95.7% of users who participated in question discussions also asked questions on Stack Overflow, and 50.9% of them answered questions.

In 60.0% of $Q_{d/a}$ (i.e., 7.8 million), askers participate in the question discussions and in 34.1% of $Q_{d/a}$ (i.e., 3.6 million), an answerer participated in the question discussion. The involvement of askers and answerers indicate that the two parties often leverage question discussions as a collaboration medium.



*This situation is technically not possible according to Stack Overflow's rules. However, due to untracked deletions, users may appear to have never participated in Q&A

Fig. 6 The number of users who participate in different types of activities on Stack Overflow, and the number and proportion of users who participate in question discussions

We further investigate the trend of the proportion of questions with askers and answerers in question discussions as the number of comments increases. **When the number of comments increases, a higher proportion of questions have askers and answerers participating.** Figure 7 shows the trend of the proportion of askers and answerers participating in question discussions as the number of comments increases. When there are at least 6 comments associated with a question (i.e., when Stack Overflow starts to hide additional comments), askers are present in at least 90.3% of the question discussions and answerers are present in at least 51.9% of the question discussions. Moreover, **when answerers are present in a question discussion, 79.3% (i.e., 2.8 million) of the answerers and 81.1% (i.e., 1.5 million) of the accepted answerers joined the question's discussions before posting the answers.** The increasing proportion and early engagements of answerers in question discussions suggest that users are actively leveraging the question discussions as a communication channel to facilitate the answering of questions.

RQ2 Summary:

2.6 million (i.e., 16.0%) users on Stack Overflow have participated in question discussions. These users overlap heavily with users who asked and answered questions on Stack Overflow. In $Q_{d/a}$, 60.0% of the questions have the asker participating in the question discussion and 34.1% of the questions have an answerer participating in the question discussion. The proportion of questions with askers and answerers participating in question discussions increases as the number of comments increases. When at least 6 comments are present, more than 90.3% of the discussions have askers participating and more than 51.9% have answerers participating. In 79.3% of $Q_{d/a}$ (81.1% of $Q_{d/aa}$), the answerer (accepted answerer) participated in the question discussion before they posted the answer (accepted answer).

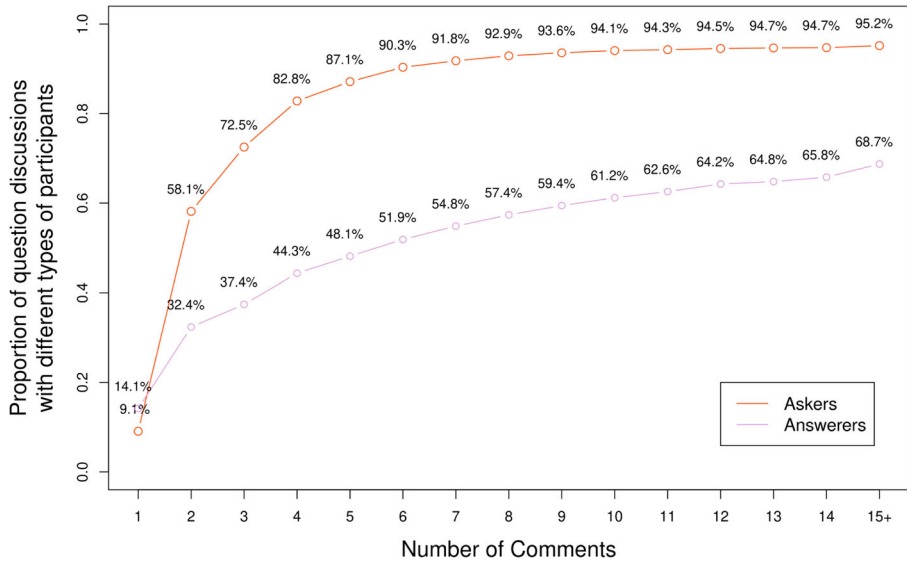


Fig. 7 The proportion of question discussions with the participation of askers and answerers

4.3 RQ3: How do Question Discussions Affect the Question Answering Process on Stack Overflow?

Motivation On Stack Overflow, questions serve as a starting point for curating crowd-sourced knowledge. Devising a good question can also be a challenging task (Calefato et al. 2018). To encourage users to ask high-quality questions, in late 2019 Stack Overflow modified its reputation system to reward more reputation points on upvotes for questions, increasing the points rewarded from 5 to 10.¹⁰ As noted previously, a question can have several follow-up answers; also, discussions can be associated with either the question or its answers. Questions (and answers) may be edited and revised by their original author, and this happens commonly.¹¹ This may be done to reflect new knowledge learned though the Q&A process, and to improve the quality of the posts themselves. In practice, some revisions are editorial or presentational in nature, such as fixing typos and formatting content for readability; however, questions are also edited to improve the quality of the crowd-sourced knowledge (Jin and Servant 2019). Baltes et al. (2018) observed that comments have a closer temporal relationship with edits than posts (i.e., a question or an answer), that is, the time difference between comments and post edits are smaller compared to comments and post creations. Typically, this happens for clarification purposes as answers and discussions shed new light on the original problem. For example, sometimes the asker's question may not include enough technical detail to be easily answered; similarly, the asker may conflate several issues into one posting. In these cases, the asker may seek to clarify the content of their question by adding new context or editing out extraneous details. Also, sometimes new answers emerge to older questions as the accompanying technologies evolve. Thus, it

¹⁰<https://stackoverflow.blog/2019/11/13/were-rewarding-the-question-askers/>

¹¹Comments may be deleted by their author, but they may not be edited in place.

is important to recognize that the question discussions can affect the evolution of the question itself; the question version that appears to a casual reader may have evolved since its original posting.

