What Edits Are Done on The Highly Answered Questions in Stack Overflow? An Empirical Study

Xianhao Jin
Virginia Tech
xianhao8@vt.edu

Francisco Servant
Virginia Tech
fservant@vt.edu

***Abstract*—Stack Overflow is the most-widely-used online question-and-answer platform for software developers to solve problems and communicate experience. Stack Overflow believes in the power of community editing, which means that one is able to edit questions without the changes going through peer review. Stack Overflow users may make edits to questions for a variety of reasons, among others, to improve the question and try to obtain more answers. However, to date the relationship between edit actions on questions and the number of answers that they collect is unknown. In this paper, we perform an empirical study on Stack Overflow to understand the relationship between edit actions and number of answers obtained in different dimensions from different attributes of the edited questions. We find that questions are more commonly edited by question owners, on bodies with relatively big changes before obtaining an accepted answer. However, edited questions that obtained more answers in a shorter time, were edited by other users rather than question owners, and their edits tended to be small, focused on titles and in adding addendums.**

***Index Terms*—Empirical software engineering, Developer forums, Collaborative editing, Software maintenance.**

I. Introduction

Stack Overflow (SO) is one of the most popular question-and-answer (Q&A)websites for software developers, and it provides a large amount of code snippets and free-form text on a wide variety of topics [3]. It allows users to share knowledge and expertise, and it has become a large knowledge repository for developers, who use it to communicate on technical problems and resolve programming issues [2].

Given the large competition for obtaining answers to questions in SO, developers will be interested in writing high-quality questions in order to obtain more answers. Recent work found that the quality of edited questions is higher than that of those that were never edited [8]. Some examples of edits to improve question quality could be: adding code examples, fixing spelling mistakes, specifying needs, or just generally provide better explanations.

SO enables and encourages community editing (as a wiki-like system), which means that one is able to edit not only their own posts, but also others' posts. Among the edited questions, some obtain more answers than others, and some obtain their answers in a shorter time than others. This could mean that some edits are more beneficial than others in terms of their impacted questions obtaining more answers or faster answers.

In this paper we study the relationship between question edits and answer collection, as well as the types of edits that SO users perform, and how different types of edits have different relationships with answer collection. Our hypothesis is that editing questions would help collect answers, but that there is a gap between the editing actions that people normally perform and the edit questions that are more prevalent in questions that obtained more answers. We focus on four attributes of question edits: *who* edited the question, *what part* was edited, *how much* was edited, and *what type* was the edit. We study the relationship between these four attributes and the number of answers that questions obtained.

To guide our investigation, we derived the following set of research questions:

- **RQ1**: Did questions with higher number of edits also obtain higher number of answers?
- **RQ2**: How common are different question edits?
- **RQ3**: Which question edits are more common in questions with a higher number of answers?

We perform an empirical study over the dataset provided for the MSR Mining Challenge 2019 — SOTorrent [3]. We defined several metrics to measure the characteristics of the edits that developers make to questions. Then, we studied the relationship between those characteristics and the number of answers that questions obtain, as well as the time that they take to obtain them.

With RQ1, we study whether edited questions overall obtain more answers than unedited questions do. Li et al. [12] observed evidence of more edited questions obtaining more answers for a smaller dataset, and we will study whether that finding generalizes to our dataset too. If edited questions obtain more answers, then one could learn from the edits that were performed on them. To answer RQ1, we compared the answer collection of edited question and unedited question in order to understand whether the edit action on questions is indeed able to help improve the answer collection.

With RQ2, we aim to understand the edits that developers normally perform. Part of our goal is to be able to compare the edits that developers normally make with the edits that are more common in questions with more answers (RQ3) — to find if there is a gap and whether we should recommend developers to make different edits. To answer RQ2, we analyze the distribution of the four attributes of edited questions that we mentioned above over different periods of time.

With RQ3, we aim to understand the characteristics of the edits that happened in questions which successfully obtained

many answers. To answer RQ3, we analyzed the relationship between the four attributes mentioned above, and two dependent variables: the number of the answers obtained, and the time to obtain an accepted answer.

We found that edits before questions obtained an accepted answer were most commonly made by the question owner, on the body, and to clarify meaning, and were not small — which conveys that at that point edits are mostly focused on increasing question quality to obtain answers. In contrast, edits after accepted answers were most commonly made by other users, on tags, and to add related resources — which conveys that at that point efforts are focused on documenting well the question and answer. Finally, we found that the most common edits in questions with many answers are different from what developers edit before obtaining an answer. The edits that obtained more answers were done by other users, in the title, adding related resources, and they were small.

