What Do Developers Discuss about Code Comments?

Pooja Rani∗, Mathias Birrer∗, Sebastiano Panichella†, Mohammad Ghafari‡, Oscar Nierstrasz∗

∗Software Composition Group, University of Bern
Bern, Switzerland
scg.unibe.ch/staff
† Zurich University of Applied Science (ZHAW), Switzerland
panc@zhaw.ch
‡ University of Auckland, New Zealand
m.ghafari@auckland.ac.nz

Abstract—Code comments are important for program comprehension, development, and maintenance tasks. Given the varying standards for code comments, and their unstructured or semi-structured nature, developers get easily confused (especially novice developers) about which convention(s) to follow, or what tools to use while writing code documentation. Thus, they post related questions on external online sources to seek better commenting practices. In this paper, we analyze code comment discussions on online sources such as Stack Overflow (SO) and Quora to shed some light on the questions developers ask about commenting practices. We apply Latent Dirichlet Allocation (LDA) to identify emerging topics concerning code comments. Then we manually analyze a statistically significant sample set of posts to derive a taxonomy that provides an overview of the developer questions about commenting practices.

Our results highlight that on SO nearly 40% of the questions mention how to write or process comments in documentation tools and environments, and nearly 20% of the questions are about potential limitations and possibilities of documentation tools to add automatically and consistently more information in comments. On the other hand, on Quora, developer questions focus more on background information (35% of the questions) or asking opinions (16% of the questions) about code comments. We found that (i) not all aspects of comments are covered in coding style guidelines, e.g., how to add a specific type of information, (ii) developers need support in learning the syntax and format conventions to add various types of information in comments, and (iii) developers are interested in various automated strategies for comments such as detection of bad comments, or verify comment style automatically, but lack tool support to do that.

Index Terms—Mining online sources, Stack Overflow, Quora, Code Comment analysis, Software documentation

I. INTRODUCTION

Recent studies provide evidence that developers consider code comments to be the most important type of documentation for understanding code [1]. Code comments are written using natural language sentences, and their syntax is neither imposed by a programming language’s grammar nor checked by its compiler. Consequently, developers follow various conventions in writing code comments [2]. These conventions vary across development environments as developers embed different kinds of information in different environments [3], [4], [5]. This makes it hard to write, evaluate, and maintain the quality of comments (especially for new developers) as the software evolves [6], [7].

To help developers in writing readable, consistent, and maintainable comments, programming language communities, and large organizations, such as Google and Apache Software Foundation provide coding style guidelines that also include comment conventions [8], [9], [10], [11]. However, the availability of multiple syntactic alternatives, the freedom to adopt personalized style guidelines, and the lack of tools for assessing comments, make developers confused about which commenting practice to adopt [6], or how to use a tool to write and verify comments.

To resolve potential confusion, and to learn best commenting practices, developers post questions on various Q&A forums. Stack Overflow (SO) is one of the most popular Q&A forums, enabling developers to ask questions to experts and other developers. Barua et al. determined the relative popularity of a topic across all SO posts and discovered the “coding style” topic as the most popular [12]. Similarly, Quora is another widely adopted by developers to discuss software development aspects [13]. However, what specific problems developers report about code comments such as do they face challenges due to multiple writing conventions or development environments, or which commenting conventions experts recommend to them on these sources, is unknown.

Therefore, we analyze commenting practices discussions on SO and Quora, to shed light on these concerns. Particularly, we formulate the following research questions:

1) RQ1: What high-level topics do developers discuss about code comments? Our interest is to identify high-level concerns and themes developers discuss about code comments on Q&A platforms.

2https://www.stackoverflow.com
3https://www.quora.com
2) **RQ2**: What **type of questions** do developers ask about code comments? Our aim is to identify the type of questions developers frequently ask, e.g., questions such as how to write comments, or what is the problem in their comments. In addition, we aim to identify which platform they prefer to ask which type of questions.