In this RQ, we study how question discussions are associated with the evolution of questions. More specifically, we study the association between the number of comments and question revisions; we do so to better understand how question discussions affect the evolution of the question content. We also study the association between the number of comments and the *answer-receiving-time* to explore how question discussions affect the Q&A process.

Approach To understand how question discussions affect the evolution of questions, we first study the correlation between question discussions and question revisions. Here, we are mainly interested in the scale of question edits in terms of the size of question content change in the question body. Specifically, we calculate the change in the number of characters in the question body between its initial version and the current version. We also categorize all questions into three groups, i.e., questions with no discussions (Q_{nd}), questions with comments (Q_{disc}), and questions with chat rooms (Q_{chat}). For each question from any category, we calculate the character length difference between the current version of the question and its initial version to investigate how question discussions are associated with the changes in the question content over a question's lifetime.

To understand how question discussions associate with the speed of question answering, we study the correlation between the number of received comments before answering activities and the *answer-receiving-time*. Similar to RQ1, here we investigate the *answer-receiving-time* of two different answering events: the *answer-receiving-time* for the first answer (i.e., t_{FA}) and the *answer-receiving-time* for the accepted answer (i.e., t_{AA}). For each question, we compute both t_{FA} and t_{AA} . We then group the questions by the number of received comments before the first answer and accepted answer respectively. Finally, we measure the Spearman correlation (Spearman 1961) between the number of comments and the median t_{FA} (t_{AA}) for questions with the same number of received comments before the first answer (accepted answer) is posted.

Results **Questions with chat rooms are more likely to be revised than questions without chat rooms, with a median size increase of 114 characters.** Questions without chat rooms, on the other hand, do not exhibit a net change in size, although such questions may still receive edits. Thus, the existence of a chat room attached to a question makes it more likely that the question will undergo significant revision. Figure 8 shows the distribution of questions by the change in question body length after the question is posted, according to different levels of question discussion activities. From the figure, we can observe that while Q_{nd} and Q_{chat} share the same median and modal of zero characters change in question body length, a higher proportion of questions with comments receive revisions that lead to an increase in the question body length.

Overall, **the number of comments is strongly correlated with both t_{FA} (i.e., $\rho = 0.709$, $p \ll 0.05$) and t_{AA} (i.e., $\rho = 0.806$, $p \ll 0.05$).** Figure 9 shows the median t_{FA} and t_{AA} of questions with respect to the number of received comments before their respected answering events. Questions with many discussions also take a longer time to answer. One possibility is that the difficulty of these questions is also higher, therefore requiring more effort by the users to have an extended discussion before the question can be answered. At the same time, for the *answer-receiving-time* of Q_{chat} , we find that it takes a median of 5,935 s (i.e., 1.6 h) and 8,438.5 s (i.e., 2.3 h) to receive the first answer and the accepted answer. The answering time involving chat rooms aligns with the observation that

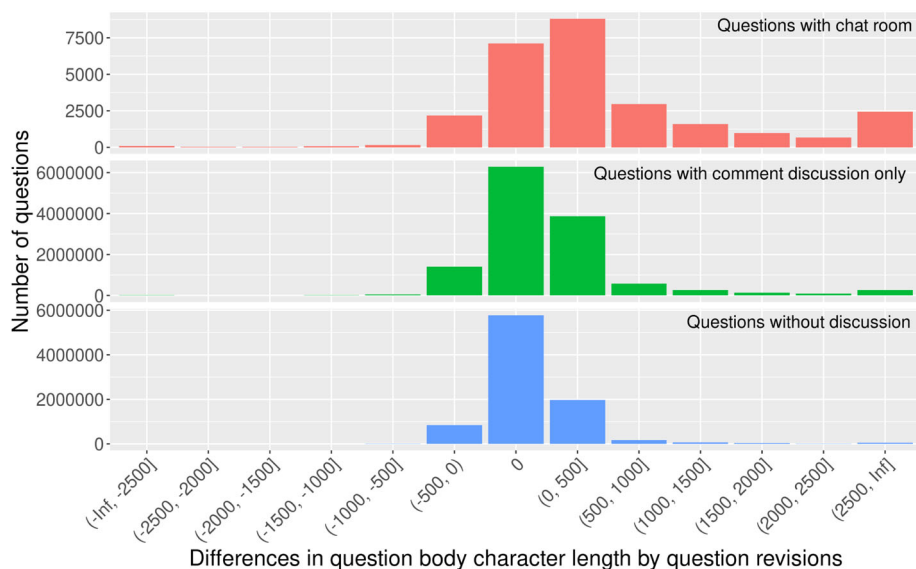


Fig. 8 The distribution of the number of questions to the change in question body character length after the question is posted at different levels of question discussion activity

questions with more discussions tend to require a longer time to answer. The strong correlation between the number of comments that a question receives and the *answer-receiving-time* suggests a close relationship between question discussions and creating answers. Our findings suggest that after a question is asked, interested users may offer help first in comments when an answer can't be created immediately. Therefore, they begin the Q&A process by discussing with the asker through commenting. This is also supported by our observations in RQ1 and RQ2 where discussions mainly begin before answering and a high proportion of answerers participate in question discussions.