II. Related Work

Other research studied the characteristics of SO questions and the efforts to improve question quality [9]. Yang et al. [16] provided a two-step approach for the automatic suggestion of the most likely editing actions for a newly created question. Other work by Duijn et al. [8] classified good and bad quality questions to understand how code quality influences question quality. In other earlier work, Correa and Sureka performed the first large scale study on poor quality (deleted) questions [7]. There are also some following up works [1] [17] studying duplicate questions in SO.

Other related research tried to understand edit actions in SO. In a recent study, Vargo et al. [14] found that edit actions by high-reputation users can be used to identify bad questions. Yang et al. found that the majority of edits are body edits, and that grammar and spelling types of edits happen more frequently than other types of edits, and then developed an edit-assistance tool [6] as well as a proactive policy assurance mechanism [5]. In this paper, we also study how common different edit types are, but we do it for multiple periods of time to understand which edits are favored under different contexts. Li et al. did an empirical study on SO to understand the trade-offs of introducing a collaborative-editing model to Q&A sites. They found that the benefits of collaborative editing significantly outweighed the risks of losing contributions [12]. We provide more details about the properties of edited questions, such as the distribution of edits at different points in time.

Our research complements existing studies by focusing on comparing the distribution of edit actions overall with their distribution in questions with a high number of answers, and with their distribution at different points in time in the evolution of the question.

III. Data set and Data preprocessing

We perform our empirical study over the dataset provided as a part of the MSR Mining Challenge 2019 called SOTorrent [3]. SOTorrent, an open dataset based on data from the official SO data dump and the Google BigQuery GitHub (GH) dataset, provides access to the version history of SO content at the level of whole posts and individual post blocks [6]. The current release of the dataset contains the version history of all 40,606,950 questions and answers in the official SO data dump with 63,914,798 post versions, 122,673,430 text block versions, and 77,578,494 code block versions, ranging from the creation of the first post on July 31, 2008 until the last edit on June 3, 2018 [4]. For our analysis, we randomly sampled 10,000 questions with accepted answers every year from 2013 to 2017 — 50,000 in total. In this random sample, there were 51.22% questions that had been edited, receiving 2.38 edits on average. We applied LOD [13] to remove outlier questions that had higher or lower than three standard deviations above or below the mean number of edits — leaving us with 49,323 questions to study.

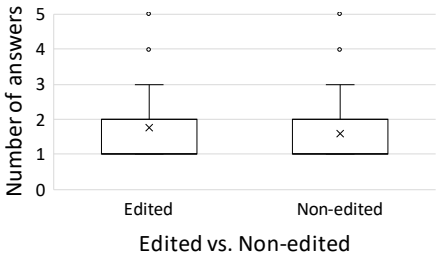

Fig. 1: Comparison on answer collection.

IV. RQ1: Did questions with higher number of edits also obtain higher number of answers?

We first study whether edited questions obtain more answers than unedited questions. The goal of this first investigation is to understand if editing posts can be beneficial, and to replicate what Li et al. [12] found for a smaller dataset: that more edited questions obtain more answers.

Result RQ1: Questions with more edits also obtained (slightly) more answers. We plot the distribution of number of answers obtained by a question, both for unedited questions and for edited questions. Figure 1 shows that box plot. We found that edited questions obtained a mean of 1.75 answers, whereas unedited questions obtained a mean of 1.58 answers. This difference was statistically significant (Mann–Whitney U test, $p < 0.01$). This shows that edited questions do get more answers, but the difference is relatively small.

V. RQ2: How common are different question edits?

To better understand the kinds of edits that people make in questions, we measured the distribution of four characteristics of question edits: *who* edited them, *what part* was edited, *how much* was edited, and *what type* was the edit. We measured who edited questions in two categories: the original creator of the question (owner), and other other users - more than 90% of other users are trusted contributors [6]. We characterize the type of edits into to the four common edits in Stack Overflow's edit guidelines: 1) to fix grammatical or spelling mistakes, 2) to clarify the meaning of a post without changing it, 3)

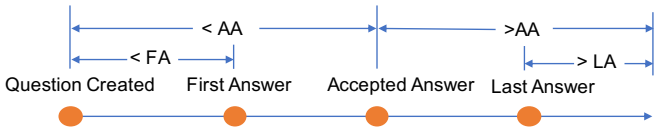

Fig. 2: Time point and interval

TABLE I: Distribution of kinds of question edits.

