

Studying the Rise and Fall of Stack Overflow

SOEN-6611 Course Project

Md Manik Hossain
Concordia University

ID: 40108891

manik.ruet08@gmail.com

Montreal, Quebec, Canada

* good motivation. we can discuss the increased popularity of Githubs etc that may affect SO

Through

Abstract—Stack Overflow accumulates an enormous amount of software engineering knowledge on different kinds of programming languages and programming related topics over the last decade. However, at the beginning of the years of the stack overflow users contribute on the community to make it popular and prove it as a helpful as well life changing resource to the programmers, but since 2013 it seems that the activities of the users have been decreased significantly after analyzing the engagement of the users. Similarly, we found the number of posted answers of questions are declining over the years on Stack Overflow which accounted around 30% answers have been decreased since 2013. In this paper, we investigate evolution of Stack Overflow and user engagement as well over the years through two RQs. We found that: 1) the number of answers decreased 30% and users activity on stack overflow is decreasing drastically from 43.02% (2008) to 1.96% (2018), although users are posting questions per years in steady way. Therefore, in 2018, only 5% of the questions got accepted answers and 95% numbers of questions remain unanswered. 2) Users are frequently changing the topics of discussion on Stack Overflow due to the popularity of the tags or programming language. To demonstrate, in the first year, 2008, C# and .net were popular. Consequently, people posted over 10% of total questions in 2008 related to C# and .net but in 2018, programmer posted new technologies (e.g. R, Python) related questions which accounted over 12% by replacing C# and .net. In addition, recently users are loving to post answer of questions related to R and Python programming language which indicates the changing of discussion topics on SO.

Index Terms—Q&A Website, Stack Overflow, Popular Tags, Knowledge Sharing, Comments

I Introduction

Technical Q&A websites are gaining popularity because of sharing knowledge and solving programming issues over internet by the programmers. When programmers face any programming problems they search it over internet and if they fail to receive any answers, by searching, they post questions on Q&A websites. On this websites, community users are helping each other to make the websites popular as well as users gain reputation and badges after answering the questions. As a result, when a developer solved a problem after getting answer from Q&A websites, then they automatically motivate himself or herself to contribute the community users by posting questions, answers, comments and review questions and answers. Because of users continuous effort of helping the other programmers, they are frequently awarded with badges

not always true → who is "they"

and reputation by Stack Overflow authority. By this way, over the past decade, Stack Overflow has become the largest Q&A website for developers. Since Stack Overflow's launch in 2008, it has accumulated 9,810,623 (around 10 million) users, 16,867,398 (over 16.8 million) questions and 26,056,531 (over 26 million) answers. In terms of accepted answer, when askers receive a good answer of their posted questions on SO that users has the option to accept an answer from all the posted answers of a question. This accepted answer means that the author of the question on SO received an answer that worked for them personally, but not every users come back to accept an answer. Hence, Stack Overflow received over 9 million (9,001,281) accepted answer of total Questions. The whole community of SO makes the site popular and improve the quality of questions and answers.

However, over few decades ago we have experienced that there were few website (e.g. Yahoo, MySpace, MSN, Geocities, AOL, Napster etc) have declined¹, but once they were ruled the internet during that time. For example, Geocities was huge and third most popular site on the web with a vast global community in past, but over the years it has declined and most of the people have forgot about those popular websites. Analogously, after analyzing the SO data dump we found that people may start to lose interests on SO because of the decreasing trend of posted questions, answers, comments, user activity or participation. In addition, we also found that the quality of posting questions and answers are decreasing over the years. Although people are posting questions steadily for getting the accepted answer on SO, but we found that the rate of getting an accepted answer of a posted questions is decreasing because the number of unanswered questions are increasing significantly over the year.

This is due to the fact that users may not be interested on Stack Overflow community because of the other technical Q&A websites (e.g. Slack, Quora, Discord etc) which may be helping more compared to SO. Actually, at present there is not only SO community to help programmers or developers but also there are few others competitive websites such as Slack, Quora, discord etc that attract developer due to different kind of factors. In addition, users may receiving speedy answers

¹ Details information are available on this site <https://www.t3.com/features/10-forgotten-websites-that-once-ruled-the-internet>

→ this can be the motivation, but we don't know if SO's decline is related to this

again, this can be a motivation

from other websites and getting money for posting questions and answers. For instance, Quora providing money to the users for posting questions and answers on the Quora community. To illustrate, in 2018, top 10 earning partners of Quora² earned an average of \$9,189 and asked 11,256 questions each, averaging 61 questions per day, whereas the top-earning partner has earned \$26,897 and asked over 8,000 questions, averaging about 45 questions per day [1].

In this paper, we study growth trend of Stack Overflow in order to analyze the evaluation of this crowd-sourcing knowledge based website and how its changed over the years. Therefore, we structure our study by answering the following Research Questions:

RQ1: What is the growth trend of Stack Overflow?

In this RQ, we study the growth trend of Stack Overflow. Our goal is to study the characteristics of how SO evolves over the years and the growth trend of Stack Overflow because of the increasing demand for software development jobs and the booming IT industry. More than half of the questions remain unanswered of total posted questions on Stack Overflow on each year. Therefore, the number of answers is decreasing steadily every year and the posted questions and answers up votes are decreasing. Further, zero votes and down votes are also increasing in reverse trend that certainly indicates the posting quality of both questions and answers as well as the engagement and participation of users on SO.

RQ2: How does the user's activity change over time?

