e-mentoring Activities in Online Programming Communities: An Empirical Study on Stack Overflow

Elham Kariri and Carlos Rodríguez

UNSW Sydney, Sydney 2052, NSW, Australia
{e.kariri, carlos.rodriguez}@unsw.edu.au

Abstract. Mentoring is widely acknowledged as an effective method for professional and academic development.The advances in the area of Information Technologies (IT) have positively impacted the mentoring process through a more technology-mediated form of mentoring known as e-mentoring or online mentoring. This form of mentoring has particularly had a great impact in improving learning opportunities in the context of online communities where mentors and mentees from around the world interact with each other in a mutually beneficial collaboration and learning experience. In this paper, we focus on online programming communities and we aim at identifying and understanding e-mentoring activities carried out in this context. We performed a qualitative study on a sample of 400 Q&A threads (i.e., questions and their corresponding answers) from Stack Overflow and identified a total of 31 different activities organized into 10 categories of activities. The results of our study provide insights into the e-mentoring activities performed in Stack Overflow, which can benefit both researchers and practitioners interested in understanding and improving e-mentoring in similar contexts.

Keywords: e-mentoring activities · Stack Overflow · qualitative data analysis.

1 Introduction

Mentoring has been applied as a personal empowerment as well as a developmental tool that addresses the main concerns of mentees through the provision of knowledge and advises that is critical in boosting competency and morale [3]. A key difference between e-mentoring and traditional mentoring is that they apply different approaches to the mentor-mentee relationships and interactions. For instance, in e-mentoring, mentors and mentees interact through online interaction platforms, unlike in traditional mentoring where mentors and mentees are usually located in one physical place [17]. Internet and computers, thus, play a major role in facilitating the online mentoring process.

The advancement in online programming communities has brought together a growing interest in leveraging such communities for mentoring purposes [11]. One example is Stack Overflow[1] (SO), where mentoring happens in the form of questions and answers (Q&A) [7, 14]. Here, users (mentees) can ask questions about their programming concerns, while other expert users (mentors) provide the corresponding answers. Mentoring activities in this context include users sharing code examples, helping each other

[1] https://stackoverflow.com

The final authenticated version is available online at https://doi.org/10.1007/978-3-030-32242-7_10

in code debugging, sharing best practices, among other activities. In order to effectively leverage on online programming communities like SO for the purpose of supporting e-mentoring, it is therefore of utmost importance to identify and understand such activities as well as to identify and categorize the profile of mentees and mentors involved.

Starting from the premise that Q&A is a form of mentoring in itself [7, 14], in this paper, we aim to identify and understand e-mentoring activities happening in SO, as well as to characterize the skills of users that play the roles of mentors and mentees. To this end, we collected the top-voted 400 Q&A threads (i.e., questions and their corresponding answers) during the first quarter of 2018. We employed a qualitative research method based on content analysis described in [19]. The results of our analysis helped us identify a total of 31 different activities (e.g., *adding code examples*), which can be grouped into 10 different categories (e.g., *adding content for enriching answers*) according to the nature of the analyzed activity. The results of our study provide insights into the e-mentoring activities performed in this community, which can be leveraged by both researchers and practitioners interested in understanding and improving e-mentoring in similar contexts.

The rest of this paper is organized as follows. Section 2 presents related work. Section 3 describes the research methodology employed in this study. Next, Section 4 presents our findings. We elaborate on such findings in Section 5, where we also discuss identified opportunities (for both researchers and practitioners) and limitations of this study. We conclude this paper with Section 6 where we also discuss future work.

2 Related Work

Mentoring in online communities can be considered an example of peer-learning [23]. It is a form of knowledge and expertise acquisition where people help each other without necessarily having a formal relation, or received professional teacher training [23], and has been frequently observed in application software development [20].

From the *crowdsourcing* and *service science* perspective, researchers investigated the use of crowdsourcing for education [2, 5], and education and learning as a service [1, 4]. In the former, Anderson [2] explores crowdsourcing for higher education and proposes a design for distributed learning. Bradley et al. [5], instead, explore the use of both open data and crowdsourcing to create the Spectral Game, a game for molecules to interactive spectra matching. In the context of education as a service, Alabbadi [1] proposes education and learning as service where the use of cloud computing is proposed as a cost-effective alternative for sustainable education and learning initiatives. Similarly, Bora and Ahmed [4] propose the use of cloud computing for e-learning and discuss its benefits in terms of costs, software maintenance, security, among other dimensions.

In the domain of *online programming communities*, Feliciano et al. [10] examined the participation and experiences of students in GitHub[2] through a case study. The study found that GitHub can support students in their learning process through peer-reviews, comments and teaching resource suggestions. On a different front, Trainer et al. [24]

[2] https://github.com

explored the impact of social and technical dimensions of software on social ties and technical skills building in mentees. Among the main findings, the study reports that front-end, interdependent projects contribute to the development of technical skills and social ties, while back-end, modular projects contribute mainly to technical skills development.