	Atttribute	Overall %	< FA	< AA	> AA	> LA
part	tile	13.6%	39.6%	46.7%	53.3%	43.3%
	body	68.9%	45.5%	54.9%	45.1%	37.1%
	tag	17.5%	34.9%	40.8%	59.2%	49.4%
who	owner	44.7%	51.6%	63.4%	36.6%	29.6%
	other user	55.3%	35.7%	41.5%	58.5%	48.5%
type	type 1	46.7%	33.8%	40.7%	59.3%	49.0%
	type 2	15.7%	40.9%	54.2%	45.8%	36.4%
	type 3	24.9%	41.5%	47.2%	52.8%	43.7%
	type 4	12.7%	36.8%	44.3%	55.7%	50.1%

to correct minor mistakes or add addendums/updates as the post ages, 4) to add related resources or hyperlinks [6]. For this, we used Chen et al.'s heuristic [6] of analyzing keywords in the edit comments. We measured these characteristics at four points in time: before the first answer (<FA), before the accepted answer (<AA), after the accepted answer (>AA), or after the last answer (>LA). We depict these time points in Figure 2. Finally, we also measured the magnitude of each question edit by computing the edit distance using the Levenshtein distance formula.

Result RQ2: Edits had clearly different characteristics before and after the accepted answer. Table I shows the entire distribution of different edit attributes — overall for all edit actions, as well as the proportion of each edit action in each time point (FA=First Answer, AA=Accepted Answer, LA=Last Answer, "<"=before, ">"=after) of their corresponding edit attribute. Please, see Figure 2 for a visual depiction of these time periods. For example, the value of 39.6% in the second row of the column named "< FA" means that 39.6% of edits on the title happened before the first answer.

The most popular edits overall were: edits on bodies, by other users, and of type1 — to fix typos. We can observe that most of the edited parts were the body of a question which occupies 68.9%. Then, 55.3% of edits are made by the other owners, and type1 — fix typos — are the most popular edits (46.7%) overall. It looks like the most popular edits overall would be focused on fixing problems, which is intuitive — although we found it a bit surprising that other users made more edits overall than owners.

The most popular edits before the accepted answer were: edits on bodies, by owners, and of type2 — to clarify meaning. We also observed that 54.9% of body edits happen before the accepted answer, which is the highest. On the other hand, 51.6% edits by questions owners are before the accepted answer, much higher than the other users. Edits of type 2 are the most common to happen before the accepted answer among all of the four types (54.2%). This shows that edits before the accepted answer are more focused on trying to attract more answers — since the owners are more invested, and they want to clarify meaning.

TABLE II: The comparison between edits by various editors.

	Edit type				Median edit length (Levenshtein dist.)
	type 1	type 2	type 3	type 4	
owner	16.2%	42.1%	24.2%	17.4%	106
other user	56.7%	7.0%	25.1%	11.1%	17

The most popular edits after the last answer were: edits on tags, by other users, and of type4 — to add related resources. We find that tag edits happened more after the last answer (49.4%), other users are more likely to edit after the last answer (48.5%), and type 4 tend to appear more after the last answer (50.1%) This shows a very different trend to the popular edits before the accepted answer. The edits that often happen after the last answer seem to be more focused on housekeeping and documenting the question and answers well.

In terms of edit size: Owners mostly made type 2 changes that were big on average, while other users made mostly type 1 changes that were small on average. Table II shows the comparison between edits made by owners and the other users. We see that question owners made their highest proportion of changes to clarify the meaning (42.1%), and their lowest proportion to correct typos(16.2%). Other users, in turn, were more focused on correcting typos (48.6%).

This finding strengthens our previous observations that owners are probably mostly focused on collecting answers (making bigger changes to clarify), while other users are mostly focused on documenting the process well after the questions were answered (making smaller changes mostly to fix typos).

VI. RQ3: Which question edits are more common in questions with a higher number of answers?

To understand the question edits that developers did when questions that received more answers and received them more quickly, we measured these two variables. We studied the same four attributes as we did in RQ2: who, what part, how much and what type. We also measured them as we did in RQ2.