3) **RQ3**: What **information needs** do developers seek about commenting practices? We investigate SO and Quora questions in more detail (including body, tags, comments of the question) to identify the challenges and needs related to writing comments in various development environments.

4) **RQ4**: What **specific commenting conventions are recommended by developers**? We investigate the answers to specific kinds of questions, asking about best practices, to collect commenting conventions suggested by developers.

For each research question, we analyze developer questions at various levels, such as focusing only on the title of the question, the whole question body, or the answers to the question. The rationale behind each level is that future approaches for identifying and automating developers’ intent, needs, and recommendations can focus on specific aspects of comments they want to evaluate and improve. Our manually labeled questions for the research questions RQ2, RQ3, and RQ4 can serve as an initial dataset for building such approaches.

To answer **RQ1**, we use a semi-automated approach involving LDA [14], a well-known topic modeling technique used in software engineering, to explore topics from the SO and Quora posts [12], [15], [16], [17]. We then analyze a statistically significant sample set of posts from SO and Quora posts and derive a taxonomy of **types of questions** and **information needs** of developers regarding comments to answer RQ2 and RQ3 respectively. To answer **RQ4**, we manually analyze the questions asked about best practices on the selected sources and extract various commenting conventions recommended by developers in their answers.

Our results show that developers frequently ask questions on Q&A forums to discuss the best syntactic conventions to write comments, ways to retrieve comments from the code or background information about various comment conventions. Specifically, the questions about how to write or process comments (implementation strategies) in a language, tool, or IDE are frequent on SO. On the other hand, questions about background information concerning various conventions or opinions on the best commenting practices are frequently posted on Quora. Our analysis shows that developers are interested in embedding various kinds of information, such as code examples and media (e.g., images), in their code comments but lack the strategies and standards to write them. We also observe a considerable proportion of questions where developers ask about the ways of automating the commenting workflow with documentation tools or IDE features to foster commenting practices and assess them. This shows the increasing need to improve the state of commenting tools by emphasizing better documentation of the supported features and by providing their seamless integration in the development environments.

The contributions of this paper are:

1) an empirically-validated taxonomy of comment-related concerns collected from multiple sources;
2) a first study to investigate the Quora platform for code comments; and
3) a publicly available dataset including all validated data, and steps to reproduce the study in the Replication Package (RP) [18].

**Paper structure.** In section II we detail the study definition and methodology adopted to answer our research questions. In section III, we present our results and insights. We discuss our findings and their implications in section IV. We recap the threats to validity in section V, summarize related work in section VI, and conclude the paper in section VII.

**II. Study Design**

The goal of this study is to investigate information needs, practices, and problems developers discuss on various online Q&A platforms about code comments. Figure 1 illustrates the steps followed to answer the research questions.
A. Data collection

SO. To identify the relevant discussions concerning commenting practices, we used an approach similar to Aghajani [19]. We selected the initial keywords (Ik) such as comment, convention, and doc to search the SO tags page [19]. The search converged to a set of 70 potentially relevant tags, referred to as initial tags (I).

Two authors independently examined all the tags, their descriptions and the top ten questions in each tag, and selected the relevant tags. We observed that certain tags are ambiguous due to their usage in a different context, such as the comments tag being used in many settings. For example, the tag comment (5710 questions on SO) contains questions about development frameworks (e.g., Django) that provide the feature of attaching comments to a website or other external source (478 questions tagged with “wordpress”), or about the websites where users add comments to the posts (512 questions tagged with “facebook” tag). Therefore, we discarded posts where co-appearing tags were wordpress and facebook, or django-comment. The resulting set of tags (I) is therefore reduced to 55 tags out of 70 tags. The list of selected 55 tags is available in the RP.

We extracted all questions tagged with at least one tag from the I set, resulting in 19705 non-duplicate questions. For each question, we extracted various metadata fields such as ID, title, body, tags, creation date, and view count from SO using the Stack Exchange Data Explorer interface. The interface facilitates the users to query all Stack Exchange sites in an SQL-like Query language.

