Answers or no answers : studying question answerability in stack overflow

Chua, Alton Yeow Kuan; Banerjee, Snehasish

2015

Chua, A. Y. K., & Banerjee, S. (2015). Answers or no answers : studying question
answerability in stack overflow. *Journal of Information Science*, 41(5), 720-731.
doi:10.1177/0165551515590096

<https://hdl.handle.net/10356/144779>

<https://doi.org/10.1177/0165551515590096>

© 2015 The Author(s). All rights reserved. This paper was published in *Journal of
Information Science* and is made available with permission of The Author(s).

Downloaded on 26 Apr 2023 19:18:29 SGT

Answers or no answers: studying question answerability Stack Overflow

Alton Y. K. Chua Snehasish Banerjee

Abstract

Several questions posted in community question answering sites (CQAs) often fail to attract even a single answer. To address the growing volumes of unanswered questions in CQAs, the objective of this paper is two-fold. First, it aims to develop a conceptual framework known as the Quest-for-Answer to explain why some questions in CQAs could draw answers while others remain ignored. The framework suggests that the likelihood to attract answers depends on both the metadata and the content of questions. Second, the paper attempts to empirically validate the Quest-for-Answer framework through a case study of Stack Overflow. A total of 3,000 questions divided equally between those answered and unanswered were used for analysis. The Quest-for-Answer framework yielded generally promising results in predicting if questions are likely to attract answers.

Keywords

Community question answering; Stack Overflow; unanswered questions

Introduction

With knowledge-sharing platforms such as community question answering sites (CQAs), users are able to harness the collective wisdom of the online community through asking and answering questions. A typical question-answering cycle in CQAs commences when a user asks a question, and awaits answers from others. The question is said to have been resolved once the asker chooses a best answer from among the responses.

However, not every question asked in CQAs is always resolved. In fact, a significant proportion of questions fail to attract even a single answer. For example, about one-fifth of unresolved questions asked in popular CQAs such as Yahoo! Answers do not receive responses while some 43% of unresolved questions remain unanswered in Baidu Knows, one of the popular Chinese CQAs [1]. If a CQA consistently fails to return answers, it is unlikely to encourage future participation from askers. Eventually, as asking and answering activities wane, its *raison d'être* would be called into question.

The longevity of any CQA clearly depends on the goodwill of users to respond to questions posted. Recognizing the importance of voluntary contributions, some studies have examined motivation factors of users in online communities [2, 3]. Users were found to participate based on the expectation that their contribution would be worthy of the effort to create new value in the community [4, 5]. Their motivation will further be fuelled by factors related to their satisfaction and knowledge self-efficacy [6].

Besides research on motivation, the growth of unanswered questions in CQAs has recently started to pique scholarly interests on what makes questions attractive to answers. In this vein, one related study found that the use of interrogative words such as “what” and “where” enhances the chance of receiving answers [7]. Another found the lack of clarity to be a key reason why answers to a given question had not been forthcoming [8]. Yet, current efforts to predict questions’ likelihood to attract answers in CQAs have not gone beyond examining the content of questions. Factors such as the time a question is being posted and from whom it originates have yet to receive serious attention.

Building on extant literature, this paper therefore seeks to shed light on the conundrum of unanswered questions in CQAs through the following two-fold objectives. First, it aims to develop a conceptual framework known as the Quest-for-Answer to explain why some questions in CQAs could draw answers while others remain ignored. The framework suggests that the likelihood to attract answers depends on both the metadata and the content of questions.

Second, the paper attempts to validate the Quest-for-Answer framework using data drawn from Stack Overflow¹, a closed domain CQA devoted to programming and software engineering. A total of 3,000 questions divided equally between those answered and unanswered were used for analysis. In particular, Stack Overflow was chosen because it is currently one of the largest CQAs with more than some 1.6 million registered users, who contributed some one million questions and 2.8 million answers [9, 10]. Its popularity stems from the fact

that earning badges of expertise in Stack Overflow serves as credentials for computer programmers to seek jobs [11]. Even as Stack Overflow is known to attract answers promptly [12], the volume of unanswered questions has been rising steeply in recent years [13]. This makes it a suitable platform for investigation.

On the theoretical front, this paper enriches CQA research by venturing into a relatively uncharted terrain. The scope of CQA research hitherto tends to be largely confined to answers-related themes such as quality of answers [14, 15, 16], and speed at which they were returned [12, 17]. Research on questions-related themes, on the other hand, is only budding. Therefore, developing the Quest-for-Answer framework represents a timely research endeavor. On the practical front, the findings might offer prescriptive guidelines for users to post questions in CQAs so as to improve their chances of receiving answers. Moderators of CQAs might also lean on the findings to devise strategies to minimize the volume of unanswered questions. Thus, this paper offers potential implications for both theory and practice.

The remainder of this paper is organized as follows. The next section reviews extant literature, and describes the Quest-for-Answer framework. The methods for data collection, operationalization, coding and analysis are presented next. This is followed by the results and the discussion. The paper concludes by highlighting its implications.

Literature review

Community question answering sites (CQAs)

Most users rely on search engines as primary gateways to seek information. Those who are familiar with library services may also take advantage of virtual reference services to meet their needs for specialized information [18]. The advent of CQAs represents another convenient avenue for online information seeking. Users are now able to ask questions directly on the Web and harness the collective wisdom of crowds:

Ask a hundred people to answer a question or solve a problem, and the average answer will often be at least as good as the answer of the smartest member.[19 p11]

Users' reliance on CQAs for information seeking has attracted considerable scholarly interests. In particular, a burgeoning area of research focuses on questions submitted in CQAs. A common objective involves developing question clustering algorithms (e.g., [20]) so that unanswered questions could be answered using the existing corpus of responses [21]. To further mitigate the problem of unanswered questions, studies such as [22] sought to automatically recommend answer providers who could be interested in responding to specific questions.