RQ3 Summary:

Question revisions for Q_{chat} are more likely to lead to larger edits in the question body, with a median increase of 114 characters to the question body. A strong correlation exists between the *answer-receiving-time* and number of comments before the answer, suggesting its close relationship with answering activities.

5 Implications and Discussions

5.1 Feedback from the Community

We shared our empirical observations on Meta Stack Overflow,¹² where users share ideas and discuss the operation of Stack Overflow. We are glad that the users from the community

¹²<https://meta.stackoverflow.com/questions/416059/>

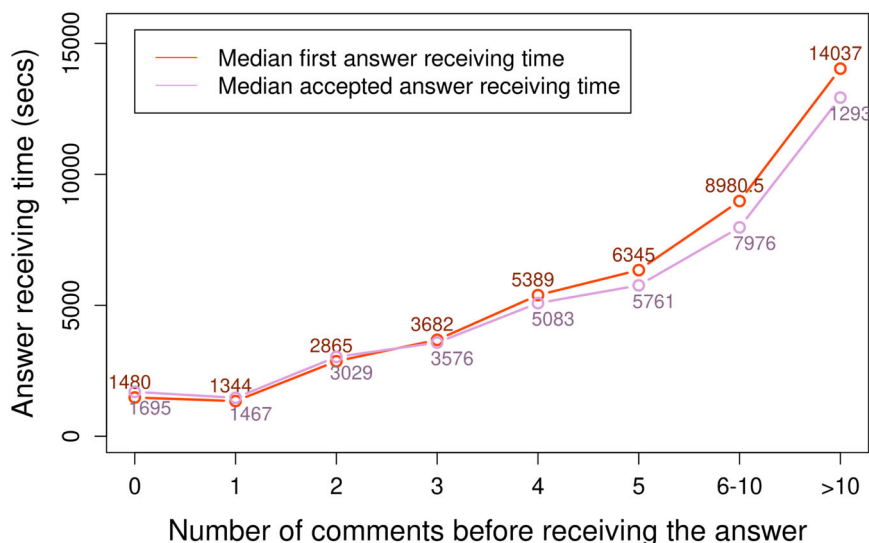


Fig. 9 Median *answer-receiving-time* with respect to the number of comments that are posted before the answer. The median is only calculated for questions with answers and questions with accepted answers respectively

find our observations align with their personal experiences with question discussions on Stack Overflow.

Some users also shared their personal experiences of leveraging question discussions. For example, one user stated “Many questions are very localized (i.e. help only the OP (Original Poster)) and very easy to answer (i.e. obvious to anyone who has any experience). For these, writing a proper answer, with explanations, seems like a waste of time.” It supports our theory that question discussions provide a means for alternative response than an answer. For questions with no answers, users may still find enough information in the question discussion that can be helpful.

Other users also noticed that question discussions may be a sign of new users not realizing the edit feature for questions, “One thing I’ve noticed is that new users don’t seem to realize they can edit their questions. When asked for clarity, they often (attempt to) dump great chunks of information in a comment.” The observation is supported by another user who stated “I always add a suggestion to [edit] the question unless I am sure the user knows how to do it. Such a suggestion is not offensive, and provides the user a convenient button to edit.” These observations also aligns with our findings that discussed questions are often edited more in RQ3.

Some users observed that comments can be deleted on Stack Overflow; future studies may wish to investigate this practice. Since Stack Overflow data dumps capture only a static copy of the website, researchers could monitor newly posted questions in real-time to capture deleted comments.

Another observation the community raised is that “easy questions are often answered in comments”. Users indicate that they find writing a quick comment can often help the asker

quickly. However, this also introduces noise to the platform, and future viewers may be uncertain where to look for the answer.

5.2 Suggestions for Researchers

While Stack Overflow is the dominating platform for Q&A activities for developers, Q&A also exists in other platforms and often in other forms. Future research can focus on the differences between Q&A platforms to better understand the developer's preferences when asking questions. A better understanding of developer's Q&A needs can help us build better platforms and tools to preserve the information from the Q&A sessions across platforms and improve the knowledge retrieval of the information for future users.

Include discussions when modeling Stack Overflow. Many current studies have considered asking and answering questions as isolated events. After a question is posted, other users will read the question and try to answer it. However, our study suggests a different story for many questions. Discussions in the form of comments occur at large scale for questions on Stack Overflow. The prevalence of question discussions with askers and answerers participating significantly in them suggests that they play a key role in the overall Q&A process; consequently, any empirical study of the Stack Overflow Q&A process has much to gain by explicitly considering question discussions in their modeling. For example, many tools have been proposed by researchers to support developers by leveraging Stack Overflow as a knowledge base (Cai et al. 2019; Uddin et al. 2020; Zagalsky et al. 2012). While, these tools mined the content of questions and answers to retrieve relevant information for developers, they do not leverage the information that is contained in question discussions. By considering question discussion in their modeling, we believe the effectiveness of these tools can be further improved with more information.