Quora. Extracting data from Quora was non-trivial due to the lack of publicly available datasets and services to access the data, its restrictive scraping policies [20], and the absence of a public API to access the data. Thus, to extract its data, we implemented a web scraper in Python using selenium to automate browsing, and BeautifulSoup to parse the HTML.

On Quora, the notion of topics is the same as tags on SO, so a question in Quora can be tagged with topics similar to SO tags. Unlike the SO tag page, Quora provides neither an index page listing all its topics, nor a list of similar topics on a topic page. We therefore used the relevant SO tags as initial Quora topics, searched for them on the Quora topic search interface, and obtained 29 topics, such as Code Comments, Source Code, Coding Style. The list of all topics and their mapping to SO tags is provided in the RP.

We scraped all questions with their meta-data such as URL, title, body, and topics of each question from the identified topics, resulting in 3671 questions in total.

B. Analysis Method

Automated analysis from LDA (RQ1). LDA infers latent discussion topics to describe text-based documents. Each document can contain several topics, and each topic can span several documents, thus making it possible for the LDA model to discover ideas and themes in a corpus. We applied LDA on the SO dataset but excluded the Quora dataset as it contains a high number of irrelevant posts (nearly 80%) based on the manually analyzed statistically significant sample set shown in Table I. Additionally, as LDA uses the word frequencies and co-occurrence frequencies across documents to build a topic model of related words, having a high number of irrelevant posts can impact the model quality. Since our objective is to discover the high-level concerns developers have, we extract only titles of the SO questions, as the title summarizes the main concern while the body of the question adds non-relevant information, such as details of development environment, what the developer has tried, or sources already referred to.

To achieve reliable high-level topics from LDA, we performed the following data-preprocessing steps on the question titles: removal of HTML tags, code elements, punctuation and stop words (using the Snowball stop word list), and applied Snowball stemming. We used the data management tool, Makar, to prepare the data for LDA. We provide the concrete steps Makar performed to preprocess the data in the Reproducibility section in the online appendix [18]. The preprocessed title field of the questions served as the input documents for LDA. We used the Topic Modeling Tool [23], a GUI for MALLET [24] that uses a Gibbs sampling algorithm, and facilitates extending the results with meta-data. We provide the input data used for the MALLET tool and the output achieved in the RP.

LDA requires optimal values for the k, α, and β parameters to be chosen, which depends on the type of data under analysis, but this represents an open challenge in software engineering tasks. Wallach et al. pointed out that choosing a smaller k may not separate topics precisely, whereas a larger k does not significantly vary the quality of the generated topics [25]. Therefore, to extract distinct topics that are both broad and high-level, we experimented with several values of k ranging from 5 to 25, as suggested by Linares-Vásquez et al. [26]. We assessed the optimal value of k by analyzing the topic distribution, coherence value (large negative values indicate that words do not co-occur together often), and perplexity score (a low value means the model correctly predicts unseen words) for each value of k from the given range [29]. This process suggested k = 10 as the most promising value for our data (with the lowest perplexity of -6.9 and high coherence score of -662.1) as fewer redundant topics were selected with these values.

References:

In the next iterations, we optimized the hyperparameters $\alpha$ and $\beta$ by using the best average probability of assigning a dominant topic to a question, inspired by the existing studies [30]. We selected the initial values of hyperparameters $\alpha = \frac{50}{\beta}, \beta = 0.01$ using the de facto standard heuristics [31] but allowed these values to be optimized by having some topics be more prominent than others. We ran the model optimizing after every ten iterations in total 1000 iterations. Thus, we concluded that the best hyperparameter configuration for our study is $k = 10, \alpha = 5, \beta = 0.01$. As LDA does not assign meaningful names to the topics, we manually inspected a sample of 15 top-ranked questions under each topic to assign topic names.

