A First Look at Information Highlighting in Stack Overflow Answers

Shahla Shaan Ahmed, Shaowei Wang
University of Manitoba
Winnipeg, Canada
{ahmeds27,shaowei.wang}@umanitoba.ca

Haoxiang Zhang
Huawei
Canada
haoxiang.zhang@acm.org

Tse-Hsun Chen
Concordia University
Montreal, Canada
peterc@encs.concordia.ca

Yuan Tian
Queen's University
Kingston, Canada
y.tian@queensu.ca

Abstract—Technical question and answer (Q&A) sites such as Stack Overflow (SO) have become increasingly important for software developers to share knowledge and contribute to communities. Despite their success and prevalence, navigating the knowledge on these sites remains challenging. To make the posts vivid to users, the Stack Overflow platform allows users to write and edit posts with Markdown or HTML, so that users can leverage various formatting styles (e.g., bold, italic, and code) to highlight the important information.

In this study, we carry out the first large-scale exploratory study on the information highlighting in SO answers. We observe that overall, information highlighting is prevalent on SO, i.e., 47.6% of the answers have information highlighted. More specifically, 38.5%, 11.3%, 7.2% of the answers use Code, Bold, and Italic, respectively. Besides source code related content (e.g., identifiers, and programming keywords), users also frequently highlight updates (e.g., updates of answers), caveats (i.e., a reminder or warn of in which context or condition the provided solution works or does not work), and reference. Our findings provide insights for future research, e.g., future research should consider the highlighted content for the downstream tasks that leverage information from the SO answers.

Index Terms—Q&A websites, information highlighting, Q&A platform

I. Introduction

Technical question and answer (Q&A) sites such as Stack Overflow (SO) have become increasingly important for software developers to share knowledge and contribute to communities. Despite Stack Overflow's success and prevalence, navigating the knowledge on it remains challenging [1]–[4]. The previous study shows that finding answers in long posts remains one of the challenges [2], [3]. 37% of all questions on Stack Overflow have more than one answer, and the average length of an answer is 789 characters [3]. To make the posts vivid to users, Stack Overflow platform allows users to edit their posts with Markdown and HTML [5], [6], so that users can leverage various formatting styles (e.g., **Bold**, *Italic*, and `Code`) to highlight text and direct other users' attention toward the most important information within posts.

Previous studies show the benefits of information highlighting in various domains (e.g., saving the reading time of humans). However, little is known about how information is highlighted on technical Q&A sites (e.g., Stack Overflow). For example, how prevalent is the information highlighted? What content is highlighted using different formatting styles?

Understanding this could provide a landscape of the usage of information highlighting on technical Q&A sites and shed light on the information that is considered important to developers.

In this paper, we performed a first large-scale study on five most commonly used information highlighting types, which are *Bold*, *Italic*, *Code*, *Delete*, and *Heading*, to understand their characteristics and what information is highlighted in the text description of SO answers[1]. We conduct our study on 55,209,643 information highlighting instances among 14,845,929 SO answers. We find that:

- Overall, information highlighting is prevalent on SO, i.e., 47.6% (14,845,929 out of 52,166,061) of the answers use the studied formatting types to highlight information. 38.5%, 11.3%, 7.2% of the answers use *Code*, *Bold*, and *Italic*, respectively, and their highlighted content is short (median length is one word).
- *Code* formatting is mainly used to highlight source code content, such as identifiers (63.5%). *Code* is also used to highlight content other than source code, such as Software (4.9%) and Equation (5.2%).
- *Italic* and *Bold* are frequently used to highlight source code content, as well as content that warns about the context where the provided solution works or does not work, updates on answers, and references to an internal or external resource.

II. Background & Related work

A. Background

Stack Overflow allows users to use Markdown and HTML to write and edit posts [5], [6]. Certain formatting types are used to highlight the information in text description with Markdown and HTML tags. In this study, we focus on five most commonly used formatting types, which are *Bold*, *Italic*, *Delete*, *Code*, and *Heading*. In table I, we present the HTML tags and their equivalent Markdown syntax for each formatting type. We group HTML tags and Markdown formatting based on their rendering effect. For example, we group <b> and <strong> as *Bold* since they render the same effect in browser when users read the posts on Stack Overflow.

---

[1]Note that an SO answer may contain both text and code block, we focus on the information highlighting in text description.