In this RQ, we wish to understand how user activities evolve over time. We desire to investigate if there are shifts in user activities; for example, people focus more on reviewing and revision of questions and answers instead of answering the questions, although reviews and revisions are also decreasing in similar trend like questions and answers. User activity on Stack Overflow is decreasing in a pessimistic way because after analyzing the activity of users we found that below 1 million users are active from 10 million users in the final years. We also found that the user are not interested in answering questions because every year over half of the questions remain unanswered of yearly total questions. In addition, users also changing the topics of discussion in a frequent manner over the years because of the popularity of programming language or technology. We found that new technologies related to Machine Learning and web development are becoming more popular on Stack Overflow, recently.

II Background

As we discussed in Section I about how user motivate themselves to participate and increase their involvement by answering and posting questions on Stack Overflow community. Now, we will expand our discussion. In background, we will discuss about the way of user engagement and participation in Stack Overflow and how they are answering questions in terms of badges and reputation system. We also expand our discussion by adding the importance of upvotes and downvotes

that indicates which questions and answers are most useful and appropriate.

II-A User engagement and participation on Stack Overflow

Stack Overflow gain popularity throughout the world because of the great feedback from the programmers. When programmer faced any problem, they got answer from the Stack Overflow community which motivate users to participate on the community to help other users and gain reputation on Stack Overflow. Therefore, they register to the community to participate and provide answers to their best knowledge so that other users can get help from their provided answers. By this way, they get involve completely and earn badge as well reputation after being registered on SO. To increase their engagement on SO, they try to answer and post questions, do review task and revision of each post(i.e. questions and answers) of other users in order to improve the quality of post (i.e. questions and answers). If any users participate in all tasks that users achieve good badges and reputations according to their participation on SO that reputation system motivate and captivate users to involve more on SO.

II-B The importance of upvotes and downvotes

Voting (i.e. upvotes, downvotes) indicates that which questions and answers are useful and appropriate as well as ensure the quality of posts in terms of problem. To identify problems, upvote and downvote works well on Stack Overflow because when user get any problem on the post either questions or answers they expose this matter through voting system. Whenever users encounter a question, answer or comment, if other users feel it as useful, they vote it up. Therefore, users make upvote when they found the post as a useful questions or answers whereas users make downvotes after realizing the questions or answers as problematic. Hence, erratic questions and answers always flagged with downvotes after manually checking by users. Upvotes indicates good and useful post (i.e. questions and answers) from the community that also ensure posting quality. When upvote flagged that means users got help from the post truly where they did not found any problem from the post. In addition, Upvotes assure the post quality because they found it as a useful answer without containing any error. Fig 1 show that a questions or answer scored 122 (positive score) through upvotes as well as due to the usefulness this answer has been accepted as an accepted answer by the author of the questions.

On the contrary, downvotes are used to indicate that the post (i.e questions and answers) is problematic and the quality of post type are not good or the post is not useful to other users. Therefore, most of the users overlook this type of post that receive downvotes from the other users. Overall, needless to say that upvotes and downvotes are important to figure out the usefulness of the post that judge the post quality by votes. Fig 2 show that a question or answer scored -29 (negative score) as well as this is not useful questions or answer to the other

²<https://quorapartners.quora.com/The-Quora-Partner-Program-in-2018>

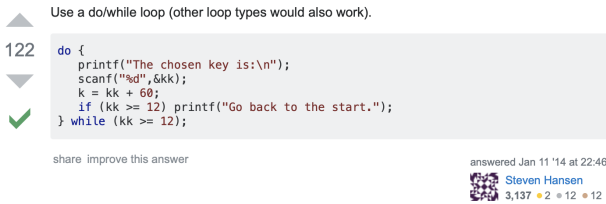


Fig. 1. A question thread where a answer or question got positive scores because of upvotes provided by other developers or programmers to ensure the post quality.

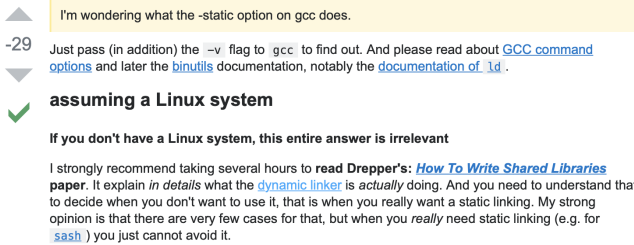


Fig. 2. A question thread where a answer or question got negative scores because of downvotes provided by other developers or programmers to ensure the post quality.

users and the quality of this post was judged by the users through downvotes.

III Evaluation metrics

III-A Evaluation metrics for answering RQ1

We analyze and answer RQ1 by selecting several metrics shown on Table I. We analyze the number of questions yearly along with how many questions have been answered and how many questions still remain unanswered throughout the years from 2008 to 2018, we also investigate the comments ratio on yearly basis. In the end, we also try to analyze the quality of posted questions and answers across the year in order to find out and analyze the overall growth of stack overflow. After analyzing those metrics we found some interesting results that indicate less user interest on Stack Overflow which will discuss through answering RQ1.

III-B Evaluation metrics for answering RQ2

In order to analyze and answer RQ2 we select several metrics shown on Table I. Hence, we evaluate the number of the new user and the active users on a yearly basis across the year. In terms of user activity, we also analyze how many review have been completed over the years. In addition, we also analyze the changing of the topics of different programming language through the top 20 popular tags on Stack Overflow from 2008 to 2018. By following this way, we will answer RQ2.