<table>
<thead>
<tr>
<th>Source</th>
<th>Extracted posts</th>
<th>Manually analyzed</th>
<th>Relevant posts</th>
</tr>
</thead>
<tbody>
<tr>
<td>SO</td>
<td>19700</td>
<td>644</td>
<td>416</td>
</tr>
<tr>
<td>Quora</td>
<td>3671</td>
<td>565</td>
<td>118</td>
</tr>
</tbody>
</table>

**Taxonomy Study (RQ2, RQ3).** After extracting all posts tagged with the relevant tags for SO, we analyzed a statistically significant sample from all (19700) posts, reaching a confidence level of 99% and an error margin of 5%. The resulting sample set contains 644 posts. We selected the sample posts using a random sampling approach without replacement to reach 644 posts. Similarly, for Quora, we selected 565 posts for our manual analysis, as shown in Table I.

**Classification:** We classified the sample posts into a two-dimensional taxonomy mapping concepts of the selected questions. The first dimension (question types) aims to answer RQ2 while the second dimension (information needs) answers RQ1. The first dimension, inspired from an earlier SO study [32], defines the categories concerning the kind of question, e.g., if a developer is asking how to do something related to comments, what is the problem with their comments, or why comments are written in a particular way. We renamed their categories [32] to fit our context, e.g., their ‘What’ type of question renamed to ‘Implementation problems.’ To classify the selected sample questions in the first dimension categories as shown in Table II, we used closed card sorting technique.

The second dimension outlines more finely-grained categories about the types of information needs developers seek [33], e.g., development environment-related needs (e.g., comments in programming languages, tools, IDEs), or about comments in general. The majority of these categories are built based on the software documentation work by Aghajani et al. [19], and the questions are classified into these categories using the hybrid card sorting technique [34]. In the development environment-related needs, we identified if a question talks about IDE & Editors (e.g., Intellij, Eclipse), Programming languages (Java, Python), or Documentation tools (Javadoc, Doxygen). The further sub-levels of the taxonomy focus on the type of information a questioner is seeking in each development environment, such as asking about the syntax to add a comment or specific information in the comment in Javadoc [33]. For instance, the question “How to reference an indexer member of a class in C# comments” [35] is about the C# language, and asking about the syntax to refer to a member in the class comment, thus gets classified into the three levels as Programming languages—Syntax & format—Class comment according to the taxonomy shown in Figure 6.

**Execution and Validation:** A Ph.D. candidate, a master’s student, and a faculty member, each having more than three years of programming experience, participated in the evaluation of the study. The sample set of questions was divided into an equal subset of questions and selected random questions for each subset to ensure that each evaluator gets a chance to look at all types of questions. We followed a three-iteration-based approach to categorize the questions. In the first iteration, we classified the posts into first and second-dimension categories. In the second iteration, each evaluator (as a reviewer) reviewed the classified questions of other evaluators and marked their agreement or disagreement with the classification. In the third iteration, the evaluator agreed or disagreed with the decision and changes proposed by the reviewers. In case of disagreements, another reviewer who had not yet looked at the classification reviewed the classification and gave his/her decision. Finally, if all evaluators disagreed, we chose the category based on the majority voting mechanism. This way, it was possible to ensure that each classification is reviewed by at least one other evaluator. In case of questions belonging to more than one category, we reviewed the other details of questions, such as tags and comments of the questions and chose the most appropriate one. We finalized the categories and their names based on the majority voting mechanism.

Based on the classification and validation approach described above, all three authors (evaluators) evaluated first the relevance of all their assigned questions, and reviewed the cases of irrelevant questions marked by other evaluators. The third author reviewed and resolved their disagreement cases using a majority voting mechanism (cohen’s $k = 0.80$). As a result, 416 questions of SO and 118 questions of Quora were considered relevant to our study, as shown in Table I. The remaining questions, marked as irrelevant, were manually inspected and no new relevant topic was identified.