TABLE I: The studied formatting types that are used to highlight information, their corresponding HTML tags, and equivalent Markdown. HTML tags are grouped based on their rendering effects.

Formatting type	HTML tags	Equivalent Markdown
Code	<code>	'example'
Bold	<b>, <strong>	**example**, __example__
Italic	<i>, <em>	*example*
Delete	<del>, <s>	None
Heading	<h1>, <h2>, <h3>, <h4>, <h5>, <h6>	# example, ## example, ### example, #### example, ##### example, ###### example

B. Related work

Several studies have investigated developers' challenges in retrieving knowledge from technical Q&A sites and provided some tools to facilitate the process [1]–[4]. For instance, Sarah and Treude investigated the possibility of using two existing techniques, wordpartten and lexrank, to identify the essential sentences from a long SO answer [3]. Gottipati et al. found that in software forums it is a painstaking process for users to manually search through answers in various threads of posts [4], and they developed a customized search engine to find relevant answers from various software forums. Similarly, Xu et al. developed an technique called AnswerBot, which takes input as the technical question and generates an answer summary for the question [2]. Zhang et al. pointed out an issue of current comment ranking and display mechanism on SO (e.g., a large portion of useful comments are not visible to users) and proposed alternative ranking mechanism to alleviate the issue [1]. Different from prior studies that focus on identifying relevant or important answers from SO, we investigate what and how information is highlighted on SO. Our study could provide insights to enhance existing techniques.

Previous research has explored the benefits of highlighting information in various domains over the last decades [7]–[11]. Wu and Yuan have shown that highlighting could reduce the cognitive load and thus reduce reading time [8]. Jorge et al. investigated the impact of highlighting text in the text classification tasks in machine learning area and found that highlighting is effective in reducing classification effort for human [10]. Similarly, Nguyen et al. developed a tool to explain the output ML models by highlighting portion of text and demonstrate its effectiveness in model explanation [11]. Different from prior studies focus on understanding the benefits of information highlighting, we focus on understanding the practice of information highlighting on SO.

III. RESEARCH QUESTIONS & DATA PREPARATION

A. Research questions

This study aims to understand how information highlighting is used in SO answers and what content is highlighted using different formatting styles. Hence, we formulate our study by answering the following three research questions:

- RQ1: How prevalent is the information highlighting in SO answers?
- RQ2: What types of information are highlighted with *Code* formatting in SO answers?
- RQ3: What types of information are highlighted with *Bold*, *Italic*, *Delete*, and *Heading* formatting in SO answers?

In RQ1, we aim to understand the prevalence of the usage of the studied highlighting formatting and their characteristics. Understanding this can provide practitioners an overview of the landscape of information highlighting in SO answers. We study *Code* formatting and the rest formatting types (i.e., *Bold*, *Italic*, *Heading*, and *Delete*) in separated RQs (RQ2 and RQ3) due to their different nature. We study the information highlighted by different formatting types. By understanding this, we can provide insights into what information is important from users' perspectives and provide insights for future research. For instance, we can provide insight for the downstream research that leverages SO information to facilitate software engineering tasks (e.g., API documentation enrichment [12], [13]). We refer to the formatting types *Bold*, *Italic*, *Heading*, and *Delete* as *Text* formatting for simplicity's sake. We use *Code* and *Code* formatting, *Text* and *Text* formatting exchangeably in the following sections.

B. Data preparation

To answer the three RQs, we downloaded a data dump of Stack Overflow from the Stack Exchange data dump dated March 2021[2]. The data dump contains details information about posts (i.e., questions and answers), as well as their revision history. In this study, we include all the answers and we ended up with 52,166,061 answer posts.

As the dataset was large for processing, we imported it into MySQL. For each answer post, we first extract the textual content from its body and exclude code block(s). Note that we extracted code block(s) using tag "<pre>". Next, we apply the regular expressions defined in Table I to identify information highlighting in each answer post, using the Python library *Regex*. In the end, we extract the content that is highlighted by the studied formatting types. We make our replication package public at https://github.com/shaoweiwang2010/REP_2022_Information_highlight_SO.

IV. RESULTS

A. RQ1: How prevalent is the information highlighting in SO answers?

Approach: To understand the prevalence of information highlighting in SO answers, we compute the percentage of answers that have the studied formatting types (see Table I) and the percentage of words highlighted in each answer with each formatting. In addition, we compute the basic statistics of information highlighting instances to understand the characteristics of each formatting type. More specifically, we

[2]https://archive.org/details/stackexchange

calculate the distribution of instances for each formatting and the number of words highlighted with them.