Despite such scholarly efforts, the reasons for which some questions in CQAs could draw answers and others remain ignored have not been comprehensively investigated. Among the few related studies, the presence of interrogative words such as "what" and "where" was found crucial in determining if a question will be answered [7]. Furthermore, vague questions that lack clarity are known to deter answers [8]. Building on such studies, this paper develops the Quest-for-Answer framework to understand what makes questions in CQAs likely to attract answers.

The Quest-for-Answer framework

The clue to attract answers could be found in features related to both the metadata and the content of questions. Metadata contain additional information about questions that can potentially affect questions' attractiveness in the CQA community [23, 24]. For example, questions asked by reputed askers, as indicated by reputation scores, might be more compelling than those submitted by askers with a mediocre reputation. Likewise, content of questions, such as the way questions have been phrased, could either engender or impede responses from answerers [7, 8]. For instance, questions that are clear are usually better poised to attract answers vis-à-vis those that are vague.

Metadata features are those that are automatically captured and displayed in CQAs. Some of these could be retrieved directly from the CQA interface, while others could be derived in the form of ratios from the

retrievable metadata. For the purpose of this paper, they are henceforth referred as primary metadata and secondary metadata respectively.

Three pertinent primary metadata are popularity and participation of askers, as well as asking time of questions. Popularity refers to the extent to which an asker is recognized in the CQA community. Most CQAs display a reputation score to summarize users' popularity. In general, any message contributed by a reputed source is viewed more favorably vis-à-vis one that is shared by an amateur source [25]. Likewise, answerers' reputation is known to shape the perceived quality of answers in CQAs [23]. Extrapolating this in the context of questions in CQAs, it would seem that askers' reputation could shape the likelihood of questions attracting answers.

Participation is a measure of the level of activity of an asker in the CQA community. Most CQAs indicate users' participation by displaying the volumes of questions asked, and answers posted. An asker, who participates regularly in asking and answering questions, could be perceived as an active member [24]. Questions posted by an active member might be deemed sincere, thereby making them likely to be answered.

Asking time refers to the hour of day as well as the day of week on which questions are posted in CQAs. It is conceivable that CQAs are not necessarily frequented by comparable user-bases throughout a day or a week [26]. If a question is posted at a time when numerous users are online, it might be likely to attract answers. Hence, it could be interesting to study the extent to which questions' asking time is related to their likelihood to attract answers.

Two secondary metadata could further determine if answers would be forthcoming in response to questions. These include askers' derived role and derived popularity. Derived role indicates the ways in which an asker contributes in the CQA community. In general, CQA users have two primary ways to contribute: asking and answering. The volumes of questions asked and answers posted by CQA users are known to shape their reputation in the community [27]. These could also suggest whether a user predominantly plays the role of an asker, or an answerer, or both. It is possible that if a question is posted by one who has the reputation of submitting frivolous enquiries indiscriminately, the CQA community might ignore it. In contrast, answers could be forthcoming in response to a question posted by one who is known to contribute questions as well as answers proportionately.

Derived popularity refers to the extent to which an asker is trusted by the CQA community. Given that most CQAs allow users to like and dislike one another's activities, the ratio of applauses and criticisms could serve as a proxy for an asker's derived popularity. Such ratios have been used in prior CQA studies as indicators for community acceptance [28]. Questions posted by askers who attract more applauses than criticisms could be likely to attract answers.

Besides metadata, content features could also engender or impede responses from questions. For the purpose of this paper, content of questions is assessed based on content structure and content quality. Content structure refers to the structural elements of questions, while content quality is a measure of the extent to which questions are well-articulated.

Two important content structure features are level of details and specificity of questions. Level of details is a measure of the extent to which questions extensively express askers' information needs. Overly sketchy questions might not attract answers. Too lengthy questions could also deter responses [13]. However, [26] suggested that questions that were overly sketchy as well as those that were too lengthy were more likely to attract answers compared with those of moderate length. Given the lack of consensus, the role played by questions' level of details in attracting answers deserves further investigation.

Specificity refers to the extent to which questions are precise in expressing askers' information needs. Precision of questions could be enhanced with the use of specific interrogative words such as "what" and "where", which in turn, could make the enquiries attractive to answers [7].

Three content quality features could be further related to questions' likelihood of attracting answers. These include accuracy, clarity and socio-emotional value. Accuracy refers to the extent to which questions are correct. Questions marred with incorrect spellings or ambiguous logic could fail to attract answers [29, 30].

Clarity refers to the understandability of questions. Clearly-articulated questions that express information needs comprehensively could be more likely to attract answers than those that are complex [8, 31].

Socio-emotional value refers to the extent to which questions express emotions and subjectivity. Askers could thank answerers in advance by expressing emotions of appreciation [32]. Moreover, subjective expressions in questions could make answers forthcoming by portraying the social context for askers' information needs [26, 30].

The features discussed above are aggregated into the Quest-for-Answer framework as shown in Table 1. Specifically, the framework identifies features related to both metadata and content that could explain why some questions remain unanswered in CQAs. Metadata includes five features, namely, popularity, participation, asking time, derived role, and derived popularity. On the other hand, content includes five features, namely, level of details, specificity, accuracy, clarity and socio-emotional value.

Table 1. The Quest-for-Answer Framework.