Storey et al. [22] investigated social and communication channels and their role in shaping and challenging participation in software development. The study shows that, while communication channels do impact the participation culture in software development, not much is known as to how the participatory culture impacts the communication channels used by developers. The study also provides recommendations on how to choose the right communication tools, which considers, among other aspects, channel affordances, timing of tools and importance of learning how to use the selected channel.

In [26], Ye found peer support to be effective in the context of open source software communities through the means of Q&As via mailing lists. Ford et al. [11], instead, explored the deployment of just-in-time mentoring program on SO. In this study, novice participants of the community were redirected to an on-site Help-Room where experienced mentors helped them redraft their questions before submitting them to the Q&A forum. It was found that mentored questions have their score increased 50% on average when compared to non-mentored questions [11]. In a different study, Zagalsky et al. [27] explored R software community to understand how knowledge creation and curation takes place in SO and mailing lists. The study revealed that this is done in a participatory (i.e., through collaborations) and crowdsourced (i.e., working independently) form. The study also shows that there are a number of prolific contributors who were responsible for providing most of the answers.

Differently from the studies above, in this paper, we focus on identifying and understanding the activities carried out during e-mentoring interactions in an online programming community based on Q&As. To the best of our knowledge, this is the first study that explore and identify such activities in this context.

3 Research methodology

This section describes the methodology used in conducting this study. In the following sections, we discuss the dataset used as well as the qualitative research method employed for the analysis.

3.1 Dataset

The dataset used in this study originated from SO. SO offers a Q&A platform that allows users to acquire and grow their programming knowledge and capabilities. The primary dataset was obtained from the threads originating from the questions asked and the corresponding answers provided by users of the platform. In addition, we also utilized the public profiles of users (in SO) to categorize the participants of such threads.

In order to select the Q&As to be included in our study, we collected the top-voted threads of Q&As from the full dataset of SO as of the first quarter of 2018. While this study is qualitative in nature, we used a minimum sample size given by $N =$

$(384.16p)/(p + 383.16)$, where p is the total number of threads in SO (this formula takes into account a margin of error of 5% and confidence interval of 95%) [12]. With $p \approx 16,000,000$, the recommended minimum sample size is $N \approx 384$. We therefore collected a sample of size $N = 400$. For each top-voted question, we collected only the corresponding top-voted answer. This allows us to focus only on high-quality question-answer pairs.

3.2 Qualitative data analysis

We used content analysis techniques described in [19] to identify e-mentoring activities happening in Q&A threads of the selected sample. We focused on identifying the emergent themes (codes) referring to e-mentoring activities from the Q&As threads, which served as the basis for identifying the activity patterns and organizing them into categories. More specifically, in order to categorize an activity as an e-mentoring activity, we considered whether the interaction was performed in the context of an emerging understanding, knowledge application, generalization, testing of ideas or organization [6]. We identified initial activities (codes) that where further refined into categories of activities (see Figure 1). Figure 2 outlines the steps for identifying e-mentoring activities and their corresponding categories. In order to categorize mentors and mentees, we analyzed the public profiles of the users that participated in the sampled threads. Here, we considered users asking questions as mentees, while users providing answers as mentors.

The coding and analysis was performed manually with the help of spreadsheets. The emerging categories for both activity identification and mentor-mentees categorization where reviewed and discussed in group meetings during various iterations to resolve any discrepancies in the meaning and interpretation of the identified categories.

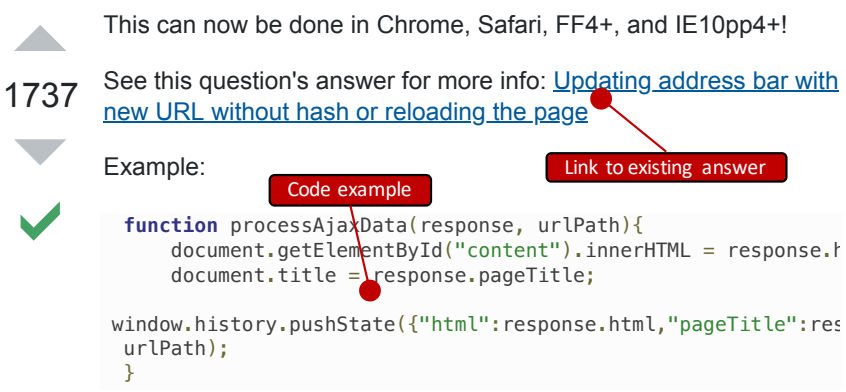

Fig. 1. Example of an answer posted on SO. Tags in red are examples of codes used during the analysis. The figure shows examples of activities such as providing a code example and adding a link to an existing answer.