Results: Overall, 47.6% (14,845,929 out of 52,166,061) of the studied answers have information highlighted. Among all the answers having information highlighted, an average of 10.6% and a median of 7.1% of the text (in words) is highlighted and each answer has an average of 3.9 and a median of 2 highlighting instances.

Code **is used the most frequently among all studied formatting followed by** *Bold* **and** *Italic***.** Table II presents the statistics of different formatting types. We observe that *Code* is the most frequently used type. 38.5% of the studied answers have content highlighted with *Code*. Moreover, 78.9% of the highlighted instances are *Code*. In addition, the studied answers have two *Code* instances on the median, which is larger than other formatting types except *Heading*. This finding is expected since users usually discuss programming problems on SO and it is common to highlight source code in the text. *Bold* and *Italic* are the most frequently used formats besides *Code*. They are used in 11.3% and 7.2% of answers, respectively. *Delete* is rarely used, only 0.07% of the answers use it.

In general, the length of highlighted content is short. The median length of the content highlighted with *Code*, *Bold*, *Italic*, *Deleting*, and *Heading* are 1, 1, 1, 8, and 2 words, respectively. Users tend to highlight single word or phrase using *Code*, *Bold*, and *Italic*. Compared with other formatting types, *Delete* instances are longer.

> Information highlighting is prevalent on SO, i.e., 47.6% of the answers use the studied formatting to highlight information. 38.5% of the answers use *Code*, which is the most frequently used format, followed by *Bold* (11.3%) and *Italic* (7.2%). In general, the length of highlighted content is short.

B. RQ2: What types of information are highlighted with Code formatting in SO answers?

Approach: In this RQ, we aim to understand what content is highlighted with the *Code*. First, we randomly sampled 385 *Code* instances with a 5% interval and 95% confidence level. Since there is no existing terminologies we could reuse for this purpose, we manually performed a lightweight open coding-like process [14], [15] to identify the type of content highlighted with *Code*. The process involves three phases and is performed by the first authors (A1 and A2) of the paper. In phase I, A1 and A2 derived a draft list of types based on 50 *Code* instances. During the phase, the types were revised and refined. In phase II, A1 and A2 independently applied the derived types to the rest 335 samples. They took notes regarding the diffidence or ambiguity during the labeling. During this phase, no new types were introduced. In phase III, A1 and A2 discussed the results from Phase II to resolve any disagreements until a consensus was reached. The coding process has a Cohen's kappa of 0.8 (measured before starting

Phase III), which indicates a substantial level of the inter-rater agreement [16]. Table III presents the definition and the corresponding example of the derived types for *Code*.

Results: *Code* **is mainly used to highlight source code elements, such as identifiers (63.5%), programming language keywords (9.9%), and statements (7.0%).** Table III presents the derived types of content that is highlighted with *Code* and the distribution of each type. *Code* is used the most frequently to highlight identifiers in source code, such as the name of classes, methods, parameters, and variables. We also observe that 59% of the highlighted identifiers appearing in the code block of their corresponding answers. That is said, *Code* is commonly used to highlight identifiers in text for referring to the corresponding identifiers in the code blocks. This is reasonable since users usually need to refer to a unit in the source code when discussing the solution. In addition, 9.9% and 7.0% of the studied instances are for Keyword, and Statement, respectively.

Code **is also used to highlight content other than source code, such as Software (4.9%), Terminology (1.8%), Equation (5.2%), and Version (0.5%).** From Table III, we observe that users also use *Code* to highlight content other than code. For instance, in 5.2% of the cases, *Code* is used to highlight content related to equations. For example, in an answer, answerer discuss the time complexity of binary search "This is O(log n) ", which is highlighted using *Code*[3]. *Code* sometimes is used to emphasize the names of a software, a framework, and a lib. For instance, the answerer of an answer mentioned "You are using Mysql as DB ..." and show a piece of code to demonstrate the database connection between Java and Mysql[4].

> Although *Code* is mainly used to highlighted the content related to source code, it is also used to highlight content other than source code, such as Software (4.9%), Terminology (1.8%), Equation (5.2%), and Version (0.5%).