	Features	Descriptions	References
Metadata	Popularity	It refers to the extent to which an asker is recognized in the community.	[23]
	Participation	It is a measure of the level of activity of an asker in the community.	[24]
	Asking time	It refers to the hour of day as well as the day of week on which questions are posted.	[26]
	Derived role	It refers to the extent to which an asker leans towards answering in relation to asking.	[27]
	Derived popularity	It refers to the extent to which an asker is trusted by the community.	[28]
Content	Level of details	It is a measure of the extent to which questions extensively express askers' information needs.	[13]
	Specificity	It refers to the extent to which questions are precise in articulating askers' information needs.	[7]
	Accuracy	It refers to the extent to which questions are correct.	[29]
	Clarity	It refers to the understandability of questions.	[31]
	Socio-emotional value	It refers to the extent to which questions express emotions and subjectivity.	[32]

Methods

Data Collection

As indicated earlier, data for this paper were drawn from Stack Overflow. The question-answering cycle in Stack Overflow commences when a user asks a question. Specifically, the CQA requires askers to post questions comprising two parts, namely, titles and descriptions. Stack Overflow allows askers to annotate

questions with one or more tags such as 'Java' or 'C++' to indicate the subject matter of the enquiries. After receiving answers to a given question, the corresponding asker can choose a response as satisfactory. The question is then said to be resolved. Throughout this cycle, users can praise one another's activity through upvotes, or criticize by casting downvotes. The extent to which users are popularly trusted by the community is indicated by their reputation scores.

For collecting data, SQL queries were executed in the Data Explorer service provided by Stack Exchange² in February, 2014. In particular, all answered and unanswered questions with the tag 'Java' that had been posted in Stack Overflow since 2008 were retrieved. The 'Java' tag was chosen because it was found to be the most popular tag³ as of January, 2014. This corroborates with the fact that Java was rated as the most popular programming language⁴ among IT professionals as of January, 2014. The data collection yielded an initial pool of 21,801 questions (12940 answered + 8861 unanswered). From the initial pool, simple random sampling was used to identify a total of 3,000 distinct questions (1500 answered + 1500 unanswered) for admission into the final dataset.

For every question, the Data Explorer returned a total of some 36 data fields. Of these, two represented content, namely, question title, and question description. The remainder included metadata, which could be grouped into four categories. The first entailed identifiers such as question identifiers, and asker identifiers. The second comprised timestamps such as those for asking questions, and the most recent activity (if any). The third entailed details about questions such as number of tags, scores, view counts. The fourth comprised details about askers such as reputation scores, as well as counts for upvote and downvote. Informed by the literature (e.g., [13, 23, 27, 31]), the features in the Quest-for-Answer framework were operationalized by a set of 20 measures that could be assessed from the retrieved data fields.

Operationalization

With respect to metadata, popularity of askers was operationalized as their reputation score, as well as volumes of upvotes and downvotes received [23]. Askers' level of participation was operationalized as their duration of membership in years in the Stack Overflow community, as well as of volumes of questions asked and answers posted [24]. For duration of membership in years, ceiling values were taken. Asking time was operationalized as the hour of a day, and the day of a week when questions were posted [26]. These values were computed from the raw timestamp of question posting using built-in functions offered in Microsoft Office Excel. Askers' derived role was operationalized as the number of answers they had posted per question [27]. A lower ratio would indicate that the asker posted questions in CQAs without attempting to answer others' queries. On the other hand, a higher ratio would indicate that the asker played an active role in the community not only by posting questions but also by answering. Askers' derived popularity was operationalized as the number of upvotes they had garnered per downvote. A lower ratio would indicate lower derived popularity of askers and vice-versa [28].

With respect to content, level of details was operationalized based on the length of questions in words [13]. Given that Stack Overflow solicits questions comprising two parts: titles and descriptions, the number of words in both was considered. Specificity of questions was operationalized as the number of tags, presence of code snippets, and presence of interrogative words such as "what", "when", "where", "which", "who", "whom", "whose", "why", and "how" [7]. Accuracy was operationalized as the extent of linguistic errors in questions (reverse-coded) [30]. Clarity was operationalized as the completeness and the complexity (reverse-coded) of questions [31]. Socio-emotional value was operationalized as the politeness and subjectivity expressed in questions [32]. The operationalized measures for the features in the Quest-for-Answer framework are presented in Table 2.

Table 2. Operationalized measures for the features in the Quest-for-Answer Framework.

	Features	Operationalized measures
Metadata	Popularity	Reputation score, # Upvotes, # Downvotes
	Participation	Membership duration in years, # Questions, # Answers
	Asking time	Hour of day, Day of week
	Derived role	# Answers per question
	Derived popularity	# Upvotes per downvote
Content	Level of details	Title length in words, Description length in words
	Specificity	# Tags, Code snippets, Interrogative words
	Accuracy	Linguistic errors (reverse-coded)
	Clarity	Completeness, Complexity (reverse-coded)
	Socio-emotional value	Politeness, Subjectivity

Coding

The operationalized measures for all the metadata, as well as those for level of details and specificity could be obtained directly from the dataset. However, the operationalized measures pertaining to accuracy (linguistic errors), clarity (completeness and complexity), and socio-emotional value (politeness and subjectivity) required content analysis [33]. For this purpose, three research associates were recruited as coders. They were graduate students of Information Systems, and users of Stack Overflow. Moreover, they had adequate knowledge of Java programming.