**Recommended comment conventions (RQ4).** Given the unstructured or semi-structured nature of comments, and varying standards to write comments, various organizations and language communities present numerous commenting guidelines to support consistency and readability of the comments. For instance, the convention “Use 3rd person (descriptive) not 2nd person in writing comments” is given in the Java Oracle style guide. However, not all of these conventions are recommended by developers in real-time and some conventions are even discouraged, depending on the development environment. Additionally, developers assumed some conventions were feasible e.g., overriding docstrings of a parent class in its subclasses,
but other developers pointed out them as a limitation of current documentation tools or environment. We attempted to collect such comment conventions recommended by developers in their answers on SO and Quora. From the classified questions in RQ2, we chose the questions categorized in the Best Practice category according to Table II. Based on the accepted answers of these questions, we identified recommendation or limitation of various comment conventions. In case a question has no accepted answer, we referred to the top-voted answer.

### III. Results

#### A. High-Level Topics Discussed about Comments (RQ1)

Table III shows the 10 topics generated from the LDA analysis, where the column Relevance denotes if a topic is relevant (R), or irrelevant (IR) and Topic name is the assigned topic label. The column Topic words shows the words generated by LDA, sorted in the order of their likelihood of relevance to the topic.

In the Syntax & Format topic, developers mainly ask about the syntax of adding comments, removing comments, parsing comments, or regex to retrieve comments from code. Occasionally, the questions are about extracting a particular type of information from comments to provide customized information to their clients, such as obtaining descriptions, or to-do comments from code comments. Depending on a programming language or a tool, strategies to add information in the comments vary, such as adding a description in XML comments [SO:9594322], in the R environment [SO:45917501], or in the Ruby environment [SO:37612124]. This confirms the relevance of recent research efforts on identifying the comment information types in various programming languages [3], [4], [36]. IDEs & Editors groups questions about commenting features provided in various IDEs to add or remove comments in the code, or setting up documentation tools to write comments. R Documentation groups questions about documentation features provided in the R language, such as creating various formats of documents including Markdown, Latex, PDF, or HTML documentation. ‘R Markdown’, a documentation format available in knitr package, provides these features in R. In the Code convention topic groups the questions about best practices, such as printing the docstrings of all functions of an imported module, or conventions to check types in Javascript. In this topic, developers also ask the reasons behind various code conventions such as the reason behind having only one class containing the main method in Java, or using particular symbols for comments.

#### Finding 4.1.1 The most relevant topics discussed by developers about commenting practices, and identified by LDA are related to Syntax & format, Documentation generation, IDEs & Editors, R documentation, Code conventions, and Seeking documentation, indicating developers interest in automating their code documentation.

#### B. Types of Questions Discussed (RQ2)

To get insights about which types of questions developers ask about comments and where they ask the specific types of questions more often, we categorized the sampled set of questions from SO and Quora according to the first dimension (question type) shown in Table II. Figure 3 shows such categories on the x-axis with respect to both sources, and the y-axis indicates the percentage of questions belonging to a category out of the total question of a source. The figure highlights implementation strategies (how-to) to be the most frequent category on SO, confirming prior study results [37], [38], [15], [12]. Differently from previous studies, we found best practice and background information questions to arise more frequently than implementation problems questions.

We also observed that different types of questions are prevalent on the investigated platforms, as highlighted by Figure 2. The figure shows that developers ask implementation strategies...
questions and implementation problems questions more on SO compared to Quora. Despite Quora being an opinion-based Q&A site, we also observed questions about best practice and background information about reasons behind various implementation and symbols used in comments. This shows how developers rely on Quora to gather knowledge behind numerous conventions and features provided by the development environment. Such types of questions are also found on SO but to a lesser extent. For instance, we observed the question: *What are these tags @ivar @param and @type in python docstring [SO:379346]* on SO. Based on the thousands of views count of the post, we can say that the question has attracted the attention of many developers. Uddin et al. gathered developers’ perceptions about APIs in SO by mining opinions automatically [39]. Our study provides the evidence to include Quora as another source to validate their approach and mine developer’s opinions.

Finding 4.2.1 Different kinds of questions are prevalent on SO and Quora e.g., implementation strategies and implementation problems questions are more common on SO whereas best practice and background information questions apart from opinion-based questions are more prevalent on Quora. This suggests that Quora can be a useful resource to understand how developers perceive certain development aspects, while SO is useful to understand what technical challenges they face during its development.