C. RQ3: What types of information are highlighted with Bold, Italic, Delete, and Heading formatting in SO answers?

Approach: In this RQ, we aim to understand what content is highlighted with the studied *Text* formatting (i.e., *Bold*, *Italic*, *Delete*, and *Heading*). Similar to RQ2, we performed an manual study. We randomly sampled 385 *Text* instances with a 5% interval and 95% confidence level. We ended up with 177 *Bold*, 169 *Italic*, 3 *Delete*, and 36 *Heading* instances. We then identified the types of highlighted content by following the methodology used in RQ2. The coding process has a Cohen's kappa of 0.78 (measured before starting Phase III), which indicates a substantial level of the inter-rater agreement [16]. Table IV presents the definition and the corresponding example of the derived types for *Text* formatting.

Results: Both *Bold* **and** *Italic* **formatting are most frequently used to highlight content related to source code.**

[3]https://stackoverflow.com/questions/17117375
[4]https://stackoverflow.com/questions/24586043

TABLE II: The answer post-wise and highlighted instance-wise statistics of different formatting types.

Formatting type	%Answers	#Instance per answer (mean/median/max)	%Highlighted words per answer (mean/median/max)	%Instances	#Words per instance (mean/median/max)
Bold	11.3%	2.0/1.0/241	9.0%/4.1%/100%	12.0%	2.9/1.0/436
Italic	7.2%	1.9/1.0/354	5.9%/2.3%/100%	7.3%	2.9/1.0/477
Delete	0.07%	1.2/1.0/36	18.1%/10.6%/100%	0.04%	15.7/8.0/701
Code	38.5%	3.7/2.0/490	8.7%/6.3%/100%	78.9%	1.4/1.0/4,669
Heading	1.6%	2.1/2.0/168	9.3%/3.9%/100%	1.7%	3.4/2.0/327

TABLE III: The definition and distribution of types of content highlighted with *Code* formatting.

Type	Definition (example)	Count (%)
Identifier	The identifier in source code, e.g, "ArrayList".	245 (63.6%)
Keyword	The keywords in programming languages, e.g., "for" and "public".	38 (9.9%)
Statement	A statement of source code, e.g., "plot = last.plot()".	27 (7.0%)
Equation	Mathematical equation, operator, or number, e.g., "O(log n)".	21 (5.2%)
Software	The name of softwares, frameworks, tools, and libs, e.g., "Tensorflow" and "MySQL".	19 (4.9%)
Path	Path and files name, e.g., "src/main/java".	17 (4.4%)
Terminology	Terminologies related to programming, e.g., "hash table".	7 (1.8%)
Cmd	Command, e.g., "cloud-init init"	5 (1.3%)
Version	Version information of a software, e.g., "62.1".	2 (0.5%)
Other	Other than the above defined types.	4 (1%)

Table IV presents the distribution of each *Text* type. We observe for certain types, *Bold* and *Italic* share similar patterns. For instance, in 28.8% of the *Bold* instances and 30.8% of the *Italic* instances, content related source code (i.e., Source code) is highlighted. That is said, users also frequently use *Bold* and *Italic* to highlight source code other than using *Code* formatting. Users probably use *Bold*, *Italic*, and *Code* exchangeably, although SO suggests users to tag inline code using *Code* format [5]. Interestingly, we also observe in some cases, the community members changed formatting of the code-related content from *Bold* and *Italic* to *Code*. For instance, in an answer, a user change the formatting for a funciton "strncmp()" from *Italic* to *Code*[5].

Users frequently use both *Bold* and *Italic* to highlight Caveat, Reference, and Terminology. In some cases, it is important to mention the condition or context in which a provided solution works (i.e., Caveat). We notice that such information is usually highlighted in answers. For instance, in a SO answer, the answerer reminded readers "Donot forget add jmtp.dll files (that comes up with jmtp download) as a native library for more info see my answer on ..." using *Bold*.[6]. *Bold* and *Italic* are used to highlight Reference and Terminology, although *Italic* is used more often than *Bold*.

Users tend to highlight the content of types Update and Heading with *Bold*, while *Italic* is more likely to

[5]https://stackoverflow.com/posts/18437465/revisions
[6]https://stackoverflow.com/questions/6498179

be used for the type of Extent. From Table IV, we can see differences between *Bold* and *Italic*. For instance, *Bold* is more frequently used to highlight Update (12.4%) and Heading (15.3%) compare with *Italic*. Interestingly, users also use *Bold* to highlight a heading. One possible reason is that the rendering effect for *Bold* and *Heading* are similar. Prior studies reveal that answers on SO are easy to become obsolete [15], [17]. Therefore, it is not surprising to observe that users highlight the update using *Text* formatting. Different from *Bold*, *Italic* is used more frequently to highlight a word/phrase to express the extent (i.e., Extent).