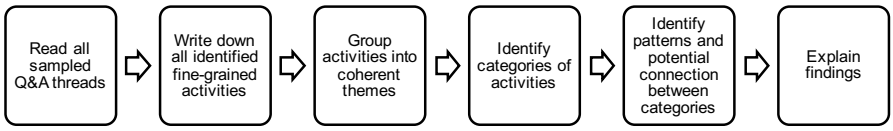

Fig. 2. Steps for the identification of e-mentoring activities and categories.

4 Findings

4.1 Mentors and mentees in Stack Overflow

Figure 3 presents the background information of mentors and mentees that have been identified in the study. Among mentors, we identified that *moderator* represents the largest category (44%). *Full-stack developers* follow with a total of 32%. We noticed that *students* also play the role of mentors, although to a lesser extent (7%). Furthermore, only a handful of *educators / academic researchers* are involved as mentors. The analysis shows that this category represents only 1% of the total number of mentors in SO. Other categories of mentors include *database administrator* (5%), *mobile developer* (7%), *managing director* (1%), *system administrator* (1%), *c-suite executive (CEO, CTO, etc.)* (1%) and *other* (1%).

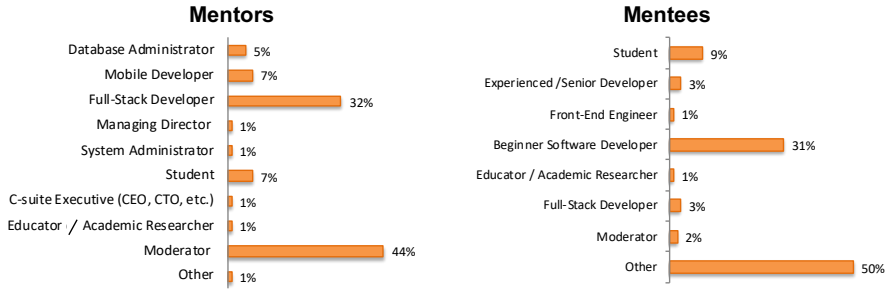

Fig. 3. Categorization of mentors and mentees in SO.

The largest category of mentees falls under *other* (50%). This category consists of users that left SO and the ones that did not provide any relevant information in their public profile that can help us with their categorization. *Beginner software developer* category is the second largest in number (31%). These two categories together make up the majority of the mentees found on this Q&A platform. Furthermore, *students* make up only 9% of all the mentees. It is also worth noting that even *experienced/senior developers* (3%) and *moderators* (2%) participate in the platform as mentees. Other categories of mentees include *front-end engineer* (1%), *educator/academic researcher* (1%) and *full-stack developer* (3%).

4.2 e-Mentoring activities in Stack Overflow

Using the methodology described in the previous section, we identified a total of 31 different activities from the 400 Q&A threads selected from SO. Such activities were organized into 10 categories as shown in Table 1. Below we discuss each of the categories and their corresponding activities.

Providing critiques and feedbacks This category was found to have the highest number of observations, totalling to 490. 81% were made by users (both mentors and mentees) who wanted to vote on an answer, a question or a comment. This category of activities is particularly important in e-mentoring in that it provides an opportunity for mentors and mentees to analyze different answers and comments to their questions, and provide constructive criticism and structured feedbacks. According to study [8], such feedbacks are critical in SO because they enable the community to discuss issues in depth and therefore ensure that the best answers are found.

Motivating and encouraging good community practices This category has the second highest number of observations (394). About 382 of them (97%) were made by users who wanted to provide compliments, offer thanks, opinions and encouraging comments. Motivating and encouraging users of online community to continue posting help enhance e-mentoring in that it promotes the interaction between mentors and mentees through continued communication [18]. For example, the mentor-mentee relationship may grow as a result of positive feedbacks and compliments. To a much lesser extent, 12 of them (3%) were made by users who wanted to put questions on hold in order to encourage improvements or edits on the answers given.

Adding content for enriching answers One of the e-mentoring categories found is *adding content for enriching answers*. This category of e-mentoring activities had a total of 368 observations, making it the third largest category. It includes activities such as *adding code examples, summaries, visual artifacts* such as bug/error image, data frame, state machine diagram, snapshot feature or expected best run output, *adding a link to an existing solution* such as an external application, project module, table, figure or graph, as well as adding a list of *external references*. This category is an important one in that it allows the virtual interaction between mentors and mentees by providing an opportunity for peer mentors to add content and enrich answers which have already been provided by other mentors. Improving answers in this way can help, in turn, improve the quality of the knowledge acquired from the platform [3].