IV Data Collection

To understand the growing trend of Stack Overflow we collect the data dump from [archive.org](#)³. The data was published on March 04, 2019 by the Stack Exchange community. The data contains information about badges, comments, post history, post links, posts, tags, users, and votes and we import the data dump into MSSQL to facilitate our study. Finally, we generate the CSV file using SQL query in MSSQL. Therefore, we ended up with 9,810,623 (around 10 million) users, 16,867,398 (over 16.5 million) questions and 26,056,531 (over 26 million) answers along with 9,001,281 (over 9 million) accepted answer of total Questions from 2008 to 2018 on yearly basis. We published our query script on github repository. We indicate links of all evaluation metrics data and source code of R as well as SQL Query of the project on README file at Github.com⁴.

V Case Study Results

V-A RQ1: What is the growth trend of Stack Overflow?

Motivation : We wish to study the count of questions and answers that represents the growth of SO because it reflects the issues of developers face and issue that are solved after posting questions and receiving answers on SO. In addition, we desired to study the growth of the Stack Overflow because of the increasing demand for IT industry and software development jobs. At present, a huge number of people is getting involved in software industry. Hence, whenever they are facing the problem, they post Questions on the Q&A community and also helping other users of the community by posting an answer on Q&A websites. In addition, there are a number of Q&A websites are available online for sharing knowledge and many more have been recently added. As a result, we want to investigate the growing trend of Stack Overflow as we know that the crowd-sourcing knowledge-based websites like Stack Overflow is popular among other questions and answer based websites.

Approach: Based on our investigation during the data collection process, we study the data dump and study the trend of Stack Overflow. In particular, we look at several dimensions of the growth of Stack Overflow such as investigating the quality of questions by analyzing the votes of each question on a yearly basis which posted on Stack Overflow, analyze the number of posted questions and answers and comments as well over the years. We performed quantitative analysis. In the quantitative analysis, we captured an overall picture of the growth trend of Stack Overflow in terms of questions, answers and quality of questions and answers which posted on Stack Overflow.

Quantitative Analysis: Evaluation Metrics

1. Study number of questions yearly

We study the number of questions on a yearly basis and we

³<https://archive.org/details/stackexchange>

⁴https://github.com/MMH08/SOEN6611_Project

TABLE I
LIST OF EVALUATION METRICS FOR RQ1 AND RQ2.

Metrics	RQ1 (What is the growth trend of Stack Overflow?)	RQ2 (How does user activities change over time?)
# 1	Study number of questions yearly	Study new and active users yearly
# 2	Study number of answers yearly	Study change of discussion topics through top most 20 tags
# 3	Study post quality of questions and answers	Study review of questions and answers
# 4	Study comments ratio of questions and answers	-

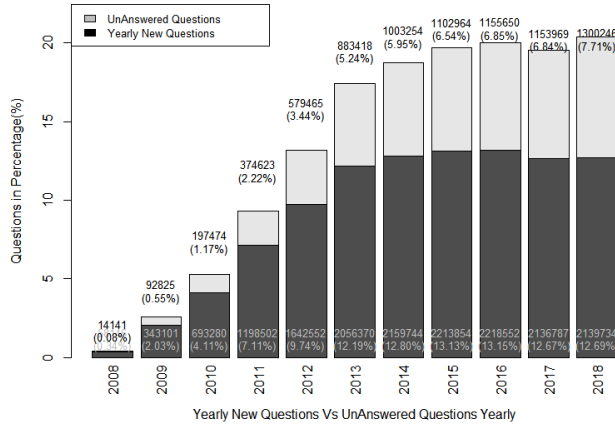


Fig. 3. Posted questions Vs unanswered questions yearly On Stack Overflow

found from Fig. 3 that in 2018, users posted over 2 million questions (12.69% questions of total questions), but among this 12.7% questions, 7.71% questions remain unanswered that means only under 5% (12.69% - 7.71%) questions got their answers. This figure indicates that the number of unanswered questions is increasing steadily over the year and the number of posted questions have dropped, but not so significant.

2. Study number of answers yearly

We study the number of answers throughout the years but we found that the number of answers has been decreasing since 2013. Fig. 4 represent that in 2013, answerer posted 12.79% answers of total answers, after that the number of answers decreased steadily over the years. Therefore, in the final year, 2018, only 9.93% answers posted of total answers. On the other hand, the number of answers increased significantly from 2008 to 2013 (12.79%) before after peaked at 12.8% in 2013 and after that the number of answers dropped to 9.93% in 2018.

3. Study posting quality

we analyzed question quality and find the proportion (%) of positive, negative, zero scores over the years. From Fig. 5, we found that over 75% of the questions got positive scores in 2008 but it has been decreased to around 30% in 2018. In terms of zero scores, over 50% of questions got zero scores in 2018 from nearly 10% in 2008. In addition, negative scores also increasing steadily because in 2008 negative score was

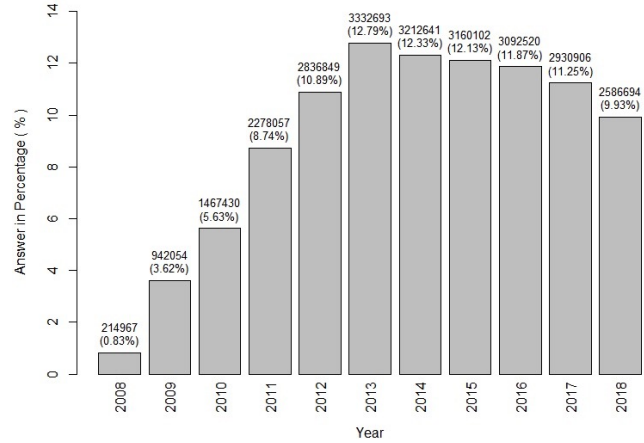


Fig. 4. Show the number of answers posted yearly On Stack Overflow in percentage (%)

0%, but in 2018 it increased to around 10%, as opposed to positive scores of questions have decreased over the year and zero scores increased significantly in reverse trend. Hence, it seems due to the huge amount of zero and negative scores the quality of post are decreasing over the year steadily. Fig. 5 describe our findings regarding the posting quality in terms of upvotes and downvotes of question. Here, positive score increase due to upvotes and negative score increase due to downvotes.