The coders were asked to code the quality of questions in terms of the five dimensions, namely, linguistic errors, completeness, complexity, politeness and subjectivity on a scale of 1 (lowest) to 5 (highest). The coding comprised two stages. In the first stage, the coders conferred among themselves to get acquainted with the five dimensions, and code an initial set of 180 questions (90 answered + 90 unanswered) randomly drawn from the dataset. The mean pair-wise Cohen's kappa among the coders indicated a non-chance level of inter-coder agreement as follows: linguistic errors ($\kappa = 0.744$), completeness ($\kappa = 0.752$), complexity ($\kappa = 0.786$), politeness ($\kappa = 0.751$), subjectivity ($\kappa = 0.792$).

In the second stage, the coders separately coded the remaining 2,820 questions. Each coder received comparable volumes of answered and unanswered questions. To avoid biases, coders were kept unapprised of whether a given question was answered or unanswered throughout the duration of the content analysis. Moreover, they were requested not to look up for the questions' status on the Internet.

Data analysis

Hierarchical logistic regression was used for data analysis. The dependent variable (DV) included whether questions were answered or unanswered. It was dummy-coded such that 1 indicated answered questions, and 0 denoted unanswered ones. The independent variables (IVs) included the 20 operationalized measures of the features in the Quest-for-Answer framework. Four of these IVs were categorical in nature. These included hour of day (24 categories) and day of week (7 categories) of asking questions, as well as the presence or

absence of code snippets and interrogative words in questions (2 categories for each). For each categorical IV, indicator contrast was used for analysis with the last category as the reference.

The hierarchical logistic regression analysis comprised three models. The first model included three control variables, namely, questions' age (in days) and view count, as well as askers' profile view count. The second model comprised the IVs pertaining to metadata. The third entailed the IVs pertaining to content. Such a hierarchical approach allowed examining the extent to which metadata and content could independently predict the likelihood of questions to attract answers.

Prior to analysis, the logistic regression model was checked for multicollinearity. A variance inflation factor value of less than 10 for all the IVs confirmed that multicollinearity was not a problem [34]. Thereafter, the performance of each of the three models was probed using Omnibus test, and Hosmer-Lemeshow goodness-of-fit test. The extent to which the IVs in the models could account for the variability in the DV was examined using Cox and Snell, as well as Nagelkerke pseudo- R^2 measures. The classification accuracy was also checked. The relationship between each IV and the DV was examined in terms of odds ratio.

Results

Table 3 presents the descriptive statistics of answered questions ($N = 1,500$) and unanswered ($N = 1,500$) questions in the dataset pertaining to the continuous IVs. The distributions of answered and unanswered questions in the dataset across hour of day as well as day of week are depicted in Figure 1 and Figure 2 respectively. A total of 1,820 questions (827 answered + 993 unanswered) in the dataset were found to contain code snippets. Of the remaining 1,180 questions, 673 were answered and 507 were unanswered. Besides, a total of 669 questions (409 answered + 260 unanswered) contained interrogative words. Of the remaining 2,331 questions, 1,091 were answered and 1,240 were unanswered.

As evident from Table 3, several IVs had standard deviations greater than means. This in itself is not an anomaly. Several prior studies had documented variables, for which standard deviations exceeded means [35, 36, 37, 38]. Nonetheless, this hints at possible non-normality, which might interfere with the multivariate analysis of logistic regression. To mitigate this problem, variable transformation approaches such as logarithm transformation or square root transformation is advocated [39, 40, 41]. For the purpose of this paper, square root transformation was applied on nine IVs, namely, reputation score, upvotes, downvotes, membership, questions, answers, answers per question, upvotes per downvote, and description length. Logarithm transformation was deliberately avoided as it was found to result in multicollinearity problems.

Table 3. Descriptive statistics (Mean $\pm$ SD) of answered and unanswered questions.

IVs	Answered questions	Unanswered questions
Reputation score	545.6 $\pm$ 4790.3	467.6 $\pm$ 2000.8
Upvotes	65.8 $\pm$ 212.4	58.3 $\pm$ 208.2
Downvotes	7.1 $\pm$ 50.3	4.6 $\pm$ 36.6
Membership	0.6 $\pm$ 0.8	0.9 $\pm$ 1.1
Questions	21.6 $\pm$ 57.2	18.8 $\pm$ 51.4
Answers	16.4 $\pm$ 126.4	16.8 $\pm$ 84.6
Answers/question	1.2 $\pm$ 11.8	0.8 $\pm$ 2.8
Upvotes/downvote	13.8 $\pm$ 55.3	13.5 $\pm$ 45.4
Title length	8.1 $\pm$ 3.7	8.8 $\pm$ 3.6
Description length	104.3 $\pm$ 138.6	167.0 $\pm$ 197.1
Tags	2.9 $\pm$ 1.2	3.6 $\pm$ 1.1
Linguistic errors	1.2 $\pm$ 0.6	1.3 $\pm$ 0.6
Completeness	3.7 $\pm$ 1.0	3.5 $\pm$ 1.0
Complexity	2.6 $\pm$ 0.9	3.4 $\pm$ 0.8
Politeness	2.0 $\pm$ 1.0	2.1 $\pm$ 1.2
Subjectivity	3.1 $\pm$ 1.2	3.4 $\pm$ 1.0

Figure 1. Distribution of answered and unanswered questions in the dataset across hour of day.

Figure 2. Distribution of answered and unanswered questions in the dataset across day of week.

Table 4 presents the results of the three logistic regression models. For the sake of brevity, only IVs with statistically significant odds ratio in the final model are shown. The control variable of questions' age was positively related to the DV ($\text{Exp}(\beta) = 1.002$, $p < 0.001$), indicating that questions that had been posted earlier were more likely to attract answers vis-à-vis those submitted recently.