For *Delete*, it is all used to highlight the obsolete or incorrect content. In terms of *Heading*, we observe that in 91.6% of the instances, users used it to highlight heading, which is expected. However, we also observe in one case for highlighting Source Code and one case for emphasizing Caveat.

> Apart from source code, both *Italic* and *Bold* are used frequently to highlight content in types of Caveat, Reference, Terminology, and Update.

V. DISCUSSION

A. Implication of our findings

Future search should consider the highlighted content for the downstream tasks that leverage information from the SO answers. In RQ3, we observe that users tend to highlight important information for a provided answer using *Text* formatting, such as a reminder or warn of in which context or condition the provided solution works or does not work (i.e., Caveat), an update and new edit in answers (i.e., Update). Such content is informative and important to users when learning and applying knowledge from the answers. Therefore, it is substantial to consider such information for the downstream tasks that leverage information extracted from SO answers, such as answer summarization [18], and API documentation enrichment [12], [13].

VI. THREATS TO VALIDITY

Internal Validity Our study involved qualitative analysis in RQs. To reduce the bias, each instance was labeled by two of the authors and discrepancies were discussed until a consensus was reached. We also showed that the level of inter-rater agreement of the qualitative studies is high.

External Validity One external threat is that it is not clear whether our findings still hold on other Q&A websites. We needed to conduct several qualitative analysis in our RQs;

TABLE IV: The definition and distribution of different types of content highlighted with *Bold, Italic, Heading, Delete*.

Type	Definition (example)	Bold	Italic	Delete	Heading
Update	The update and new edit in the post, e.g., "**Update:** According to your comments, I think you will want to try gzfile() in read.table()".	22 (12.4%)	2 (1.2%)	0	0
Equation	Same as the definition of Equation in Table III.	3 (1.7%)	2 (1.2%)	0	0
Source code	The union of Identifier/Keyword/Statement/Operator/Cmd/Path in Table III, e.g., "I'd suggest the usage of **startupOrder** to configure the startup of route though".	51 (28.8%)	52 (30.8%)	0	1 (2.8%)
Caveat	A reminder or warn of in which context or condition the provided solution works or does not work, e.g., "**It won't work in that shell though, you need to open a new one**".	33 (18.6%)	43 (25.4%)	0	1 (2.8%)
Reference	Reference to internal or external resource, e.g., ".not()".	12 (6.8%)	17 (10.1%)	0	0
Results	Results or output.	3 (1.7%)	2 (1.2%)	0	0
Terminology	Same as the definition of Terminology in Table III.	12 (6.8%)	21 (12.4%)	0	0
Heading	The title of a section, e.g., "**WORKING EXAMPLE**".	27 (15.3%)	2 (1.2%)	0	33 (91.6%)
Options	Options in software, e.g., "you have to enable Less **Secure Sign-In** in your google account".	4 (2.3%)	1 (0.6%)	0	0
Extent	A word/phrase to express extent, e.g., "at least with Hibernate it assumes you want to use a global "hibernate" sequence for *all* tables, which is just stupid."	4 (2.3%)	14 (8.3%)	0	0
Version	Same as the definition of Version in Table III.	0	3 (1.8%)	0	0
Delete	Deleted outdated/wrong description, e.g., "Just use ~~KVO~~ KVC".	0	0	3 (100%)	0
Other	Other than the above defined types.	6 (3.4%)	10 (5.9%)	0	1

however, it is impossible to manually study all instances. To minimize the bias when conducting our qualitative analysis, we took statistically representative random samples of all relevant revisions, in order to ensure a 95% confidence level and 5% confidence interval for our observations.

VII. CONCLUSION

This paper is the first large-scale study of information highlighting in the text description of SO answers. We find that information highlighting is prevalent, with 47.6% of the 52,166,061 studied answers having text highlighted. We propose a terminology to categorize highlighted text in SO answers and find that source code content (e.g., identifiers and programming keywords) are frequently highlighted using *Code* and other highlighting formats (i.e., *Bold* and *Italic*). Besides code, users also tend to highlight updates (e.g., updates of answers), caveats (i.e., a reminder or warning of in which context or condition the provided solution works or does not work), and references.