C. Developer Information Needs (RQ$_1$)

We analyzed the questions from two different perspectives. The first-dimension categories (question types) (Y-axis in Figure 3) show the types of question (e.g., implement strategies, problems) developers ask, whereas the second-dimension categories (X-axis in Figure 3) highlight the kinds of problems they face with the code comments in the development environment. For example, Figure 3 shows how in the second dimension analysis the most frequent category implementation strategies contains questions about how to do something (comment-related) in a development environment, be it specific to a programming language, a documentation tool, or to the IDE itself. On the other hand, developers discuss the possible features of the documentation tools and IDEs in the limitation & possibilities category. This category highlights the developer’s struggle in locating the feature details from
of writing comments in a programming language or a tool without stating the kind of comment (class/function/block), or concerning the intent of syntax conventions. Such background questions are more often posted on Quora compared to SO.

**Finding 4.2.2** One-third of the developer commenting practices questions are about their specific development environment. The top two frequent questions concern the categories Syntax & format and Asking for feature indicating developers’ interest in improving their comment quality. The rest focus on setting up or using documentation tools in IDEs to generate comments automatically.

Figure 5 shows that developers frequently need to add different kinds of information to the code comments, such as code examples: How to add code description that are not comments? [SO:45510056], performance-related: Javadoc tag for performance considerations [SO:39278635], and media [SO:43556442]. Additionally, developers ask about features to add comments automatically, detect various information from the comments, or synchronize the comments with code changes [SO:23493932]. These questions show the worthiness of devoting research efforts to the direction of identifying information types from comments, detecting inconsistent comments, and assessing and generating comments automatically to improve code comments [3], [40]. We separated the feature-related questions (different features of the tools and IDEs) into two categories, Using feature and Asking for feature, based on the user awareness. In the former category, the user is aware of the existence of a feature in the environment but finds problems in using it, as shown in Listing 1. In the latter category, users inquire about the existence of a feature, or try to locate it, as shown in Listing 2.

**Finding 4.2.3** The implementation strategies category questions are the most frequently viewed questions on SO and limitation & possibilities (is it possible) questions are the second most viewed questions based on the view count of questions. On the other hand, based on the answer count of questions, best practice questions trigger the most discussions along with implementation strategies.
Finding 4.2.4 Developers often ask about the syntax to write function (method) comments compared to other kinds of comments (class, package), showing the trend of increasing effort towards API documentation. Another frequently asked question on SO concerns the conventions to add different kinds of information to code comments, such as code examples, media, or custom tags, indicating developers’ interest in embedding various information in comments.

Finding 4.2.5 Apart from questions related to comment syntax and features, developers ask about adopting commenting styles from other programming languages, modifying comment templates, understanding comments, and processing comments for various purposes.