With respect to metadata, volume of downvotes was positively related to questions' likelihood to attract answers ($\text{Exp}(\beta) = 1.124$, $p < 0.05$). Interestingly, answers were generally forthcoming in response to questions

submitted by askers with numerous downvotes. Duration of askers' membership was negatively related to the DV ($\text{Exp}(\beta) = 0.524$, $p < 0.001$). This indicates that questions posted by askers, who were relatively new in the community, were attractive to answers. Volume of questions asked by askers was however positively related to the DV ($\text{Exp}(\beta) = 1.102$, $p < 0.001$). This shows that questions posted by askers, who had submitted queries in large volumes, were likely to be answered. In terms of asking time, questions posted during the following hours of the day were likely to attract answers: hour 20 ($\text{Exp}(\beta) = 2.229$, $p < 0.05$), hour 21 ($\text{Exp}(\beta) = 2.063$, $p < 0.05$), and hour 22 ($\text{Exp}(\beta) = 1.790$, $p < 0.05$). Moreover, questions posted on the following days of the week deterred responses: Monday ($\text{Exp}(\beta) = 0.459$, $p < 0.001$), Tuesday ($\text{Exp}(\beta) = 0.619$, $p < 0.05$), Wednesday ($\text{Exp}(\beta) = 0.599$, $p < 0.01$), and Thursday ($\text{Exp}(\beta) = 0.551$, $p < 0.01$).

With respect to content, title length of questions was negatively related to questions' likelihood to attract answers ($\text{Exp}(\beta) = 0.950$, $p < 0.001$). Likewise, description length of questions was negatively related to the DV ($\text{Exp}(\beta) = 0.919$, $p < 0.001$). These indicate that questions with succinct titles and descriptions were more likely to attract answers compared with those that were verbose. The number of tags provided in questions was also negatively related to the DV ($\text{Exp}(\beta) = 0.703$, $p < 0.001$). The fewer the tags associated with a question, the more likely was it to receive answers. The absence of code snippets was positively related to the DV ($\text{Exp}(\beta) = 1.641$, $p < 0.001$). In other words, questions without code snippets were receptive to answers. The absence of interrogative words was negatively related to the DV ($\text{Exp}(\beta) = 0.642$, $p < 0.001$). Stated otherwise, questions without interrogative words deterred responses. Completeness of questions was positively related to the DV ($\text{Exp}(\beta) = 2.045$, $p < 0.001$). This indicates that questions that expressed information needs comprehensively were more likely to attract answers vis-à-vis those that were incomplete. Complexity of questions was negatively related to the DV ($\text{Exp}(\beta) = 0.363$, $p < 0.001$), pointing out that less complex questions were likely to attract answers. Politeness in questions was also negatively related to the DV ($\text{Exp}(\beta) = 0.901$, $p < 0.05$). In other words, less polite questions were ironically found attractive to answers.

The final logistic regression model corresponding to the Quest-for-Answer framework was found statistically significant based on the Omnibus test in predicting if questions are likely to be answered ($\chi^2 = 1347.32$, $df = 48$, $p < 0.001$). A non-significant Hosmer-Lemeshow goodness-of-fit test ($\chi^2 = 6.24$, $df = 8$, $p > 0.05$) further confirms that the model fits well with the data. Based on Cox and Snell, as well as Nagelkerke pseudo- R^2 measures, the model accounted for about 36% to 48% of variability in the DV. It recorded a classification accuracy of 77.50% in distinguishing between answered and unanswered questions.

Table 4. Results of the hierarchical logistic regression analysis.

IVs	Model 1	Model 2	Model 3
Question age	1.002 ***	1.002 ***	1.002 ***
Downvotes ^a		1.176 **	1.124 *
Membership ^a		0.498 ***	0.524 ***
Questions ^a		1.128 ***	1.102 ***
Hour of day: 20		2.202 *	2.229 *
Hour of day: 21		1.882 *	2.063 *
Hour of day: 22		1.938 *	1.790 *
Day of week: Mon		0.541 ***	0.459 ***
Day of week: Tue		0.732	0.619 *
Day of week: Wed		0.704 *	0.599 **
Day of week: Thu		0.613 **	0.551 **
Title length			0.950 ***
Description length ^a			0.919 ***
Tags			0.703 ***
Code snippets: Absent			1.641 ***
Interrogative words: Absent			0.642 ***
Completeness			2.045 ***
Complexity			0.363 ***
Politeness			0.901 *
Omnibus test χ^2	354.31 ***	582.39 ***	1347.32 ***
Hosmer-Lemeshow test χ^2	152.99 ***	27.36 **	6.24

Cox & Snell R ²	0.11	0.18	0.36
Cox & Snell incremental R ²	0.11	0.07	0.18
Nagelkerke R ²	0.15	0.24	0.48
Nagelkerke incremental R ²	0.15	0.09	0.24
Classification accuracy	64.50%	68.00%	77.50%

*** p < 0.001; ** p < 0.01; * p < 0.05; ^a Square root transformed variables

Discussion

The Quest-for-Answer framework seems promising in predicting if questions are likely to attract answers. From the results, three main findings could be gleaned. First, questions posted by askers with numerous downvotes ($\text{Exp}(\beta) = 1.124$, $p < 0.05$), and short duration of membership ($\text{Exp}(\beta) = 0.524$, $p < 0.001$) were found likely to be answered. Additionally, questions posted by askers who had previously submitted large volumes of queries were found more attractive to answers vis-à-vis those posted by askers who had submitted few queries ($\text{Exp}(\beta) = 1.102$, $p < 0.001$). The CQA community welcomes questions from newly registered users as well as those with several downvotes perhaps out of altruism. Besides, the community seems to favor questions from those who are active in posting queries in the community. Interestingly, it was found that askers' reputation score, volumes of upvotes and downvotes attracted, derived role as well as derived popularity did not play a significant role in making answers forthcoming. Overall, the CQA community seems to accept questions from all users – novices and experts alike – without engaging in online hegemony. This finding is at odds with studies such as [26], which suggested that questions posted by expert askers were more likely to be answered compared with those asked by ingénues.