Design automated tools to highlight information in question discussions. Stack Overflow's overwhelming success with the international software development community is due largely to the high quality of its content, in the form of questions and answers with accompanying discussions. However, maintaining the quality and relevance of such a large knowledge base is a challenging task; a recent study found that low quality posts hurt the reputation of Stack Overflow (Ponzanelli et al. 2014). Because programming technologies evolve quickly, the detailed information in the questions and answers can become obsolete (Zhang et al. 2021a) and requires continual updating. Therefore maintaining a high quality and up to date knowledge base is very important for its users. For this reason, Stack Overflow allows users to edit questions and answers even after a clear consensus has arisen. Stack Overflow, as a major source of information for developers, currently does not have any mechanisms that are dedicated to the maintenance of knowledge on the platform. Since knowledge maintenance is essential to the community, our study shows that users leverage question discussion to aid the maintenance of knowledge in the question content. Previous studies have also observed similar phenomena in answers (Zhang et al. 2021a; Soni and Nadi 2019). We suggest future research to focus on the evolution of knowledge on Stack Overflow via commenting behavior to extract best practices of the process. By understanding the evolution of knowledge content on Stack Overflow, we can design better mechanisms on the platform to better support the community effort in maintaining knowledge. For example, there could be report buttons for questions and answers that can raise flags regarding false information, legacy information, or potential security flaws. Questions with such flags can be then examined by other users and therefore maintaining a knowledge base that is up to date.

5.3 Suggestions for Q&A Platform Designers

Stack Overflow uses a gamification system based on reputation and badges to reward users who participate in the Q&A process; for example, upvotes on questions and answers reward the original poster with reputation points. However, at present upvotes for comments do not boost the reputation of the commenter, so the system does not currently reward participation in discussions.¹³ Since so much effort is put into discussions—as evidenced by the presence of 43.6 million comments and 1.5 million chat messages in the 2021 data dump—this seems like a missed opportunity. Stack Overflow could reward those users who, through their participation in discussions, help to clarify, explore, and otherwise improve the questions and answers themselves; our studies here have shown just how influential question discussions can be on improving the quality of the questions and answers. Rewarding participation in discussions would create a positive feedback loop in the Stack Overflow gamification system, which would in turn encourage more users to engage in discussions.

Acknowledge discussions as essential in the Q&A process, and design systems that incorporate the users’ need for discussions. A good piece of shareable knowledge starts with a good question, and Stack Overflow has practices to help ensure high quality questions. For example, when novice users (i.e., users with newly registered accounts) first ask questions, they are led through an interactive guide on how to ask a good question. The guide includes both conventions (e.g., *tag the question*) and best practices for asking questions (e.g., *include what has been attempted to solve the question*). Although Stack Overflow already has a detailed walkthrough on how to ask a good question, we observed that in practice, discussing and revising questions remains commonplace. At the same time, crowdsourced Q&A is a labor intensive process; for example, a question may take time to attract the “right” answerers or a question may be hard to understand without clarification. In exploring RQ3, we observed that questions with extended discussions—especially those that continue into a chat room—tend to receive more edits to the question body. We conjecture that question discussions can serve as a feedback loop for the asker, resulting in improvements to the questions through subsequent edits. Our observation also echoes a previous study which shows that tutoring novice users before they post their questions can improve the quality of their question (Ford et al. 2018). We wonder if a question quality assurance “bot” might be able to leverage the question discussion data and mining the discussion patterns to further support askers in efficiently getting answers through crowdsourced Q&A.

Offer real-time Q&A for urgent question asking, and encourage users to organize the information for future reading. Question discussions offer a means for askers and answerers to communicate with each other during the Q&A process. Currently, chat rooms are triggered automatically once three back-and-forth comments occur between two users. However, there are cases where two users may wish to start a live conversation immediately. For example, traditionally in the open source community, it is suggested to ask urgent questions in an IRC channel to receive an immediate response (Raymond 2019). However, when users do so, the information during the Q&A session will be buried in the IRC chat log. On the other hand, if a user were to ask the question on Stack Overflow, in exchange for not having an instant response, the Q&A information will remain easily accessible by the public. While Stack Overflow already offers chat rooms as a means for instant and real-time communication, currently the chat room triggering mechanism in posting comments

¹³<https://meta.stackexchange.com/questions/17364/>

is an inefficient communication channel for such need. There exists a potential for users to choose between a synchronous or asynchronous discussion through chat rooms or comments, respectively. For example, Stack Overflow could build in a feature that allows users to indicate if they are available online, and are waiting for an answer. When other users see the indicator, they could directly start discussions in chat rooms, and later update the content of the question based on the discussion. An intelligent communication channel selection bot could be designed to help users seek an effective type of communication by mining the historical data of communication preferences. Furthermore, a content summarization tool could be designed to extract pertinent information from both comments and chat rooms, for future users to better understand the context of the evolution of a question.

6 Threats to Validity

External Validity Threats to external validity relate to the generalizability of our findings. In our study, we focus on question discussions on technical Q&A on Stack Overflow, which is the largest and most popular Q&A platform for programming related questions. As a result our results may not generalize to other Q&A platforms (e.g., CodeProject¹⁴ and Coderanch¹⁵). To mitigate this threat, future work can consider studying more Q&A platforms.