D. Recommended Comment Convention (RQ4)

There are various syntactic and semantic commenting guidelines mentioned in the style guides, and developers are often confronted with several conventions, or are unable to find any for a specific purpose. We collect various comment conventions recommended by developers in their answers on SO and Quora in Table IV. For example, a developer asks Should .net comments start with a capital letter and end with a period?[SO:2909241], concerning grammar rules in the comments. The accepted answer confirms the recommendation and describes how it helps to improve readability. We, therefore, constructed the recommendation [.net] long inline comments should start with a capital letter and end with a period.[SO:2909241], concerning grammar rules in the comments. The accepted answer confirms the recommendation and describes how it helps to improve readability. We, therefore, constructed the recommendation [.net] long inline comments should start with a capital letter and end with a period.[SO:2909241], concerning grammar rules in the comments. The accepted answer confirms the recommendation and describes how it helps to improve readability. We, therefore, constructed the recommendation [.net] long inline comments should start with a capital letter and end with a period.[SO:2909241], concerning grammar rules in the comments. The accepted answer confirms the recommendation and describes how it helps to improve readability. We, therefore, constructed the recommendation [.net] long inline comments should start with a capital letter and end with a period.[SO:2909241], concerning grammar rules in the comments. The accepted answer confirms the recommendation and describes how it helps to improve readability. We, therefore, constructed the recommendation [.net] long inline comments should start with a capital letter and end with a period.[SO:2909241], concerning grammar rules in the comments. The accepted answer confirms the recommendation and describes how it helps to improve readability. We, therefore, constructed the recommendation [.net] long inline comments should start with a capital letter and end with a period.[SO:2909241], concerning grammar rules in the comments. The accepted answer confirms the recommendation and describes how it helps to improve readability. We, therefore, constructed the recommendation [.net] long inline comments should start with a capital letter and end with a period.[SO:2909241], concerning grammar rules in the comments. The accepted answer confirms the recommendation and describes how it helps to improve readability. We, therefore, constructed the recommendation [.net] long inline comments should start with a capital letter and end with a period.
the comments is another concern highlighted in the study, for example, how to differentiate the variables within a comment [SO:2989522] [SO:47089022], where to put class implementation details (in the class comment or in inline comments) [SO:35957906], and which tag to use for a particular type of information [SO:21823716]. We also found developer concerns regarding grammar rules and word usage in all the sources we analyzed (SO [SO:2909241], and Quora [43]). Although various style guidelines propose comment conventions, there are still many aspects of comments for which either the conventions are not proposed or developers are unable to locate them. Developers commonly ask questions, such as Any coding/commenting standards you use when modifying code? [SO:779025] on SO and Why is there no standard for coding style in GNU R? on Quora. There is therefore a need to cover detailed aspects of comments in the coding style guidelines to help developers write high-quality comments.

On the Impact of Comment Conventions. Various commenting conventions are presented in the style guides to support consistent and readable comments. However, extracting these conventions automatically from style guidelines and customizing them according to the project requirements is still a challenge and not explored much. Additionally, which of these conventions play a more important role for comment comprehension and which ones do not, is not yet explored. Binkley et al. evaluated the impact of identifier conventions on code comprehension, but the conventions were limited to identifiers [44]. Smit et al. identified the relative importance of 71 code conventions, but the majority of the comment conventions were limited to detecting missing documentation comments [45]. Therefore, assessing the impact and importance of comment conventions depending on a specific domain and project, and on various development tasks appears to be another potential direction.

Tools to Assess Comment Quality. Our results show that developers are interested in various automated strategies, such as automatic generation of comments, detection of bad comments, identification of information embedded in comments, and the quality assessment of comments, lack tools that can be integrated into their IDE, especially to verify the comment style automatically [SO:14384136]. However, a limited set of documentation tools support comment quality assessment or adherence of comment to the commenting conventions. For example, current style checker tools, such as Checkstyle, RuboCop and pydocstyle provide support for formatting conventions but lack support for comprehensive checks for grammar rules and content. There is a need to survey current automated style checker tools. Additionally, some languages with advanced style checkers don’t support comment checkers at all, such as OCLint for Objective-C, and Ktlint for Kotlin. We found instances of developers asking about the existence of such tools [SO:8834991] in Figure 5 and Asking tool existence in Table III in the online appendix [18]. Therefore, more tool support is needed to help developers in verifying the high-quality of comments.

V. Threats to Validity

Threats to construct validity concern the relationship between theory and experimentation. In our study, they mainly relate to potential imprecision in our measurements. To mitigate potential bias in the selection of developer discussions on SO, we relied on SO tags to perform initial filtering. However, it is possible that this tag-based filtering approach misses some relevant posts concerning comment convention practices and topics. We therefore investigated the co-appearing tags to find similar relevant tags. Aghajani et al. studied software documentation-related posts, including code comments on SO and other sources [19]. We extracted their documentation-related tags from the given replication package and compared them to our tags (It) to verify if we missed any. On Quora, we mapped the selected SO tags as keywords and searched these
keywords on Quora search interface. To avoid eventual biased in this manual process, we also adopted LDA, to investigate high-level topics emerging in the SO and Quora posts. Thus, a mix of qualitative and quantitative analysis was performed to minimize potential bias in our investigation, providing insights and direction into the automated extraction of relevant topics.