Moreover, we extend our study by analyzing the answers quality in similar way and found the decreasing trend for answers as well. Here we found positive scores decreasing from around over 70% in 2008 to nearly 50% in 2018, whereas negative scores are increasing from over 20% to around 50% in the final year, 2018. Additionally, negative scores seems steady throughout the year. Therefore, this kind of decreasing trend indicate less quality of the answers that describe in Fig. 6. This figure describe our findings regarding the posting quality in terms of upvotes and downvotes of answer. For answers quality same approach also followed like questions, hence positive score increase due to upvotes and negative score increase due to downvotes.

4. Study comments ratio yearly

In order to answer RQ1 we analyze comment ratio that calculate all the comments of each questions and answers in yearly basis. From 2008 the number of comments increase significantly until 2015 from 0.3% to 22.36%, after that the

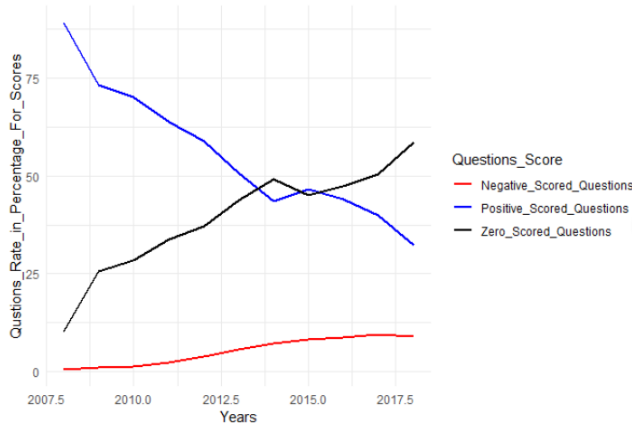


Fig. 5. Show how many questions got positive(up votes), negative(down votes), zero(no-votes) scores in percentage(%) which indicate questions post quality on Stack Overflow across the year.

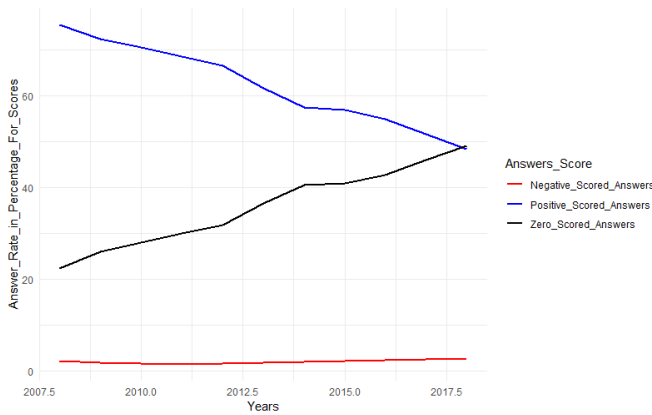


Fig. 6. Show how many answers got positive(up votes), negative(down votes), Zero(No-Votes) scores in percentage(%) which indicate answers quality on Stack Overflow across the year

comments ratio steadily decreased until 2017 that depicts Fig. 7. From 2017 the comments keep dropping to 19.3% in final year, 2018. Therefore, we found decreasing trend from 2015 to 2018.

The number of posted questions on SO are steadily decreasing since 2016, whereas unanswered questions are also increasing steadily in reverse trend. The number of answer on SO are decreasing steadily since 2013, although users are posting their questions. It seems users are not caring about questions anymore. Scores ensure the quality of posts but from this decreasing trend, it is really hard to say about the quality of posted questions and answers. However, it indicates decreasing trend and users inactivity. In addition, the number of comments on each post has decreased in a similar trend.

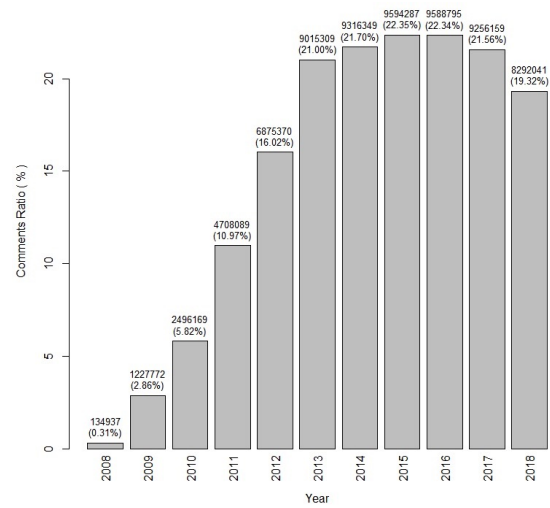


Fig. 7. The proportion of comments that have been posted for each questions and answer in yearly basis.

V-B RQ2: How does user activities change over time?