First, we measure *#Answers*, which is defined as the number of answers that a question obtained after the edit happened, and before the last answer to the question. We also studied *how much* questions were edited by measuring the Levenshtein distance of each edit. We divided questions into clusters, according their edit lengths. We used K-Means clustering [15] because of its simplicity and relatively high accuracy on one dimensional data. We used the elbow method [11] to determine the optimal number of clusters. In this case, we divided questions into 3 clusters, with: smaller, medium and larger edits (see Table III). Finally, we also measured *time* as the time period from the edit action to the accepted answer.

Table III shows our result for RQ3. We report the mean value for *#Answers*, i.e., the number of obtained answers (because the median value was 1 for most attributes), and we report the median value for *time*.

Result RQ3: The edits that were common in the questions that achieved the highest number of answers were not

TABLE III: Mean number of answers and median time to obtain them for different edit attributes.

Attributes			#Answers	Time
part	title		1.59	3,332.0
	body		1.57	3,351.5
	tag		1.51	4,421.0
who	owner		1.65	1,600.0
	other user		1.47	6,739.0
type	type1		1.55	3,131.0
	type2		1.42	7,227.5
	type3		1.55	2,622.5
	type4		1.46	8,136.0
how much	title	smaller edits	1.61	2,728.0
		medium edits	1.56	4,745.5
		larger edits	1.46	8,438.0
	body	smaller edits	1.58	3,226.0
		medium edits	1.40	6,661.5
		larger edits	1.24	9,284.0
	tag	smaller edits	1.54	3,470.0
		medium edits	1.54	5,676.0
		larger edits	1.50	6,030.0

the same as the edits that developers most often made presumably trying to achieve more answers — i.e., before accepted answers.

What part: Questions with edits on titles obtained more answers. The edits with the highest mean number of answers were those on the title (1.59), then on the body (1.57), and then on the tag (1.51). Edits on title also obtained the median quickest answers (3,332 seconds). These differences, although admittedly small, were statistically significant (Mann–Whitney U test, $p < 0.01$).

These findings contrast with the fact that developers mostly change question bodies before obtaining an accepted answer. Our findings suggest that, when applicable, editing question titles may be a better strategy to obtain more answers.

Who: Questions with edits by other users obtained more answers. Edits by other users had the highest mean number of obtained answers (1.65), as opposed to edits by question owners (1.47). In contrast, questions with edits by the owner had a dramatically shorter median time to obtain an answer (1,600s vs. 6,739s). These differences, although sometimes small, were statistically significant (Mann–Whitney U test, $p < 0.01$).

In this case, our findings suggest that, when applicable, developers may want to make different kinds of edits than the common ones before finding an accepted answer — since questions with edits from owners had dramatically shorter times to obtain answers.

What type: Questions with edits to add addendums/updates obtained more answers. From Table III, the questions with edits to add addendums/updates had the highest mean number of answers (1.55), which is the same number as edited questions to fix spelling mistakes (1.55). In terms of median time to obtain an answer, questions with edits to add addendums/updates obtained answers faster (median 2,622.5s) than questions with edits to fix typos (median 3,131s). Again, although these differences are not too substantial, the were statistically significant (Mann–Whitney U test, $p < 0.01$).

Again, our observation for edits in questions with more answers is different from the edits that developers generally perform in their questions before obtaining their accepted answer — since they mostly clarify the question's meaning. Our findings reveal that fixing typos and/or add addendums/updates when it is applicable may be more important to achieve more and faster answers.

How much: Questions with relatively small edits obtained more answers. For each category of *What part* of the question was edited, we cluster its corresponding edits in terms of their size — edit distance. We observed that, regardless of the part that was edited, smaller edits (shorter distance) concentrated the questions that obtained more answers. In terms of the median time to obtain an accepted answer, we observe the same trend — smaller edits averaged shorter times to obtaining an accepted answer.

This finding suggests that making small edits may be more important to obtain more and faster answers.

VII. Conclusions and Future work

We performed an empirical study on the Stack Overflow SO Torrent dataset to understand the edits that users make to SO questions. We found that questions with more edits also had obtained more answers. Then, we also found that developers make very different kinds of edits to questions, depending on whether the question has already received an answer or not. Before receiving an answer, the most popular edits to questions were made by the question owner, on the body, and to clarify meaning, and were not small. After receiving an answer, the most popular edits to questions were made by other users, on tags, and to add related resources. This difference conveys that there may be different intents before and after questions are answered — aiming to obtain more answers vs. aiming to document the question and answer well.