Improving the quality of questions and answers Another category of activities with a high number of observations is that of improving the quality of questions and answers. This category had a total of 310 observations, of which 68% were edited by either rewording it or improving the vocabulary used in the post. The rest of the observations (32%) targeted at correcting typos and code formatting. All these can be considered

#	e-Mentoring Activities	Total number of Observations
1	**Providing critiques and feedbacks**	**490**
	– Voting on a question, answer or comment (400) – Marking a question, answer or comment as a favourite (77) – Acknowledging/Citing useful answers for future readers (13)	
2	**Motivating and encouraging good community practices**	**394**
	– Providing a compliment (e.g., thanks) (382) – Marking a question on hold to encourage editing (12)	
3	**Adding content for enriching answers**	**368**
	Adding: – Code examples (280) – Summaries (e.g., a brief before a long explanation) (14) – Visual artifacts, such as bug/error image, data frame, state machine diagram, snapshot feature or expected output (24) – Links to an existing answer (e.g., external application, project module, table, figure or graph) (41) – External references (9)	
4	**Improving quality of questions and answers**	**310**
	– Editing a question or answer (e.g., to reword a sentence) (212) – Correcting spelling, grammar error, code formatting (98)	
5	**Helping in code debugging**	**210**
	– Providing a correct code or solution (97) – Suggesting a different method to correct a bug (96) – Providing a possible solution for a bug (17)	
6	**Explaining and clarifying answers**	**186**
	– Clarifying a concept meaning (42) – Providing an explanatory tutorial (95) – Using a video/demo to explain an implementation (14) – Providing code snippets with explanations (35)	
7	**Sharing best development practices**	**179**
	Sharing: – Documentation guides about a platform (46) – Developer blogs, articles, tutorials or projects (121) – Links to academic publications such as e-book (2) – Reference documentations (10)	
8	**Managing posted information**	**177**
	– Marking a question as duplicate (13) – Closing a question (21) – Creating a new tag or editing an existing one (143)	
9	**Helping in organizing meetings and communications**	**112**
	– Providing contact information (e.g., e-mail address) (2) – Collaborating in discussions (107) – Suggesting a new communication channel (e.g., chats) (3)	
10	**Offering suggestions for improvement**	**55**
	– Providing advice or tip (e.g., API versions) (14) – Sharing other posted answers (41)	

Table 1. Categories of e-mentoring activities in SO. The number shown in parenthesis for each activity is the number of threads on which the activity was performed. The last column reports on the number of observations of activities under each category. Notice that activities are non-exclusive, which means that a thread can fall into more than one type of activity. The number of observations for a category can be therefore higher than the total number of threads. For example, category *providing critiques and feedbacks* has 490 observations, which is computed as the sum of the observations for each activity under this category.

to be e-mentoring activities in that they are geared toward ensuring effective communication between mentors and mentees, thereby ensuring successful mentoring in the absence of face-to-face interaction [3].

Helping in code debugging The results of our analysis shows, furthermore, that 210 observations are about *helping in code debugging*. This is a category of activities that involves identifying and understanding potential coding errors in programs and then providing a solution to fix them. The corresponding activities contribute to e-mentoring in SO in that they allow mentors to provide insights regarding how to debug and fix code errors, which is considered a key task in software development.

Explaining and clarifying answers A total of 186 observations aimed at *explaining and clarifying answers*. Activities under this category help enhance the communication between mentors and mentees in SO. It can be viewed as a practice through which the answers of mentors to questions posted by mentees are further elaborated to favor clarity. For example, explaining and clarifying answers ensures that the essential meaning of the answers provided by the mentors is understood by the mentees (and the whole community), thereby mitigating the risks of potential misunderstandings.

Sharing best development practices This category has a total of 179 observations. Most of the observations within this category involves sharing a blog or article with best practices. Some of the activities under this category included providing guide documentation regarding a platform, link to academic publications (e.g., e-book) and reference documentations. These are considered e-mentoring activities in that they enable mentors to remotely provide useful learning materials to mentees. Sharing best practices within a domain is important in e-mentoring because it provides both mentors and mentees with an opportunity to learn from well-established practices and techniques in a particular domain.

Managing posted information A total of 177 observations were found under this category. 80% of them involved creating a new tag for a question. The main activities under this category included marking questions that were found to be duplicate, closing questions that have already been answered, creating new tags for a question and editing an existing question. These activities help enhance the quality of e-mentoring in SO because it contributes to the curation of posts (and, therefore, knowledge) shared between mentors and mentees, as well as teach how to properly make contributions to the community. The latter is particularly important in SO, which rely on a code of conduct for participating in the community[3].

Helping in organizing meetings and communications Another important category of e-mentoring activities in SO consists in *helping in organizing meetings and communications*, which was observed 112 times. Activities under this category include

[3] https://stackoverflow.com/conduct

collaborative discussion about positive experiences and opinions, alternative communication channels and sharing of contact information. All these activities contribute to e-mentoring in that they help establish proper and effective communication channels between mentors and mentees [6].