Another threat is that the studied Stack Overflow data dump captures only the current copy of Stack Overflow's website data. For example, users are allowed to delete their comments, answers, and questions. This means that when users delete their comments, they are expunged from the dataset, and we are unaware of how those comments might have affected the rest of the discussion. This concern is also shared by community members as one user stated "the majority of the comments ever posted on Stack Overflow are probably deleted." Meanwhile, since there is always a valid reason for a comment to be removed, another users suggested that "it's actually good that deleted comments are not public and Stack Overflow data dumps only capture the snapshot at the time it was taken. We don't want this kind of comments (i.e., rude/abusive comments¹⁶) to linger for more than a quarter..." Since Stack Overflow releases their data dump quarterly, we perform a comparison between the data dump from Dec. 2019 and the data dump from Dec. 2021. From the 32.9 million question comments in 2019, only 2.1% (i.e., 689,476) comments have been deleted in the newer data dump. So in other words, we are unable to monitor comments that were posted and deleted within the releases of two data dumps. But if the comment survived initially, it'll likely last.

Internal Validity Threats to interval validity relate to experimental errors and bias. Our analysis is based on the data dump of Stack Overflow from December 2021 (the comment dataset) and web crawling in December 2021 (the chat room dataset). While the difference between the data dump and chat room crawling is only a month, Stack Overflow as a dynamic platform is subject to change and the data itself can evolve. Future work can assess our observations on new data and evaluate whether our findings continue to hold over time.

Construct Validity Since the Stack Overflow data dump not include chat room-related data, we mined that data directly from the Stack Overflow website. This means that our crawler

¹⁴<https://www.codeproject.com/>

¹⁵<https://coderanch.com/>

¹⁶<https://meta.stackoverflow.com/questions/326494/>

and the collected data may be subject to errors (e.g., crawler timeout). We mitigate this issue by manually checking a subset of the collected data and verified the correctness of the scripts.

7 Related Work

7.1 Discussion Activities on Stack Overflow

While Stack Overflow is mainly a Q&A platform, in addition to question and answering, it also has many other mechanisms to help with the Q&A process (e.g., the gamification system through reputation points and commenting). In our work, we consider users posting comments associated with questions as question discussions. However, in many other works, a discussion on Stack Overflow can have different meanings. For example, some studies (Linares-Vásquez et al. 2014; Shcherban et al. 2020) have considered the question as a discussion (e.g., the question, all its associated answers, and any comment associated with the question or its answers). In our work, we use discussions to describe commenting activities associated with a specific post (i.e., a question or an answer).

Most previous works on Stack Overflow discussions have a primary focus on answer discussions. Their aim is to better understand the community efforts in improving the crowd-sourced knowledge on Stack Overflow. Zhang et al. (2021a) conducted an empirical study to understand answer obsolescence on Stack Overflow. In their study, comments are used as an indicator of obsolescence for their associated answer. A follow up study by Zhang et al. (2019) examined answer comments and highlighted that the information contained in the comments should not be overlooked when reading their associated answers. After acknowledging the importance of answers, Zhang et al. (2021b) focused on the current commenting mechanism on Stack Overflow and observed that the current presentation of comment information is not optimal for readers. The comment hiding mechanism on Stack Overflow only displays the top five comments with the most upvotes. However, due to most comments never receiving any upvotes, later comments, which are likely to be more informative, are hidden from readers by default.

Comments are also viewed as triggers for post updates. Baltes et al. (2018) observed that post edits often occur shortly after comment posts and suggests that comments and post edits are closely related. Based on this observation, a study by Soni and Nadi (2019) further analyzed how comments affect answer updates on Stack Overflow. Their observation echoes the finding by Zhang et al. (2021a) that unfortunately users do not update their answers even with comments directly suggesting so.

Compared to the current studies on Stack Overflow discussions that mostly focus on answers from the perspective of knowledge maintenance, our study focuses on the question discussions that mainly begin and occur during the Q&A process. In other words, previous works have focused on preserving the knowledge while our work tends to focus more on the creation of the knowledge.

7.2 Leveraging Discussions in Software Engineering

During software development, communication between members of the team is important for the long-term success of the project. Online discussions are a core part of the process, especially in open source projects where developers may be scattered around the world and

rely on a variety of channels to communicate with each other (Storey et al. 2017). Since the advent of ubiquitous e-mail in the 1980s, developers have used mailing lists for discussions about the projects they are working on and interested in. Studies show that the use of mailing lists facilitates the gathering of people with similar interests, and many open source projects still run mailing lists today (Shihab et al. 2009a) (e.g., the Gnome mailing list¹⁷). The mailing list archive is an informative resource for researchers to understand the development of the project. Rigby and Hassan (2007) studied the Apache developer mailing list to learn about the personality traits of developers and how the traits shift during the development of the project. Sowe et al. (2006) studied three Debian mailing lists and constructed social networks of the mailing list to investigate how knowledge is shared between expert to novice participants.

In addition to the asynchronous email exchanges, developers also use real-time communication channels such as IRC for discussions. IRC channels are often used by open source projects as a complement to their mailing list operations (e.g., the #emacs channel on Freenode exists in addition to the project's mailing list). Shihab et al. investigated *GNOME GTK+* (2009a, b) and *Evolution* (2009b) IRC channels to better understand how developers discuss in IRC. Although e-mail and IRC are still in use today, newer and more efficient platforms have also emerged to better support the need for communication. For example, developers report bugs and feature requests on issue trackers (e.g., Jira¹⁸), and ask questions on Stack Overflow (Vasilescu et al. 2014). Vasilescu et al. (2014) observed that in the *R* community, developers are moving away from the *r-help* mailing list to sites like Stack Overflow in the Stack Exchange network since questions are answered faster there. Prior studies examined different communication channels aiming to better understand and improve the communication among developers. Alkadhi et al. (2017) applied content analysis and machine learning techniques to extract the rationale from chat messages to better understand the developers' intent and the decision making process during software development. Lin et al. (2016) studied the usage of Slack by developers and noticed that bots are in discussions to help software developers.