**Threats to internal validity** concern confounding factors, internal to the study, that can affect its results. In our study, they mainly affect the protocol used to build the taxonomy, which could directly or indirectly influence our results. To limit this threat, we used different strategies to avoid any subjectivity in our results. Specifically, all posts were validated by at least two reviewers and, in case of disagreement, a third reviewer participated in the discussion to reach a consensus. Thus, for the definition of the taxonomy, we applied multiple iterations, involving different authors of this work.

**Threats to conclusion validity** concern the relationship between theory and outcome. In our study, they mainly relate to the extent to which the produced taxonomy can be considered exhaustive. To limit this threat, we focused on more than one source of information (SO and Quora), so that the resulting taxonomy has a higher likelihood to be composed of an exhaustive list of elements (i.e., comment convention topics).

**Threats to external validity** concern the generalizability of our findings. These are mainly due to the choice of SO and Quora as the main sources. SO and Quora are widely used for development discussions to date, although specific forums, such as DZone and Reddit could be considered for future works. Moreover, besides all written sources of information, we are aware that there is still a portion of the developer communication taking place about topics that are not traceable. Thus, further studies are needed to verify the generalizability of our findings.

VI. RELATED WORK

Studying developer activities related to various development tasks from various sources can guide the development of tools that help developers to find the desired information more easily. Therefore, researchers have recently focused on leveraging the useful content of these sources i.e., Git, CVS [21], archived communications online forums and CQA (Community Question Answer) sites [38], [12], [30], [16], [13], [19] to comprehend developers information needs. SO and Quora are the most popular platforms that researchers have studied to capture developers questions about trends and technologies [38], security-related issues [16], and documentation issues etc. [19]. Recently researchers have started investigating Quora to get more insight into developer communities [13], e.g., finding and predicting popularity of the topics [46], [13], finding answerability of the questions [47], detecting experts on specific topics [20], [48], [49], or analyzing anonymous answers [50]. Our study is first to investigate this platform for code comments.

Research shows that interesting insights can be obtained from combining these sources [51], [19]. Aghajani et al. studied documentation issues on SO, Github, and mailing lists [19]. They reported a taxonomy of documentation issues developers face. However, they do not focus on the style issues of the code comments. Our study focuses on the all aspects of code comments i.e., the content and style aspect of the code comments. In a previous study, Barua et al. found coding style/practice among the top share on SO [12]. They considered the topic among common English language topics instead of a technical category due to usage of generic words in this topic. As their focus was on technical categories, they did not explore the coding style questions further. Our study complements their work by exploring the specific aspects of coding style, focusing on comment conventions.

VII. CONCLUSIONS

In this study, we investigated commenting practices discussions occurring in SO, and Quora. We first performed automated analysis (LDA) on extracted discussions and then complemented it with a more in-depth manual analysis on the selected sample set. From the manual analysis, we derived a two-dimensional taxonomy. The first dimension of the taxonomy focuses on the question types, while the second dimension focuses on five types of first-level concerns and 20 types of second-level concerns developers express. We qualitatively discussed our insights, and presented implications for developers, researchers and tool designers to satisfy developer information needs regarding commenting practices. We provide the data used in our study, including the validated data and the detailed taxonomy, in the replication package [18]. In the future, we plan (i) to verify the completeness and relevance of gathered comment conventions, (ii) to survey practitioners and tool designers to learn which rules are more important than others, and (iii) to explore ways to improve and assess tool support in this direction. This investigation will help us to see which comment convention affects them most, and during which specific development activity.

ACKNOWLEDGMENTS

We gratefully acknowledge the financial support of the Swiss National Science Foundation for the project “Agile Software Assistance” (SNSF project No. 200020-181973, Feb. 1, 2019 - April 30, 2022).

REFERENCES