Offering suggestions for improvement The activities under this category involve providing suggestions on how the answers and solutions provided by mentors to the questions from mentees could be further improved, e.g., through tips, advices and previously posted answers to similar questions. This is an essential category of activities for the enhancement of the effectiveness of e-mentoring because it helps in ensuring that the provided answers and solutions are as accurate as possible. Overall, 55 observations were made where *offering suggestions for improvement* was the main aim of the activities.

5 Discussion

To the best of our knowledge, this is the first study aims at identifying mentoring-related activities in SO as well as characterizing the profile of users that play the roles of mentors and mentees. In the following, we discuss and reflect upon the activities identified in the previous section, the opportunities for researchers and practitioners interested in these communities and platforms, and the limitations of this work.

5.1 e-Mentoring activity areas in Stack Overflow

The findings discussed in Section 4 allowed us to identify three key e-mentoring activity areas in SO, namely, *knowledge creation*, *knowledge curation*, and *mentor-mentee communication*, *organization* and *encouragement*. In the following, we discuss each of these areas along with the implications for e-mentoring.

Knowledge creation This area encompasses e-mentoring activities that are geared toward creating knowledge. While answering question is the main mechanism for creating knowledge in SO [27], other means also exist in the platform for complementing such knowledge creation instrument. One such alternative is *adding content for enriching answers*, which allows peer mentors to complement each others' contribution toward richer answers to mentees' questions. Artifacts used for enriching answers include code examples, external references (e.g., blogs) and visual artifacts such as diagrams and screenshots. Another form of knowledge creation includes *helping in code debugging*. This form of knowledge creation is not limited to just providing a solution for a code bug, but typically also involves detailed discussions on the underlying causes as well as key programming concepts related to a bug. Next, *sharing best development practices* can also be considered a mechanism for knowledge creation. This activity involves a range of artifacts including user and reference documentation, academic publications and tutorials that specifically target recommendations and best practices in the context of programming.

e-Mentoring activity areas in Stack Overflow		Knowledge creation			Knowledge curation					Mentor-mentee communication, organization & encouragement		# observations by categories of mentors and mentees	% participation by categories of mentors and mentees
Categories of mentors and mentees \ Categories of e-mentoring activities		Adding content for enriching answer	Helping in code debugging	Sharing best development practices	Providing critiques and feedbacks	Improving quality of questions and answers	Explaining and clarifying answers	Managing posted information	Offering suggestions for improvement	Motivating and encouraging good community practices	Helping in organizing meeting and communications		
Mentors	Database Administrator	1	7	25		10	19	6	1			69	5%
	Mobile Developer	28	6	2	5	18	8	6	14			87	7%
	Full-Stack Developer	221	69	43	2	28	42	7	13		2	427	32%
	Managing Director	5	1	5			1					12	1%
	System Administrator	1	5	3	1							10	1%
	Student	17	43	15			11	2	3			91	7%
	C-suite Executive	3			2	2		1	1			9	1%
	Educator or Academic Researcher	18										18	1%
	Moderator	71	79	86	4	125	50	85	11	12	52	575	44%
	Other				3	2		6	2			13	1%
Mentees	Student				44	19	22	4	5		7	101	9%
	Experienced/Senior Developer				19	2				5	8	34	3%
	Front-End Engineer				2		7					9	1%
	Full-Stack Developer					3		14		21		38	3%
	Beginner Software Developer				167	52	2	2	1	139		363	31%
	Educator or Academic Researcher				4					7		11	1%
	Moderator	3			8	3	9	4		1		28	2%
	Other				229	46	15	40	4	209	43	586	50%
# observations (by category of activity)		368	210	179	490	310	186	177	55	394	112	2481	

Table 2. Activities performed by mentors/mentees (represented as a table heatmap). The number in each cell reports the number of times a given mentor/mentee category performed the corresponding activity.

Knowledge creation in SO is mostly started by mentors (see Table 2). More specifically, it is widely performed by *moderators*, *full stack developers*, *students*, *mobile developers*, and *database administrators*. For example, we noticed that 60% of *adding content for enriching answers* activity is done by *full stack developers*, 19% by *moderators* and 21% by the remaining categories of mentors. In addition, we found that 38% of *helping in code debugging* activities is done by *moderators*, 33% by *full stack developers* and 21% by *student*. Moreover, sharing *best development practices* activities are done by 48% of *moderators*, 25% by *full stack developers* and 14% of the activities by *database administrators* from mentor categorization. Our analysis shows that the two most active categories of mentors in this area and across the three corresponding activity categories are therefore *full-stack developer* and *moderator*.