Storey et al. (2017) surveyed how developers leveraged communication channels and observed that real-time messaging tools and Q&A platforms such as Stack Overflow are essential for developing software. Dittrich and Giuffrida (2011) studied developers' communication across different platforms and observed that real-time messaging plays a role in the communication of developers. Their study shows that real-time messaging tools can support the usage of other communication channels (e.g., Skype calls) and provide a means for developers to form social and trust relationships with their colleagues. Chatterjee et al. (2019) analyzed characteristics of Q&A sessions in Slack and observed that they cover the same topics as Stack Overflow. Zhen Wei and Zhang (2019) applied neural networks techniques on real-time messages to automatically capture Q&A sessions. Ford et al. (2018) experimented with using real-time chat rooms for the mentoring of asking questions on Stack Overflow for novice users. Chowdhury and Hindle (2015) leveraged information from Stack Overflow to create a content filter to effectively filter irrelevant discussions in IRC channels.

In our study, we focus on question discussions on Stack Overflow to better understand how they facilitate the Q&A process.

¹⁷<https://mail.gnome.org/mailman/listinfo>

¹⁸<https://www.atlassian.com/software/jira>

7.3 Understanding and Improving Stack Overflow

Prior research investigated how developers leverage Stack Overflow and studied different mechanisms aiming to improve the design of Stack Overflow (Xia et al. 2013; Chen et al. 2018; Zhou et al. 2019; Wang et al. 2020; Ford et al. 2018). Treude et al. (2011) categorized the types of questions on Stack Overflow, and observed that Stack Overflow can be useful for code review and learning the concepts of programming. Wang et al. (2020) studied the edits of answers and observed that users leverage the gamification system on Stack Overflow to gain more reputation points. Prior studies also aimed to understand the quality of the crowdsourced knowledge on Stack Overflow. For example, Srba and Bielikova (2016) observed that an increasing amount of content with relatively lower quality is affecting the Stack Overflow community. Lower quality content on Stack Overflow may also affect how questions are answered. Asaduzzaman et al. (2013) showed that the quality of questions plays an important role in whether a question receives an answer by studying unanswered questions on Stack Overflow. An automated system to identify the quality of posts and filter low-quality content was proposed by Ponzanelli et al. (2014). To improve the quality of the crowdsourced knowledge on Stack Overflow, prior studies aimed to identify artifacts with different properties (Wang et al. 2018a; Ragkhitwetsagul et al. 2019; Choetkiertikul et al. 2015; Tian et al. 2013; Vasilescu et al. 2014; Ye et al. 2017; Zhou et al. 2019). For example, Nasehi et al. (2012) examined code examples on Stack Overflow and identified characteristics of effective code examples. Their study shows that explanations for code examples have the same importance as code examples. Yang et al. (2016) analyzed code snippets of popular languages (C#, Java, JavaScript, and Python) on Stack Overflow and examined their usability by compiling or running them.

Prior studies also examined various supporting processes on Stack Overflow to better understand its operation and improve its efficiency of the crowdsourced knowledge sharing process. Chen et al. (2018) used a convolutional neural network (CNN) based approach to predict the need for post revisions to improve the overall quality of Stack Overflow posts. Several studies proposed approaches to automatically predict tags on Stack Overflow (Xia et al. 2013; Saha et al. 2013; Beyer and Pinzger 2015). Wang et al. (2014, 2018b) proposed an automatic recommender for tags based on historical tag assignments to improve the accuracy of the labeling of tags for questions.

Instead of the extensively studied artifacts on Stack Overflow (e.g., questions, answers, tags), we investigate the question discussions by an empirical study of 43.6 million comments and 1.5 million chat room messages to understand how discussions can facilitate the Q&A process.

8 Conclusions

Question discussions are an integral part of the Q&A process on Stack Overflow, serving as an auxiliary communication channel for many developers whose technical information needs are not fully met within their nominal work environment. Question discussions occur throughout all phases of the Q&A process, especially before questions are answered. In 76.2% of $Q_{d/a}$ and 80.6% of $Q_{d/aa}$, the question discussions begin before the first answer and the accepted answer is posted; furthermore, 19.4% of the question discussions begin even after the accepted answer is posted. Question discussions allow askers and potential answerers to interact and solve the question before posting an answer. In $Q_{d/a}$, askers participate in 60.0% (i.e., 7.8 million) of the questions discussions and answerers

participate in 34.1% (i.e., 3.6 million) of question discussions. When the number of comments increases, a higher proportion of questions are participated by askers and answerers. The *answer-receiving-time* of a question is strongly correlated (i.e., with a Spearman correlation of $\rho = 0.709$) with the number of comments a question receives before its first answer. We believe that our study of question discussions can be leveraged in several ways to improve the Q&A process. For example, an automated triaging system could suggest an appropriate communication channel; also, bots could be designed to warn about questions that seem unclear and might require further clarification.

Acknowledgments We would like to thank the anonymous reviewers for their insightful comments. The findings and opinions in this paper belong solely to the authors, and are not necessarily those of Huawei. Moreover, our results do not in any way reflect the quality of Huawei software products.