Second, the ways in which questions are phrased play a significant role in determining their likelihood to attract answers. Specifically, questions with short titles ($\text{Exp}(\beta) = 0.950$, $p < 0.001$), short descriptions ($\text{Exp}(\beta) = 0.919$, $p < 0.001$), and few tags ($\text{Exp}(\beta) = 0.703$, $p < 0.001$) were likely to attract answers. This finding is consistent with that of [13], which found that answered questions were shorter on average vis-à-vis unanswered ones. Interestingly, [26] found that sketchy as well as lengthy questions were more likely to be answered in Yahoo! Answers compared with those that were of moderate length. However, lengthy questions in Stack Overflow failed to attract answers. In another related study that analyzed some 385 questions from Stack Overflow, questions with code snippets were found likely to be answered [35]. In contrast, using a dataset of 3,000 questions, this paper found that such questions were less likely to be answered vis-à-vis those without code-snippets ($\text{Exp}(\beta) = 1.641$, $p < 0.001$). Even though Stack Overflow is supposed to deal with programming issues, the community appears disinclined to respond to questions containing code snippets. It is conceivable that evaluating code snippets, especially lengthy ones, and pinpointing syntactical or logical errors could be a daunting task. Moreover, if a suggested answer fails to work, the unsatisfied asker might seek further clarification from the answerer, leading to rounds of back-and-forth exchanges. Hence, questions with code snippets might deter responses.

Moreover, this paper found that the presence of interrogative words such as “what” and “where” in questions ($\text{Exp}(\beta) = 0.642$, $p < 0.001$), as well as their completeness ($\text{Exp}(\beta) = 2.045$, $p < 0.001$) increased the chance of attracting answers [7, 8]. Conceivably, complex questions were unlikely to be answered ($\text{Exp}(\beta) = 0.363$, $p < 0.001$). Consistent with prior studies, questions that maintained an overly polite tone were however found unlikely to be answered ($\text{Exp}(\beta) = 0.901$, $p < 0.05$) [26]. Askers could use polite words in questions to thank potential answerers in advance. However, the use of such words appeared to have backfired by serving as a distraction. Being a technical forum, the Stack Overflow community might not appreciate formalities in questions. Interestingly, the presence of linguistic errors in questions did not necessarily deter responses. This is inconsistent with [30], which suggested that questions plagued with linguistic errors would fail to attract answers. Stack Overflow seems to be patronized by users who are not too particular about linguistic accuracy of questions as long as the queries are technically understandable.

Third, the time of posting questions in CQAs is a key factor in determining their likelihood to attract answers. It was found that questions posted between 8pm and 11pm were generally receptive to answers ($1.790 \leq \text{Exp}(\beta) \leq 2.229$, $p < 0.05$). Consistent with [8], the volumes of unanswered questions were generally lower in week-ends vis-à-vis week days ($0.459 \leq \text{Exp}(\beta) \leq 0.619$, $p < 0.05$). In particular, answers appeared to be

forthcoming in response to questions posted on Friday, Saturday and Sunday. Therefore, in order to maximize chances of receiving answers, askers should ask questions between 8pm and 11pm on Friday, Saturday, and Sunday.

Interestingly, the majority of questions in Stack Overflow appeared to be posted 6am onwards till 7pm (Figure 1), and from Monday to Friday (Figure 2). Such an asking time strikingly overlaps with business hours to a large extent. Given that Stack Overflow is a CQA devoted to programming and software engineering, it is perhaps widely used by IT professionals as an information seeking platform for their office work. However, the findings suggest that askers should refrain from posting questions during business hours. Instead, it would be better to submit the questions between 8pm and 11pm on Friday, Saturday and Sunday to maximize chances of receiving answers. The findings gleaned by validating the Quest-for-Answer framework using data drawn from Stack Overflow are summarized in Table 5.

Table 5. Findings in Stack Overflow pertaining to the Quest-for-Answer Framework.

	Features	Findings in Stack Overflow
Metadata	Popularity	Questions posted by askers who had received several downvotes are likely to be answered.
	Participation	Questions posted by askers who are new in the community as well as those who had posted numerous queries are likely to be answered.
	Asking time	Questions posted between 8pm and 11pm on Friday, Saturday, and Sunday are likely to be answered.
	Derived role	Answered and unanswered questions do not differ in terms of askers' derived role.
	Derived popularity	Answered and unanswered questions do not differ in terms of askers' derived popularity.
Content	Level of details	Questions with short titles and short descriptions are likely to be answered.
	Specificity	Questions with interrogative words, few tags but no code snippets are likely to be answered.
	Accuracy	Answered and unanswered questions do not differ in terms of the presence of linguistic errors.
	Clarity	Complete and non-complex questions are likely to be answered.
	Socio-emotional value	Irrespective of subjectivity, questions that do not use an overly polite tone are likely to be answered.