Knowledge curation The crowdsourcing, public nature of SO makes knowledge creation in this platform open to its community members. While such characteristic is typically seen as a key to leverage the power of the crowd [16], it comes with a number of challenges from a curation and quality control perspective [9]. We review next the activities that target knowledge curation of SO contributions made by the community.

One of the main activities for knowledge curation is *providing critiques and feedback*. It includes activities such as voting on questions, answers and comments, marking questions as favourite, and citing existing answers. Such feedback mechanisms help knowledge curation in the context of e-mentoring mainly in three ways. Firstly, it helps mentors and mentees assess the quality of their own posts (i.e., on their questions and answers). Secondly, it facilitates finding good questions and answers through the statis-

tics associated to each post (e.g., number of votes and favourites). Thirdly, it allows for finding mentors (experts) in specific areas through SO's reputation system[4], which is based on the feedback mechanisms above.

Other activities in this area involve a more direct manipulation of the content of questions and answers for curation purposes. For example, the category *improving quality of questions and answer* involves activities such as question and answer editing (e.g., for understandability and technical correctness purposes) and spelling and grammar correction. These activities can be especially useful, e.g., for both mentees that are very new to a topic (e.g., for ensuring the use of the right terminology in posts) and non-native English speakers. Furthermore, *explaining and clarifying answers, managing posted information*, and *offering suggestions for improvements* are also categories that carry out knowledge curation activities. These categories involve a variety of curation activities including clarifying and explaining programming concepts (e.g., encapsulation in object-oriented programming), closing a question (e.g., out of scope question) and providing advice and tips (e.g., API versions).

While in the previous section we discussed that knowledge creation is mainly initiated by mentors, we can see that knowledge curation is carried out by both mentors and mentees in a more evenly distributed manner as compared to knowledge creation (see Table 2). If we look at each activity category individually, we can notice that the category *providing critiques and feedbacks* is dominated by mentees, while the remaining categories is performed mainly by mentors. These results are in line with the observation that mentees in SO are actively involved in voting and favoriting questions and answers they find useful. The categories of *improving quality of questions and answers, explaining and clarifying answers, managing posted information* and *offering suggestions for improvement* typically require expertise and therefore are mainly performed by mentors.

The activities discussed in this section have tangible implications for the e-mentoring experience. For example, having a question edited for improving its understandability allows the original poster of the question to *learn by example* [13] on how to properly write questions both for a topic and within SO community. Similarly, questions marked as duplicate not only allows for discovering an already existing answer to the question but it is also an opportunity to remind mentees about the good practices for participating in the community (e.g., searching and researching before posting a question[5]). Finally, the gamification mechanisms put in place by SO offer gratification and acknowledgement for both mentors and mentees, which helps them keep engaged in the community [27].

Mentor-mentee communication, organization and encouragement The third area emerged from e-mentoring activities in SO involves tasks that can be broadly classified as community management. It encompasses administrative, organizational and procedural aspects of online mentoring such as guiding discussions, participation encouragement, helping in finding consensus and promoting community best practices and policies. More specifically, the category *motivating and encouraging good community prac-*

[4] https://stackoverflow.com/help/whats-reputation

[5] https://stackoverflow.com/help/how-to-ask

tices involves activities such as the provision of compliments (e.g., thanks) and marking questions to encourage editing and improvement toward best practices. The implications of this category are two-fold. On the one hand, it serves as a incentive mechanism to promote participation of mentors and mentees through both extrinsic (e.g., increased reputation score in the platform) and intrinsic motivation (e.g., shared purpose) [9]. On the other hand, it helps encourage the improvement of quality of knowledge produced by the community through iterative enhancement [9] of both questions and answers.

From an organization and communication perspective, the category *helping in organizing meetings and communication* involves activities such as sharing contact information, collaboration in discussions and suggestions of alternative communication and collaboration channels. While these activities may not represent a direct form of mentor-mentee knowledge transfer, they do play an important role in the e-mentoring process. For example, in the context of social computing, study [25] discusses the interplay between SO and Github where, e.g., Github committers provide more answers and ask less questions, which suggests a cross-fertilization between different platforms and channels that can contribute to expertise sharing in an e-mentoring context.

Mentor-mentee communication, organization and encouragement is widely performed by *beginner software developers*, *full stack developers*, and *others* in the mentee category (see Table 2). For example, we noticed that the majority of *motivating and encouraging good community practices* activities are done by *others* (53%) and *beginner software developers* (35%) in the mentee category. These results suggest that peer mentees tend to help each other in following good community practices. Under mentor category, the results shows that *moderator* is the most involved category in this area, mainly, in helping in organization and communication matters.