Motivation : We found that SO is not friendly to new users from blog⁵ where users share their own experience about this issue. Moreover, not a great first impression on SO which leads to stop contribution on the knowledge sharing community SO completely⁶. Therefore, we wish to study how user activities change over time and investigate if there are shifts in user activities, for example, how much users are active on the community to help other users by answering the questions. But, we found that each year over half of the questions remains unanswered of total questions. Therefore, we need to study user activities over the years on Stack Overflow that how many users are answering questions and what is the current status of review for questions and answer over the years. As technology is changing day by day and people tend to work with state-of-the-art technology recently, hence we need to analyze how the discussion topics are changing over the years and how they are engaged with this popular knowledge sharing Q&A based websites like Stack Overflow.

Approach: we first analyze user activities on Stack Overflow through investigating the number of new register users on SO yearly that indicates the user's interest on the websites. In addition, we analyze the active users (who post five or more questions and answers in a year) on the Stack Overflow community. By this way, we analyze the user engagement on Stack Overflow and find out the active users over the years from total users. Moreover, study the discussion topics of the users and how they are changing the topic of discussion on the community in terms of SO

⁵Here is more details about how SO unwelcome new users <https://borngreek.com/2012/01/04/stack-overflow-hates-new-users/> and another useful source <https://borngreek.com/2012/01/04/stack-overflow-hates-new-users/>

⁶Information are available on this website related to the first impression on SO <https://meta.stackexchange.com/questions/22290/are-demands-on-new-users-continually-increasing>

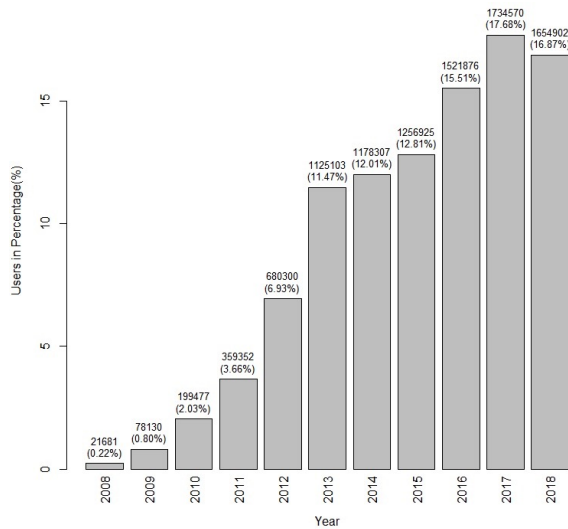


Fig. 8. Show the number of new registered users yearly on Stack Overflow.

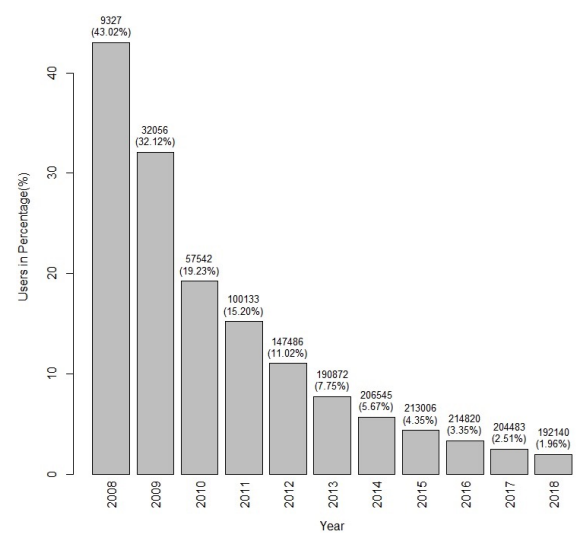


Fig. 9. Show the number of active users yearly on Stack Overflow who post 5 or more questions and answers in a year.

tags. Therefore, we find out topmost 20 tags on SO and we consider top 20 tags that have the most number of questions on the community through SQL query after importing data into MSSQL from SO data dump. we also analyze the top 20 tags that have received the highest number of questions and answers over the year on SO and how the discussion topics through tags are changing over time until 2018. After that, we find out which tags have been receiving the maximum number of questions in the current time compared to past on SO. Finally, we analyze the review task of each questions and answers which accomplished by users on a yearly basis on SO that also manifest the user's participation directly.

Quantitative Analysis: Evaluation Metrics

1. Study new and active users yearly

We wish to study the number of active users yearly and we found that new users have been registered significantly every year with the similar trend but in the final year, in 2018 the number of registered users dropped by 1% from 17.68%, in 2017 to 16.87%, in 2018 which described by Fig. 8. Moreover, we consider active users who post 5 or more questions or answers on a yearly basis. We found that in 2008 below half (43.02%) of the users of that year were active but over the years the percentage of active users has decreased significantly because in 2018, only below 2% (1.96%) users were active that means only 1 million users are active form 10 million users by the end of the year 2018. Fig. 9 depicts the scenario of the active users or activity of users.