Declarations

Conflict of Interests The authors declare that they have no conflict of interest.

References

- Alkadhi R, Laja T, Guzman E, Bruegge B (2017) Rationale in development chat messages: an exploratory study. In: Proceedings of the 14th international conference on mining software repositories, MSR '17. IEEE Press, pp 436–446
- Allamanis M, Sutton C (2013) Why, when, and what: analyzing stack overflow questions by topic, type, and code. In: 2013 10th Working conference on mining software repositories, MSR '13. IEEE, pp 53–56
- Anderson A, Huttenlocher D, Kleinberg J, Leskovec J (2012) Discovering value from community activity on focused question answering sites: a case study of stack overflow. In: Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining, KDD '12, pp 850–858
- Asaduzzaman M, Mashiyyat AS, Roy CK, Schneider KA (2013) Answering questions about unanswered questions of stack overflow. In: Proceedings of the 10th working conference on mining software repositories, MSR '13. IEEE Press, pp 97–100
- Baltes S, Dumani L, Treude C, Diehl S (2018) Sotorrent: reconstructing and analyzing the evolution of stack overflow posts. In: Proceedings of the 15th international conference on mining software repositories, MSR '18, pp 319–330
- Beyer S, Pinzger M (2015) Synonym suggestion for tags on stack overflow. In: Proceedings of the 23rd international conference on program comprehension, ICPC '15. IEEE, pp 94–103
- Cai L, Wang H, Xu B, Huang Q, Xia X, Lo D, Xing Z (2019) Answerbot: an answer summary generation tool based on stack overflow. In: Proceedings of the 2019 27th ACM joint meeting on european software engineering conference and symposium on the foundations of software engineering, FSE '19, pp 1134–1138
- Calefato F, Lanubile F, Novielli N (2018) How to ask for technical help? Evidence-based guidelines for writing questions on stack overflow. *Inf Softw Technol* 94:186–207
- Chatterjee P, Damevski K, Pollock L, Augustine V, Kraft NA (2019) Exploratory study of slack q&a chats as a mining source for software engineering tools. In: Proceedings of the 16th international conference on mining software repositories, MSR '19. IEEE Press, pp 490–501
- Chen C, Chen X, Sun J, Xing Z, Li G (2018) Data-driven proactive policy assurance of post quality in community q&a sites. In: Proceedings of the 2018 ACM human-computer interaction, vol 2
- Chen M, Fischer F, Meng N, Wang X, Grossklags J (2019) How reliable is the crowdsourced knowledge of security implementation? In: 2019 IEEE/ACM 41st international conference on software engineering, ICSE '19. IEEE, pp 536–547
- Choetkiertikul M, Avery D, Dam HK, Tran T, Ghose A (2015) Who will answer my question on stack overflow? In: Proceedings of the 24th Australasian software engineering conference, pp 155–164
- Chowdhury SA, Hindle A (2015) Mining stackoverflow to filter out off-topic irc discussion. In: Proceedings of the 12th working conference on mining software repositories, MSR '15, pp 422–425