5.2 Opportunities for researchers and practitioners

Many opportunities are envisioned in the context of understanding and improving e-mentoring in online programming communities. On the research side, one potential direction that can emerge from our findings consists in further exploring how each of the identified activities contribute to e-mentoring in SO as well as in other similar online programming communities. For example, one interesting research question in this line is the understanding of how code examples benefit the e-mentoring process in these communities and whether the mechanisms currently in place are appropriate from an e-mentoring perspective. Given the typically short-lived interactions that happen between mentors and mentees in SO while performing these activities, it is also worth exploring the implications of such type of interactions and contrast them with traditional mentoring, which usually involves longer-lasting mentoring activities and mentor-mentee relationships. In addition, given that our findings on the characterization of mentors and mentees are based solely on information provided in public user profiles, more studies are needed (e.g., interviews) in order to better identify and characterize the actual profile of the users involved in e-mentoring activities.

Practitioners, on the other hand, can benefit from the reported findings by identifying opportunities for the development of new features for the platform to support common e-mentoring activities. For example, additional or improved features can be added to support widely performed activities such *improving the quality of questions*

and answers, which can go beyond human-based curation and incorporate also automated techniques, e.g., based on NLP and AI [21, 15]. From the perspective of the category *adding content for enriching answers*, platforms such as SO could also rethink the way in which resources are managed within the platform. For instance, new features can be proposed that are able to categorize learning resources into bundles that target specific topics of interest for mentors and mentees participating in the community. Finally, more and better integration with development and collaboration tools (such as GitLab[6] and Slack[7]) can be added to the platform in order to provide an environment that allows both mentors and mentees to seamlessly switch back and forth between the tools used during their e-mentoring activities.

5.3 Limitations of this study

The study reported in this paper comes with its own limitations. Firstly, the dataset used in our analysis was limited to top-voted threads, and within each thread, to top-voted answers only. This decision, while helpful as a heuristic for choosing good questions and answers, brings together the risk of drawing conclusions that apply only to the dataset sampled for our study. Secondly, the characterization of mentors and mentees rely solely on the self-reported profile of the users participating in the selected threads. The implications of this is that our categorization may not capture the true profile of user, which may differ from the self-reported one for reasons such as lack of profile updates or simply because no profile information is provided by the user. Finally, this study is exploratory in nature and focuses on SO only. While this is a representative online programming community, the findings reported in this paper may not apply to other communities where e-mentoring may take place such as Github and Apache[8].

6 Conclusion and future work

This work explored e-mentoring activities in online programming communities through an empirical study on Stack Overflow's Q&As. The analysis of 400 threads of top-voted Q&As collected from this platform allowed us to identify a total of 31 different activities grouped into 10 categories, which create impact in three different areas, namely: knowledge creation, knowledge curation, and mentor-mentee communication, organization and encouragement. Our analysis found that, while *knowledge creation* activities are mainly performed by mentors, *knowledge curation* develops in a more participatory manner where both mentors and mentees jointly collaborate in curating knowledge in Q&As threads. Furthermore, *mentor-mentee communication, organization and encouragement* was found to be mostly initiated by peer mentees in an effort to encourage good community practices, organization and communication. We believe the results of this study will help in understanding better e-mentoring activities in this domain and motivate further research in this direction.

[6] https://gitlab.com
[7] https://slack.com
[8] https://www.apache.org

In future work, we plan to extend this research with further studies involving interviews with Stack Overflow's mentors and mentees with the aim of understanding the underlying motivations for participating in this form of e-mentoring, actual benefits of such participation, and opportunities for improving e-mentoring in online programming communities.

Acknowledgment

We would like to thank Prof. Boualem Benatallah, Dr. Ho-Pun Lam and reviewers for their valuable comments and feedbacks, which helped to improve this work.