Finally, we also found that the most popular edits in questions with many answers are different than those that people tend to make before receiving answers. The most popular edits in questions with many answers were done by other users, in the title, adding related resources, and they were small.

These findings help developers adjust their expectations in terms of the potential benefit of their edits, as well as letting them decide which edits to make when multiple ones are applicable.

In future work, we intend to extend this work with a human study to understand in more depth the intention of developers when making edits to questions, the context that takes them to perform such edits, and the expectations that they have for them. Another potential avenue for future work would be to build an automated tool to support developers with their edits, taking into consideration the factors that we discovered in this study, as well as other additional factors discovered by interviewing developers. Finally, we provide a replication package for this study [10].

References

[1] M. Ahasanuzzaman, M. Asaduzzaman, C. K. Roy, and K. A. Schneider. Mining duplicate questions in stack overflow. In *Proceedings of the*

13th International Conference on Mining Software Repositories, MSR '16, pages 402–412, New York, NY, USA, 2016. ACM.

[2] L. An, O. Mlouki, F. Khomh, and G. Antoniol. Stack overflow: A code laundering platform? In *2017 IEEE 24th International Conference on Software Analysis, Evolution and Reengineering (SANER)*, pages 283–293, Feb 2017.

[3] S. Baltes, L. Dumani, C. Treude, and S. Diehl. Sotorrent: reconstructing and analyzing the evolution of stack overflow posts. In *Proceedings of the 15th International Conference on Mining Software Repositories*, pages 319–330. ACM, 2018.

[4] S. Baltes, C. Treude, and S. Diehl. Sotorrent: Studying the origin, evolution, and usage of stack overflow code snippets. *arXiv preprint arXiv:1809.02814*, 2018.

[5] C. Chen, X. Chen, J. Sun, Z. Xing, and G. Li. Data-driven proactive policy assurance of post quality in community q&a sites. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):33, 2018.

[6] C. Chen, Z. Xing, and Y. Liu. By the community & for the community: A deep learning approach to assist collaborative editing in q&a sites. *Proc. ACM Hum.-Comput. Interact.*, 1(CSCW):32:1–32:21, Dec. 2017.

[7] D. Correa and A. Sureka. Chaff from the wheat: Characterization and modeling of deleted questions on stack overflow. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, pages 631–642, New York, NY, USA, 2014. ACM.

[8] M. Duijn, A. Kučera, and A. Bacchelli. Quality questions need quality code: Classifying code fragments on stack overflow. In *Proceedings of the 12th Working Conference on Mining Software Repositories*, MSR '15, pages 410–413, Piscataway, NJ, USA, 2015. IEEE Press.

[9] D. Ford, K. Lustig, J. Banks, and C. Parnin. 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*, page 608. ACM, 2018.

[10] X. Jin and F. Servant. What Edits Are Done on The Highly Answered Questions in Stack Overflow? An Empirical Study, Mar. 2019.

[11] T. M. Kodinariya and P. R. Makwana. Review on determining number of cluster in k-means clustering. *International Journal*, 1(6):90–95, 2013.

[12] G. Li, H. Zhu, T. Lu, X. Ding, and N. Gu. Is it good to be like wikipedia?: Exploring the trade-offs of introducing collaborative editing model to q&a sites. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 1080–1091. ACM, 2015.

[13] A. Shrivastava, V. B. Gupta, et al. Methods for the determination of limit of detection and limit of quantitation of the analytical methods. *Chronicles of young scientists*, 2(1):21, 2011.

[14] A. W. Vargo and S. Matsubara. Editing unfit questions in q amp;amp;amp;amp;a. In *2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, pages 107–112, July 2016.

[15] K. Wagstaff, C. Cardie, S. Rogers, S. Schrödl, et al. Constrained k-means clustering with background knowledge. In *Icml*, volume 1, pages 577–584, 2001.

[16] J. Yang, C. Hauff, A. Bozzon, and G.-J. Houben. Asking the right question in collaborative q&a systems. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, HT '14, pages 179–189, New York, NY, USA, 2014. ACM.

[17] W. E. Zhang, Q. Z. Sheng, J. H. Lau, and E. Abebe. Detecting duplicate posts in programming qa communities via latent semantics and association rules. In *Proceedings of the 26th International Conference on World Wide Web*, pages 1221–1229. International World Wide Web Conferences Steering Committee, 2017.