2. Study discussion topics through top 20 tags

We also study the top 20 tags on Stack Overflow in order to investigate how users changing the topics of discussion and how many questions have been posted on those popular tags over the years and we found that people frequently change their topics of discussion. From Fig. 10 we found that in

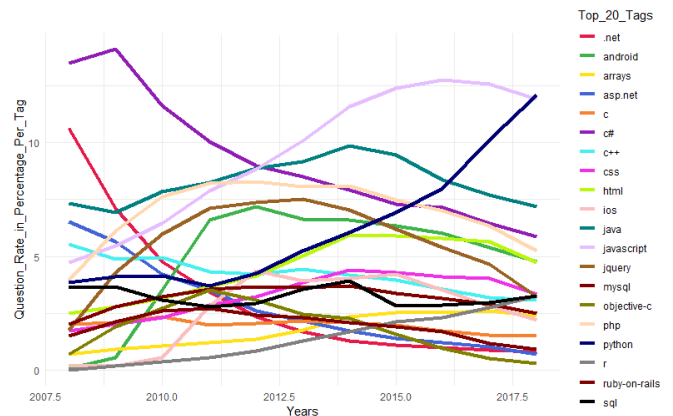


Fig. 10. Indicate how many questions have been posted in percentage(%) for each 20 popular tags on Stack Overflow and also indicate how topics of discussion have changed over the years on Stack Overflow

2008 C# and .net was very popular and users posted over 10% questions of the total questions in that year, but over the years they fall significantly and in the final year, 2018, R and python, array replace these two tags and became popular because now users are posting most of the questions related to these topics. Other tags were fluctuating over the years.

On the other hand, we also study top 20 tags that received most answers and found similar trend like questions. We observed decreasing trend of every almost every tag except python programming language. From Fig. 11, in 2008, answers that are associated with C#, C and .net tags received most answers but over the time its falls significantly and other tags fluctuate over the years. In 2018, python language has becoming more popular among top 20 tags that is recently using to develop scientific software for Machine Learning and receiving more questions and answer because of its popularity.

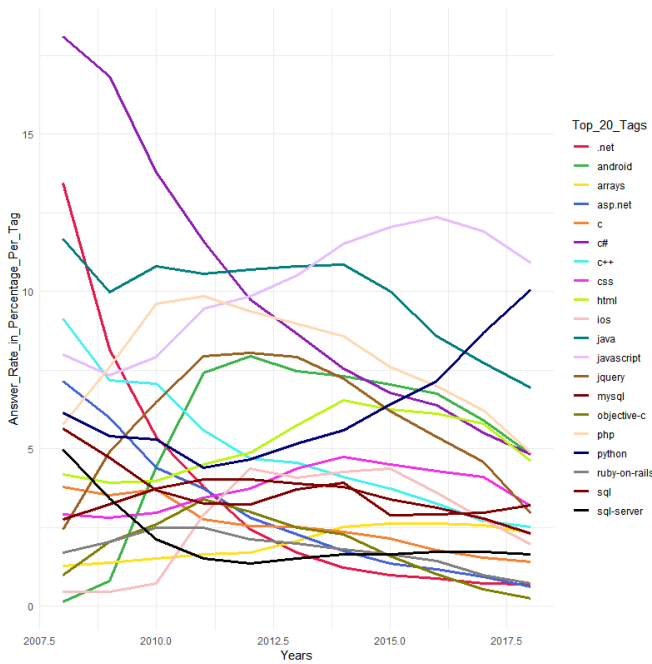


Fig. 11. Depicts how many answers have been posted for each 20 popular tags on Stack Overflow and also indicate how topics of discussion have changed over the years on Stack Overflow

3. Study review of questions and answer yearly

Stack Overflow started review part of questions and answer from 2011. In order to analyze users activity, we also study the review of all the questions and answers yearly across all the years. Hence, Fig. 12 indicate that until 2015 users were so much concern about review in order to improve the quality of questions and answers because the count of review increased in similar trend from 2011, but after 2015 users are not caring about review that manifest the users engagement and participation. As a result, they may be lose interest on review because we observe the number of reviews dropped from 18.68% in 2015 to 15.73% in 2018.

Stack Overflow welcomed a significant number of new users over the years until 2017 before dropping down to under 17% in 2018 of total users. However, in 2008, users were so much active (over 43% of the total users in that year), but over the years its decreased in similar trend before falling down to only below 2% users in the last year that means only less than 1 million are active in 2018. Therefore, the rest of the users are inactive (around 9 million from 10 million users) which indicate their participation and engagement on Stack Overflow community. In addition, it seems that new technologies related to Machine Learning and Web Development are becoming more popular on Stack Overflow recently.

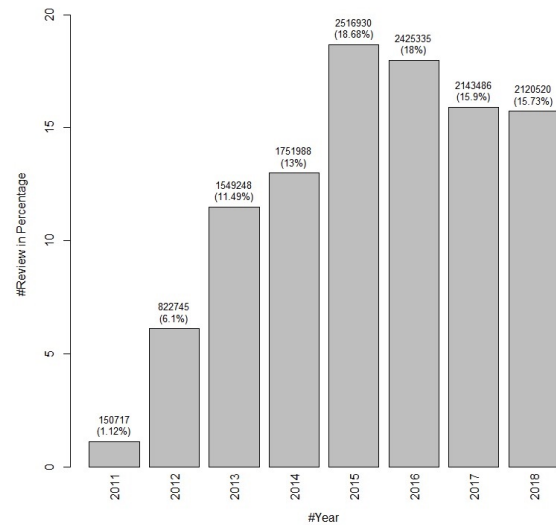


Fig. 12. Show the number of review of each questions and answer that have been done by users yearly on Stack Overflow

VI Discussion

In recent times, programmers are not getting the answer promptly and they became frustrated after posting first questions on SO which lead to stop contribution. For example, a question has been posted on SO community and the author had to wait 5 years to get the first answer ⁷. As we mentioned in Section I that there are several other technical Q&A websites (e.g. Slack, Quora, Discord etc) on internet which are competitor of Stack Overflow where users are helping to each others and they may getting first answers. Therefore, users may focusing on these sites and stop contribution on Stack overflow. Needless to say, we found that one-time asker or answerer are increasing day by day on Stack Overflow. When users do not receive any answer from the community they become frustrated, hence they stop contribution and never come back on SO community. To demonstrate, we found 5,368 one-time asker in 2008, but in 2018, this number increased to 497,473. Similarly, in 2008, we calculate 3,913 one time answerer but in 2018 that number increased to 286,038. Therefore, the one-time askers and answerers are increasing more and more over the time. We will suggest to the authority of the SO to take actions in order to allure users so that they answer the questions speedy to help other users and earn more badge as well as reputations.