Conclusion

To address the growing volumes of unanswered questions in CQAs, this paper developed the Quest-for-Answer framework to explain why some questions in CQAs could draw answers while others remain ignored. The framework posited that questions' likelihood to attract answers could be predicated by 10 features related to metadata and content of questions. The framework was empirically validated by data drawn from Stack Overflow. The results were generally promising. With respect to metadata, askers' popularity, participation and asking time of questions were found significant in predicting if answers would be forthcoming. With respect to content, level of details, specificity, clarity and socio-emotional value of questions emerged significant in enhancing or impeding responses.

This paper offers implications for both theory and practice. On the theoretical front, its significance is two-fold. First, it extends the boundaries of CQA research by developing the Quest-for-Answer framework to identify features related to the metadata and the content of questions that maximize their likelihood to attract answers. Building on prior related studies ([7, 13], the framework helps predict if questions are likely to be answered. Thus, it deepens the understanding of why some questions in CQAs could draw answers while others remain unanswered.

Second, the Quest-for-Answer framework offers a number of future research opportunities. Although this paper validated the framework through the case of Stack Overflow, the dataset was confined to only one CQA platform. Moreover, it treated all questions as an indistinguishable corpus even though queries posted in CQAs could be of several types [42]. Therefore, it could be an interesting research direction to investigate the extent to which the Quest-for-Answer framework helps determine questions' likelihood to attract answers for different types of queries. Furthermore, even though prior studies have analyzed users' motivation to answer questions in general [2, 3, 6], the extent to which the metadata and the content of questions shape users' intentions to share knowledge have not been explored. Such studies are necessary to expand the boundaries of current research on CQAs.

On the practical front, this paper offers insights to askers on how and when questions should be posted in Stack Overflow for answers to be forthcoming. Specifically, it was found that askers should ask clear and succinct questions without being overly polite. They should use specific interrogative words such as "what" and "where" in questions, but abstain from using code snippets. Moreover, they would do well to post questions between 8pm and 11pm on Friday, Saturday, and Sunday. Moderators of Stack Overflow could also lean on the findings to prepare prescriptive guidelines for educating askers on ways to ask questions. This is especially important given that Stack Overflow could be largely frequented by employees of several IT companies. The findings of this paper therefore have implications in terms of community management for effective knowledge sharing among IT professionals.

Notes

1. <http://stackoverflow.com>
2. <http://data.stackexchange.com/stackoverflow/query/new>
3. <http://stackoverflow.com/tags?tab=popular>
4. <http://www.tiobe.com/index.php/content/paperinfo/tpci/index.html>

Acknowledgements

The authors would like to acknowledge Gary Kuen for his help in preparing the initial draft of this paper.

References

- [1] Li B and King I. Routing questions to appropriate answerers in community question answering services. In: *Proceedings of the International Conference on Information and Knowledge Management*. New York: ACM, 2010, pp. 1585-1588.
- [2] Casalo LV, Flavian C and Cuinaliu M. Relationship quality, community promotion and brand loyalty in virtual communities: Evidence from free software communities. *International Journal of Information Management* 2010; 30(4): 357-367.
- [3] Chen IYL. The factors influencing members' continuance intentions in professional virtual communities – A longitudinal study. *Journal of Information Science* 2007; 33(4): 451-467.
- [4] Chiu CM, Hsu MH and Wang ETG. Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories. *Decision Support Systems* 2006; 42(3): 1872-1888.
- [5] Wasko MM and Faraj S. Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly* 2005; 29(1): 35-57.
- [6] Jin XL, Zhou Z, Lee MK and Cheung CM. Why users keep answering questions in online question answering communities: A theoretical and empirical investigation. *International Journal of Information Management* 2013; 33(1): 93-104.
- [7] Choi E, Kitzie V and Shah C. A machine learning-based approach to predicting success of questions on social question-answering. In: *iConference Proceedings*. <https://www.ideals.illinois.edu/bitstream/handle/2142/36040/224.pdf> (2013, accessed December 2014), pp. 409-421.