- Dittrich Y, Giuffrida R (2011) Exploring the role of instant messaging in a global software development project. In: Proceedings of the IEEE sixth international conference on global software engineering, pp 103–112
- Ford D, Lustig K, Banks J, Parnin C (2018) “We don’t do that here”: how collaborative editing with mentors improves engagement in social Q&A communities. In: Proceedings of the 2018 CHI conference on human factors in computing systems, CHI ’18. Association for Computing Machinery, pp 1–12
- Jin X, Servant F (2019) What edits are done on the highly answered questions in stack overflow? An empirical study. In: 2019 IEEE/ACM 16th international conference on mining software repositories, MSR ’19, pp 225–229
- Lin B, Zagalsky A, Storey M, Serebrenik A (2016) Why developers are slacking off: understanding how software teams use slack. In: Proceedings of the 19th ACM conference on computer supported cooperative work and social computing companion, CSCW ’16 companion. ACM, pp 333–336
- Linares-Vásquez M, Bavota G, Di Penta M, Oliveto R, Shihyanyk D (2014) How do api changes trigger stack overflow discussions? A study on the android sdk. In: Proceedings of the 22nd international conference on program comprehension, pp 83–94
- Nasehi SM, Sillito J, Maurer F, Burns C (2012) What makes a good code example?: a study of programming q a in stackoverflow. In: Proceedings of the 28th international conference on software maintenance, ICSM ’12, pp 25–34
- Ponzanelli L, Mocchi A, Bacchelli A, Lanza M, Fullerton D (2014) Improving low quality stack overflow post detection. In: Proceedings of the 30th international conference on software maintenance and evolution, ICSE ’14, pp 541–544
- Raghithwetsagul C, Krinke J, Paixao M, Bianco G, Oliveto R (2019) Toxic code snippets on stack overflow. *IEEE Trans Softw Eng*
- Raymond E (2019) How to ask questions the smart way. <http://www.catb.org/esr/faqs/smart-questions.html>. Accessed 21 Oct 2019
- Rigby PC, Hassan AE (2007) What can oss mailing lists tell us? A preliminary psychometric text analysis of the apache developer mailing list. In: Fourth international workshop on mining software repositories, MSR ’07, pp 23–23
- Saha AK, Saha RK, Schneider KA (2013) A discriminative model approach for suggesting tags automatically for stack overflow questions. In: Proceedings of the 10th working conference on mining software repositories, MSR ’13, pp 73–76
- Sengupta S, Haythornthwaite C (2020) Learning with comments: an analysis of comments and community on stack overflow. In: Proceedings of the 53rd Hawaii International conference on system sciences
- Shcherban S, Liang P, Tahir A, Li X (2020) Automatic identification of code smell discussions on stack overflow: a preliminary investigation. In: Proceedings of the 14th ACM/IEEE international symposium on empirical software engineering and measurement (ESEM), pp 1–6
- Shihab E, Jiang ZhenMing, Hassan AE (2009a) On the use of internet relay chat (irc) meetings by developers of the gnome gtk+ project. In: Proceedings of the 6th international working conference on mining software repositories, MSR ’09, pp 107–110
- Shihab E, Jiang ZM, Hassan AE (2009b) Studying the use of developer irc meetings in open source projects. In: Proceedings of the 25th IEEE international conference on software maintenance, ICSM ’09. IEEE, pp 147–156
- Soni A, Nadi S (2019) Analyzing comment-induced updates on stack overflow. In: 2019 IEEE/ACM 16th international conference on mining software repositories, MSR ’19. IEEE, pp 220–224
- Sowe S, Stamelos I, Angelis L (2006) Identifying knowledge brokers that yield software engineering knowledge in oss projects. *Inf Softw Technol* 48(11):1025–1033
- Spearman C (1961) The proof and measurement of association between two things. *Am J Psychol (AJP)*
- Stack Exchange Data Dump (2021) <https://archive.org/details/stackexchange>. Accessed 29 Dec 2021
- Storey M, Zagalsky A, Filho FF, Singer L, German DM (2017) How social and communication channels shape and challenge a participatory culture in software development. *IEEE Trans Softw Eng* 43:185–204
- Srba I, Bielikova M (2016) Why is stack overflow failing? Preserving sustainability in community question answering. *IEEE Softw* 33:80–89
- Tian Q, Zhang P, Li B (2013) Towards predicting the best answers in community-based question-answering services. In: Proceedings of the seventh international AAAI conference on weblogs and social media
- Treude C, Barzilay O, Storey M-A (2011) How do programmers ask and answer questions on the web? (NIER Track). In: Proceedings of the 33rd international conference on software engineering, ICSE ’11. ACM, pp 804–807
- Uddin G, Khomh F, Roy CK (2020) Mining api usage scenarios from stack overflow. *Inf Softw Technol* 122:106277

- Vasilescu B, Serebrenik A, Devanbu P, Filkov V (2014) How social q&a sites are changing knowledge sharing in open source software communities. In: Proceedings of the 17th ACM conference on computer supported cooperative work & social computing, CSCW '14. Association for Computing Machinery, pp 342–354
- Wang S, Lo D, Vasilescu B, Serebrenik A (2014) Entagrec: an enhanced tag recommendation system for software information sites. In: Proceedings of the 30th IEEE international conference on software maintenance and evolution, ICSME '14. IEEE Computer Society, pp 291–300
- Wang S, Chen T-H, Hassan AE (2018a) Understanding the factors for fast answers in technical q&a websites. *Empir Softw Eng* 23(3):1552–1593
- Wang S, Lo D, Vasilescu B, Serebrenik A (2018b) Entagrec ++: an enhanced tag recommendation system for software information sites. *Empir Softw Eng* 23:800–832
- Wang S, Chen TP, Hassan AE (2020) How do users revise answers on technical q&a websites? A case study on stack overflow. *IEEE Trans Softw Eng* 46(9):1024–1038
- Xia X, Lo D, Wang X, Zhou B (2013) Tag recommendation in software information sites. In: Proceedings of the 10th working conference on mining software repositories, MSR '13. IEEE Press, pp 287–296
- Yang D, Hussain A, Lopes CV (2016) From query to usable code: an analysis of stack overflow code snippets. In: Proceedings of the 13th international conference on mining software repositories, MSR '16. ACM, pp 391–402
- Ye D, Xing Z, Kapre N (2017) The structure and dynamics of knowledge network in domain-specific q&a sites: a case study of stack overflow. *Empir Softw Eng* 22(1):375–406
- Zagalsky A, Barzilay O, Yehudai A (2012) Example overflow: using social media for code recommendation. In: 2012 Third international workshop on recommendation systems for software engineering. IEEE, pp 38–42
- Zhang H, Wang S, Chen T, Hassan AE (2019) Reading answers on stack overflow: not enough! *IEEE Trans Softw Eng* 1–1
- Zhang H, Wang S, Chen TP, Zou Y, Hassan AE (2021a) An empirical study of obsolete answers on stack overflow. *IEEE Trans Softw Eng* 47(4):850–862
- Zhang H, Wang S, Chen T-H, Hassan AE (2021b) Are comments on stack overflow well organized for easy retrieval by developers? *ACM Trans Softw Eng Methodol (TOSEM)* 30(2):1–31
- Zhen Wei YG, Zhang J (2019) Automating question-and-answer session capture using neural networks. In: 2019 KDD workshop on deep learning for education, DL4ed
- Zhou J, Wang S, Bezemer C-P, Hassan AE (2019) Bounties on technical q&a sites: a case study of stack overflow bounties. *Empir Softw Eng* 06