Further studies could put more effort into investigating how to use the highlighted content for downstream tasks (e.g., answer summarizing) that leverage information from SO answers, and provide tools that can suggest highlighted text for SO users.

REFERENCES

[1] H. Zhang, S. Wang, T.-H. Chen, and A. E. Hassan, "Are comments on stack overflow well organized for easy retrieval by developers?" *ACM Transactions on Software Engineering and Methodology (TOSEM)*, vol. 30, no. 2, pp. 1–31, 2021.

[2] B. Xu, Z. Xing, X. Xia, and D. Lo, "Answerbot: Automated generation of answer summary to developers' technical questions," in *2017 32nd IEEE/ACM International Conference on Automated Software Engineering (ASE)*. IEEE, 2017, pp. 706–716.

[3] S. Nadi and C. Treude, "Essential sentences for navigating stack overflow answers," in *2020 IEEE 27th International Conference on Software Analysis, Evolution and Reengineering (SANER)*. IEEE, 2020, pp. 229–239.

[4] S. Gottipati, D. Lo, and J. Jiang, "Finding relevant answers in software forums," in *2011 26th IEEE/ACM International Conference on Automated Software Engineering (ASE 2011)*. IEEE, 2011, pp. 323–332.

[5] StackExchange, "How do I format my posts using Markdown or HTML?" https://meta.stackexchange.com/help/formatting.

[6] S. Overflow, "Markdown help," https://stackoverflow.com/editing-help.

[7] H. Strobelt, D. Oelke, B. C. Kwon, T. Schreck, and H. Pfister, "Guidelines for effective usage of text highlighting techniques," *IEEE transactions on visualization and computer graphics*, vol. 22, no. 1, pp. 489–498, 2015.

[8] J.-H. Wu and Y. Yuan, "Improving searching and reading performance: the effect of highlighting and text color coding," *Information & Management*, vol. 40, no. 7, pp. 617–637, 2003.

[9] S. Wilson, F. Schaub, R. Ramanath, N. Sadeh, F. Liu, N. A. Smith, and F. Liu, "Crowdsourcing annotations for websites' privacy policies: Can it really work?" in *Proceedings of the 25th International Conference on World Wide Web*, 2016, pp. 133–143.

[10] J. Ramírez, M. Baez, F. Casati, and B. Benatallah, "Understanding the impact of text highlighting in crowdsourcing tasks," in *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, vol. 7, 2019, pp. 144–152.

[11] A. T. Nguyen, B. C. Wallace, and M. Lease, "Combining crowd and expert labels using decision theoretic active learning," in *Third AAAI conference on human computation and crowdsourcing*, 2015.

[12] H. Li, S. Li, J. Sun, Z. Xing, X. Peng, M. Liu, and X. Zhao, "Improving api caveats accessibility by mining api caveats knowledge graph," in *2018 IEEE International Conference on Software Maintenance and Evolution (ICSME)*. IEEE, 2018, pp. 183–193.

[13] C. Treude and M. P. Robillard, "Augmenting api documentation with insights from stack overflow," in *2016 IEEE/ACM 38th International Conference on Software Engineering (ICSE)*. IEEE, 2016, pp. 392–403.

[14] C. B. Seaman, "Qualitative methods in empirical studies of software engineering," *IEEE Transactions on software engineering*, vol. 25, no. 4, pp. 557–572, 1999.

[15] H. Zhang, S. Wang, T.-H. Chen, Y. Zou, and A. E. Hassan, "An empirical study of obsolete answers on stack overflow," *IEEE Transactions on Software Engineering*, vol. 47, no. 4, pp. 850–862, 2019.

[16] A. J. Viera, J. M. Garrett *et al.*, "Understanding interobserver agreement: the kappa statistic," *Fam med*, vol. 37, no. 5, pp. 360–363, 2005.

[17] C. Ragkhitwetsagul, J. Krinke, M. Paixao, G. Bianco, and R. Oliveto, "Toxic code snippets on stack overflow," *IEEE Transactions on Software Engineering*, vol. 47, no. 3, pp. 560–581, 2019.

[18] X. Ren, Z. Xing, X. Xia, G. Li, and J. Sun, "Discovering, explaining and summarizing controversial discussions in community q&a sites," in *2019 34th IEEE/ACM International Conference on Automated Software Engineering (ASE)*. IEEE, 2019, pp. 151–162.