Stack overflow should improve the approach of welcoming new users because community show the hostile behavior towards them. At present, users are experiencing hostile behaviors from the community which leads to stop the contribution on SO and sometimes users became victims of the trolls by the Stack Overflow community. These kind of

⁷Details information are available here <https://medium.com/@trieloff/yes-its-getting-harder-to-get-an-answer-on-stackoverflow-3f8e5954d324>

bad experience⁸ share by the victims of the community after facing this kind of unexpected attitude which pessimist them to contribute or participate on the world largest community. Therefore, the activity of users on SO are decreasing day by day because we found that recently below 2% (only below one million users active from 10 million users) users are active on SO community, but at the beginning of the year, 2008, almost 43% users were active and this number has been dropped steadily over the year. Hence, the authority of the Stack Overflow should take action to increase the generosity between the users on the community. By this measure, participation and engagement of users would be increased and they will be more active on this site instead of stopping contribution on Stack Overflow.

New users rarely ever follow the guideline of the SO community which leads to low-quality questions or answers on Stack Overflow. When user registered for the first time they immediately post questions or answers without following any rules and regulations of the community. Therefore, the quality of post are declining and other existing users gives downvotes to their posts to indicate that this post is problematic and difficult to follow. In 2018, we observe over 60% of questions got zero scores and positive scores decrease from over 75% to nearly 30% in 2018 that manifest the poor quality of posting questions or answers. We found nearly same scenario for the quality of answers where SO users know that post score (upvote and downvote) indicates the posting quality on SO. As a result, the authority of Stack Overflow should improve approach of registering new users on the community. We will suggest to promote a "Take a Tour" guide for new users about the community rules and regulations. By following this approach, new users would be able to learn about the content or quality of post from the beginning without facing any difficulties.

VII Thread to validity

External validity: Threats to external validity are related to the generalization from our study. In this study, we focus on Stack Overflow that is the most popular technical Q&A sites in the world. However, our findings may not observe well to other Stack Exchange sites. In order to curtail the external validity, in future, we will expand our studies on other Stack Exchange websites such as Server Fault, Super User, and Ask Ubuntu. We will find how the users of this websites are contributing and the current growth trend in their specific domains.

We conduct two quantitative analysis for our case study in Section V. Firstly, we investigate count of questions, answers and comments to represent developers face and overall growth trend of Stack Overflow. In addition, we expand this analysis by studying post (i.e. questions and answers) quality that associate with upvote and downvote. Secondly, we conduct study on finding the new users and the user activities (i.e.

posting questions, answers and do review tasks) on this largest knowledge sharing technical website, Stack Overflow. Moreover, we expand this study with the investigation of changing the discussion topics of top most 20 tags for questions and answers both. As it is not possible for us to study manually all the questions, answers, comments, users and each user activities in this study, we select statistically representative random samples of all questions threads with a 95% confidence level and a 5% confidence interval in order to minimize the bias.

Internal validity: Threats to internal validity are related to experimental errors. We conduct qualitative analysis in both RQs and tried to alleviate experimenter bias and error. To assuage the bias and experimental error, evaluation metrics for both RQs were determined by the two authors of this study and discrepancies were discussed until a consensus was reached.

VIII Related works

VIII-A Improvement of the Posting Quality on Stack Overflow

As we mentioned in Discussion Section VI that user should go through a obligatory "Take a Tour" guide by Stack overflow before posting any questions or answers in order to increase the quality of posts (i.e. questions and answers) that help to follow the posts by other users. By following this mandatory approach, new users will be able to know the way of posting questions and answers on Stack Overflow community in better way which improve the quality of questions and answers. If users do not know the procedure of posting questions and answer, negative and zero scores will increase. Actually, we found that in 2008 below 10% questions got zero-scores, but in 2018 it increased to around 60%, whereas positive score have deceased drastically from over 75% in 2008 to nearly 30% in 2018. Similarly, negative scores are also increasing which accounted below 10% in 2018. In addition, negative and zero scores of answers are increasing, while positive scores are decreasing in similar fashion likewise question from 2008 to 2018. As a result, posting (i.e. questions and answers) quality has been deteriorating on Stack Overflow.

Wang et al. study how one user is communicating or interacting with other users on Stack Overflow and analyzed how they are acting on SO who are looking for answers and who are posting answers to understand in a better way than how knowledge is transformed to individual users [4]. Srba et al. analyzed low-quality of content created by undesired groups of users on Stack Overflow which has negative impacts on the community and also proposed a way to solve the problems [8]. Wang et al. analyzed how users revise answers on Stack Overflow under the current badge and reputation system [6] and produce some important suggestions in order to improve Stack Overflow revision system. Vasilescu et al. analyzed how users transfer questions in order to support R user mailing list to crowdsourcing knowledge sharing platforms [5] and the author found that users can get faster answers from technical Q&A sites instead of mailing list. Ponzanelli et al. find a captivating way to recognize low-quality questions and answer on Stack Overflow posted by developers [7].