References

1. Alabbadi, M.M.: Cloud computing for education and learning: Education and learning as a service (ELaaS). In: 2011 14th International Conference on Interactive Collaborative Learning. pp. 589–594. IEEE, Piestany, Slovakia (2011)
2. Anderson, M.: Crowdsourcing higher education: A design proposal for distributed learning. MERLOT Journal of Online Learning and Teaching **7**(4), 576–590 (2011)
3. Bierema, L.L., Merriam, S.B.: E-mentoring: Using Computer Mediated Communication to Enhance the Mentoring Process. Innovative Higher Education **26**(3), 211–227 (2002)
4. Bora, U.J., Ahmed, M.: E-Learning using Cloud Computing. International Journal of Science and Modern Engineering **1**(2), 9–12 (2013)
5. Bradley, J.C., Lancashire, R.J., Lang, A.S., Williams, A.J.: The Spectral Game: leveraging Open Data and crowdsourcing for education. Journal of Cheminformatics **1**(1), 9 (2009)
6. Churchill, D., King, M., Webster, B., Fox, B.: Integrating learning design, interactivity, and technology. In: Carter, H., Gosper, M., Hedberg, J. (eds.) Proceedings of 30th ASCILITE-Australian Society for Computers in Learning in Tertiary Education Annual Conference. pp. 139–143. Macquarie University, Sydney, Australia (2016)
7. Cuban, L.J.: How Teachers Taught: Constancy and Change in American Classrooms, 1890-1990. Longman (1984)
8. Dalip, D.H., Gonçalves, M.A., Cristo, M., Calado, P.: Exploiting User Feedback to Learn to Rank Answers in Q&A Forums: A Case Study with Stack Overflow. In: Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 543–552. SIGIR '13, ACM, Dublin, Ireland (2013)
9. Daniel, F., Kucherbaev, P., Cappiello, C., Benatallah, B., Allahbakhsh, M.: Quality Control in Crowdsourcing: A Survey of Quality Attributes, Assessment Techniques, and Assurance Actions. ACM Computing Surveys **51**(1), 7:1–7:40 (2018)
10. Feliciano, J., Storey, M., Zagalsky, A.: Student Experiences Using GitHub in Software Engineering Courses: A Case Study. In: 2016 IEEE/ACM 38th International Conference on Software Engineering Companion (ICSE-C). pp. 422–431. IEEE, Austin, TX, USA (2016)
11. Ford, D., Lustig, K., Banks, J., Parnin, C.: "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. pp. 608:1–608:12. CHI '18, ACM, Montreal QC, Canada (2018)
12. Garcia, C., Jha, G., Verma, R.: The Ultimate Guide to Effective Data Collection. Online: https://europa.eu/capacity4dev/ict4dev/discussions/free-ebook-ultimate-guide-effective-data-collection
13. van Gog, T., Rummel, N.: Example-Based Learning: Integrating Cognitive and Social-Cognitive Research Perspectives. Educational Psychology Review **22**(2), 155–174 (2010)

14. Hargreaves, A., Fullan, M.: Mentoring in the New Millennium. Theory Into Practice **39**(1), 50–56 (2000)
15. Hirschberg, J., Manning, C.D.: Advances in natural language processing. Science **349**(6245), 261–266 (2015)
16. Howe, J.: The Rise of Crowdsourcing. Wired **14**(6), 1–4 (2006)
17. Packard, B.W.L.: Web-based Mentoring: Challenging traditional models to increase women's access. Mentoring and Tutoring: Partnership in Learning **11**(1), 53–65 (2003)
18. Ren, Y., Kraut, R.E., Resnick, P.: Encouraging commitment in online communities. In: Building Successful Online Communities: Evidence-Based Social Design, pp. 77–124. The MIT Press (2012)
19. Renner, M., Taylor-Powell, E.: Analyzing qualitative data. University of Wisconsin - Extension, Program Development and Evaluation (2003), https://learningstore.uwex.edu/assets/pdfs/g3658-12.pdf
20. Schilling, A., Laumer, S., Weitzel, T.: Train and Retain: The Impact of Mentoring on the Retention of FLOSS Developers. In: Proceedings of the 50th Annual Conference on Computers and People Research. pp. 79–84. SIGMIS-CPR '12, ACM, Milwaukee, WI, USA (2012)
21. Socher, R., Bengio, Y., Manning, C.D.: Deep Learning for NLP (without Magic). In: Strube, M. (ed.) Tutorial Abstracts of ACL 2012. p. 5. ACL 2012, Association for Computational Linguistics, Jeju Island, Korea (Jul 2012)
22. Storey, M., Zagalsky, A., Filho, F.F., Singer, L., German, D.M.: How Social and Communication Channels Shape and Challenge a Participatory Culture in Software Development. IEEE Transactions on Software Engineering **43**(2), 185–204 (2017)
23. Topping, K.J.: Trends in Peer Learning. Educational Psychology **25**(6), 631–645 (2005)
24. Trainer, E.H., Kalyanasundaram, A., Herbsleb, J.D.: E-Mentoring for Software Engineering: a Socio-Technical Perspective. In: 2017 IEEE/ACM 39th International Conference on Software Engineering: Software Engineering Education and Training Track (ICSE-SEET). pp. 107–116. IEEE, Buenos Aires, Argentina (2017)
25. Vasilescu, B., Filkov, V., Serebrenik, A.: StackOverflow and GitHub: Associations between Software Development and Crowdsourced Knowledge. In: 2013 International Conference on Social Computing. pp. 188–195. IEEE, Alexandria, VA, USA (2013)
26. Ye, Y.: Peer to peer support for the reuse of Open Source Software libraries. In: 2009 IEEE International Conference on Information Reuse Integration. pp. 284–289. IEEE, Las Vegas, NV, USA (2009)
27. Zagalsky, A., German, D.M., Storey, M.A., Teshima, C.G., Poo-Caamaño, G.: How the R community creates and curates knowledge: an extended study of stack overflow and mailing lists. Empirical Software Engineering **23**(2), 953–986 (2018)