- [8] Asaduzzaman M, Mashiyat AS, Roy CK and Schneider KA. Answering questions about unanswered questions of stack overflow. In: *Proceedings of the International Workshop on Mining Software Repositories*. IEEE, 2013, pp. 97-100.
- [9] Anderson A, Huttenlocher D, Kleinberg J and Leskovec J. Discovering value from community activity on focused question answering sites: A case study of Stack Overflow. In: *Proceedings of the International Conference on Knowledge Discovery and Data Mining*. New York: ACM, 2012, pp. 850-858.
- [10] Morrison P and Murphy-Hill E. Is programming knowledge related to age? An exploration of stack overflow. In: *Working Conference on Mining Software Repositories*. IEEE, 2013, pp. 69-72.
- [11] Carey K. Show me your badge. The New York Times. <http://www.nytimes.com/2012/11/04/education/edlife/show-me-your-badge.html> (2012, November 2, accessed December 2014).
- [12] Mamykina L, Manoin B, Mittal M, Hripcsak G and Hartmann B. Design lessons from the fastest Q&A site in the west. In: *Proceedings of the Conference on Human Factors in Computing Systems*. New York: ACM, 2011, pp. 2857-2866.
- [13] Saha RK, Saha AK and Perry DE. Toward understanding the causes of unanswered questions in software information sites: A case study of stack overflow. In: *Proceedings of the Joint Meeting on Foundations of Software Engineering*. New York: ACM, 2013, pp. 663-666.
- [14] Chua AYK and Banerjee S. So fast so good: An analysis of answer quality and answer speed in community question-answering sites. *Journal of the American Society for Information Science and Technology* 2013; 64(10): 2058-2068.
- [15] Fichman P. A comparative assessment of answer quality on four question answering sites. *Journal of Information Science* 2011; 37(5): 476-486.
- [16] Oh HJ, Yoon YC and Kim HK. Finding more trustworthy answers: Various trustworthiness factors in question answering. *Journal of Information Science* 2013; 39(4): 509-522.
- [17] Shah C. Effectiveness and user satisfaction in Yahoo! Answers. *First Monday* (2011 February 7, accessed January 2014); 16(2). <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3092/2769>
- [18] Shah C and Kitzie V. Social Q&A and virtual reference - Comparing apples and oranges with the help of experts and users. *Journal of the American Society for Information Science and Technology* 2012; 63(10): 2020-2036.
- [19] Surowiecki J. The wisdom of crowds. New York: Anchor Books, 2004.
- [20] Blooma MJ and Kurian JC. Clustering similar questions in social question answering services. In: *Proceedings of the Pacific Asia Conference on Information Systems*. <http://www.pacis-net.org/file/2012/PACIS2012-073.pdf> (2012, accessed October 2014).
- [21] Shtok A, Dror G, Maarek Y and Szpektor I. Learning from the past: Answering new questions with past answers. In: *Proceedings of the International Conference on World Wide Web*. New York: ACM, 2012, pp. 759-768.
- [22] Zheng X, Hu Z, Xu A, Chen D, Liu K and Li B. Algorithm for recommending answer providers in community-based question answering. *Journal of Information Science* 2012; 38(1): 3-14.
- [23] Agichtein E, Castillo C, Donato D, Gionis A and Mishne G. Finding high-quality content in social media. In: *Proceedings of the International Conference on Web Search and Web Data Mining*. New York: ACM, 2008, pp. 183-194.
- [24] Jeon J, Croft WB, Lee JH and Park S. A framework to predict the quality of answers with non-textual features. In: *Proceedings of the SIGIR Conference on Research and Development in Information Retrieval*. New York: ACM, 2006, pp. 228-235.
- [25] Wathen CN and Burkell J. Believe it or not: Factors influencing credibility on the web. *Journal of the American Society for Information Science and Technology* 2002; 53(2): 134-144.
- [26] Yang L, Bao S, Lin Q, et al. Analyzing and predicting not-answered questions in community-based question answering services. In: *Proceedings of the Association for the Advancement of Artificial Intelligence Conference*. Palo Alto, CA: AAAI, 2011, pp. 1273-1278.
- [27] Bian J, Liu Y, Zhou D, Agichtein E and Zha H. Learning to recognize reliable users and content in social media with coupled mutual reinforcement. In: *Proceedings of the International Conference on World Wide Web*. New York: ACM, 2009, pp. 51-60.
- [28] Agichtein E, Liu Y and Bian J. Modeling information-seeker satisfaction in community question answering. *ACM Transactions on Knowledge Discovery from Data* 2009; 3(2): 10: 1-27.

- [29] Ignatova K, Bernhard D and Gurevych I. Generating high quality questions from low quality questions. In: *Proceedings of the Workshop on the Question Generation Shared Task and Evaluation Challenge*. <http://qg.cs.memphis.edu/papers/QG2008/PDFs/8-Ignatova-QG08.pdf> (2008, accessed August 2014).
- [30] Kitzie V, Choi E and Shah C. Analyzing question quality through intersubjectivity: World views and objective assessments of questions on social question-answering. In: *Proceedings of the American Society for Information Science and Technology*. 2013, 50: 1-10.
- [31] Riloff E, Wiebe J and Phillips W. Exploiting subjectivity classification to improve information extraction. In: *Proceedings of the National Conference on Artificial Intelligence*. Palo Alto, CA: AAAI, 2005, pp. 1106-1111.
- [32] Kim S and Oh S. Users' relevance criteria for evaluating answers in a social questions and answers site. *Journal of the American Society for Information Science and Technology* 2009; 60(4): 716-727.
- [33] Krippendorff K. Content analysis: An introduction to its methodology. Beverly Hills, CA: Sage, 2004.
- [34] Myers RH. Classical and modern regression with applications (Vol. 2). Belmont, CA: Duxbury Press, 1990.
- [35] Ajiferuke I, Wolfram D and Famoye F. Sample size and informetric model goodness-of-fit outcomes: A search engine log case study. *Journal of Information Science* 2006; 32(3): 212-222.
- [36] Krampen G, Fell C and Schui G. Psychologists' research activities and professional information-seeking behaviour: Empirical analyses with reference to the theory of the intellectual and social organization of the sciences. *Journal of Information Science* 2011; 37(4): 439-450.
- [37] Luo J, Ba S and Zhang H. The effectiveness of online shopping characteristics and well-designed websites on satisfaction. *MIS Quarterly* 2012; 36(4): 1131-1144.
- [38] Whitty MT, Buchanan T, Joinson AN and Meredith A. Not all lies are spontaneous: An examination of deception across different modes of communication. *Journal of the American Society for Information Science and Technology* 2012; 63(1): 208-216.
- [39] Johnson DE. Applied multivariate methods for data analysts. Pacific Grove, CA: Duxbury Press, 1998.
- [40] Tabachnick BG and Fidell LS. Using multivariate statistics (3rd edition). New York, NY: HarperCollins College Publishers, 1996.
- [41] Fox AB, Bukatko D, Hallahan M and Crawford M. The medium makes a difference: Gender similarities and differences in instant messaging. *Journal of Language and Social Psychology* 2007; 26(4): 389-397.
- [42] Treude C, Barzilay O and Storey MA. How do programmers ask and answer questions on the web?: Nier track. In: *Proceedings of the International Conference on Software Engineering*. IEEE, 2011, pp. 804-807.