⁸More information are available about this experience shared by the SO users here <https://stackoverflow.blog/2018/04/26/stack-overflow-isnt-very-welcoming-its-time-for-that-to-change/>

A number of techniques have been introduced so that tags can automatically generate when they post questions to help new users [9] and also helping users to find out the proper websites to ask proper questions [9]. Our study aims to find out the growth trend and users activity over the years on Stack Overflow by investigation different dimension with several evaluation metrics. Different from prior studies that only focus on the questions and answers, we focus on an overall growth trend and users activity as developers are not getting answers fastly. We evaluate questions, answers, users, review and users activities on SO.

VIII-B Receiving fast answer in technical Q&A websites

At present, developers are not getting answers fastly after posting questions on So, even though S. Wang et al. provides several approaches to the authorities of Stack Overflow to improve their badges and reputations system in order to captivate users to get answers fast [2]. A set of metrics shown by Asaduzzaman et al. [11] to propose that they are related to the question and poster degree to consider the remaining time for an unresponsive question (did not received any answer) to be answered. A study accomplished by Asaduzzaman et al. [11] on those questions which have not been answered yet on Stack Overflow and also analyzed the problems why a huge number of questions remain unanswered that means they did not receive any accepted answer so far. The author explain this way, for example, questions are complex to understand, too vague, concise questions, nonetheless difficult to identify as a questions for the Stack Overflow community. A very good prediction model proposed by Rahman and Roy [12] employing metrics related to user behavior, topics, and popularity of a question to determine unresolved questions. Different from their work, we study the overall growth trend by analyzing questions, answers comments users and user activity on Stack Overflow. We also consider 20 different popular tags on Stack Overflow and how the discussion topics are changing over the year.

IX Conclusion

The overall growth trend of Stack Overflow is decreasing because we found a declining trend through our evaluation metrics of two RQs. In terms of first RQ, we investigate the yearly questions, answers as well as we also analyze the quality of the post. We found that the quality of posting questions on Stack Overflow are decreasing over the year because recently around 50% of the questions are getting negative and zero scores as we discussed in our first RQ. We also found that the number of questions and answers are reducing with a similar decreasing trend in each year. With regards to the second RQ, we analyze the new register users on Stack Overflow and find that users are registering in the increasing trend until 2017 (slightly decreased in 2018) but only 2% (below 1 million) users are active from around 10 million users on Stack Overflow. Moreover, a significant number of users are inactive (98%, around 9 million) which indicates that the users are not involve with SO completely, hence their activities

dropped significantly. As a result, programmers or coders are not getting answers promptly because users are not interested more compared to the past and they are not caring anymore about answering questions. Furthermore, we also analyzed the review of questions and answers, but the review also decreased over the years which potentially indicates users engagement and inactivity. Finally, we study the changing topics in terms of the top 20 popular tags to analyze the discussion areas and how they are changing over the years. we found programmers are changing their discussion topics frequently. For example, in 2008, .net and C# were popular topics to discuss but over the years that have been changed in 2018 and replaced by python, r and array that means new technologies related to machine learning and web development are becoming more popular on Stack Overflow.

References

- [1] The Quora Partner Program in 2018 [Online]. Available: <https://quorapartners.quora.com/The-Quora-Partner-Program-in-2018>
- [2] Shaowei Wang, Tse-Hsun Chen, Ahmed E. Hassan. Understanding the Factors for Fast Answers in Technical Q&A Websites: An Empirical Study on Four Stack Exchange Websites. *Empirical Software Engineering* (2017).
- [3] H. Zhang, S. Wang, T. P. Chen, Y. Zou, and A. E. Hassan, "An empirical study of obsolete answers on Stack Overflow," 2018. [Online]. Available: <https://doi.org/10.5281/zenodo.1320311>
- [4] S. Wang, D. Lo, and L. Jiang, "An empirical study on developer interactions in StackOverflow," in *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, ser. SAC '13, 2013, pp. 1019–1024.
- [5] B. Vasilescu, A. Serebrenik, P. Devanbu, and V. Filkov, "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& 38: Social Computing*, ser. CSCW '14, 2014, pp. 342–354.
- [6] S. Wang, T. P. Chen, and A. E. Hassan, "How do users revise answers on technical Q&A websites? A case study on Stack Overflow," *IEEE Transactions on Software Engineering*, pp. 1–15, 2018.
- [7] L. Ponzanelli, A. Mucci, A. Bacchelli, M. Lanza, and D. Fullerton, "Improving low quality Stack Overflow post detection," in *2014 IEEE International Conference on Software Maintenance and Evolution*, 2014, pp. 541–544.
- [8] I. Srba and M. Bielikova, "Why is Stack Overflow failing? preserving sustainability in community question answering," *IEEE Software*, vol. 33, no. 4, pp. 80–89, July 2016.
- [9] S. Wang, D. Lo, B. Vasilescu, and A. Serebrenik, "Entagrec ++: An enhanced tag recommendation system for software information sites," *Empirical Software Engineering*, vol. 23, no. 2, pp. 800–832, 2018.
- [10] X. Xia, D. Lo, X. Wang, and B. Zhou, "Tag recommendation in software information sites," in *Proceedings of the 10th Working Conference on Mining Software Repositories, MSR '13*, San Francisco, CA, USA, May 18–19, 2013, 2013, pp. 287–296.
- [11] Asaduzzaman M, Mashiyat 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*, pp 97–100.
- [12] Rahman MM, Roy CK (2015) An insight into the unresolved questions at stack overflow. In: *Proceedings of the 12th working conference on mining software repositories, MSR '15*, pp 426–429.