How Many Answers Are Enough? : Optimal Number of Answers for Q&A Sites

Pnina Fichman
Indiana University, Bloomington, Indiana, United States
fichman@indiana.edu

Abstract

With the proliferation of the social web, questions about information quality and optimization attract the attention of IS scholars. Question-answering (QA) sites, such as Yahoo!Answers, have the potential to produce good answers, but at the same time not all answers are good and not all QA sites are alike. When organizations design and plan for the integration of question answering services on their sites, identification of good answers and process optimization become critical. Arguing that 'given enough answers all questions are answered successfully,' this paper identifies the optimal number of posts that generate high quality answers. Based on content analysis of Yahoo! Answers' informational questions (n=174) and their answers (n=1,023), the study found that seven answers per question are 'enough' to provide a good answer.

Keywords

Q&A sites, QA, CQA, optimization, Web 2.0, information quality.

1. Introduction

One of the goals of IS research is to find ways "to increase the timeliness, accuracy, and completeness of information at a minimum of costs---economic, cognitive, political, social, affective, and physical. At the heart of IS research, then, is a complex optimization problem" [1, p. 13]. As such, it is not surprising that information quality is a focus of much IS research (e.g., [2-3]). The challenges associated with information quality, both conceptual and practical, are not new but, with the adoption of information technology, organizations are faced with additional challenges. This complexity further intensifies as organizations try to leverage the potential of the social Web, mass collaboration, and free and open source software (FOSS). Thus, scholars have examined the potential and challenges associated with organizations using FOSS [4], and the potential of cost reduction and innovation by means of crowdsourcing [5-7]. With these complexities in mind, optimization is still one of the core challenges in IS research and practice.

The proliferation of mass information production on the social Web (e.g., Wikipedia, Yahoo! Answers) raises many questions about the reliability of user-created content. Empirical support for the potential of crowdsourcing, for example, is provided by consistent reports that the quality of Wikipedia entries is as good as those in traditional encyclopaedias (e.g., [8]) and that the Wikipedia Reference Desk is as good as reference services provided by libraries [9]. At the same time, concerns about the rise of a culture of mediocrity fostering a cult of amateurs [10] where everything is miscellaneous [11].

Scholars try to explain why and how the participatory nature of Web 2.0 provides an infrastructure for achieving high quality knowledge production. A popular explanation suggests that it is the "wisdom of crowds" [12]. Another explanation comes in the form of Linus' Law: "given enough eyeballs, all bugs are shallow" [13]. However, 'enough' may mean some but not too many, as the cliché argues that too many cooks can spoil the broth. In the context of FOSS, this rationale leads to Brooks' Law [14], which claims that increasing the number of developers in a project can introduce inherent coordination complexity that may hinder group performance.

Like FOSS and Wikipedia, Question Answering (QA) sites draw on mass collaboration and user participation. They are based on the idea that "everyone knows something" [15, p. A01], and that through collaborative knowledge production, users can provide answers to questions that are being asked. The growing popularity of these sites in terms of the number of users, questions, and answers is fascinating. For example, Yahoo! Answers is among the most frequently consulted reference sites, second only to Wikipedia. By the end of 2009, Yahoo! Answers boasted 1 billion questions and answers, 179 million users, and over 200 million visitors worldwide [16]. If QA sites provide high quality information while reducing costs, then organizations can utilize similar mechanisms for mass user participation to improve their services; specifically, information intermediation services can leverage this potential through crowdsourcing their services. While the potential benefit of QA sites providing quality information has been empirically documented (e.g., [9, 17]), great caution must be advised because information quality varies between answers and across different QA sites (e.g., [17-18]). Assuming that answer multiplication[1] is beneficial, a few questions should be addressed: Is there an optimal number of answers/answerers per question that leads to the best outcomes in terms of information quality? How many answers per question are 'enough' to produce a good answer? Is it possible that after an optimal number of answers have been posted, the added value of additional answers is minimal or may even hinder answer quality? Is it likewise possible that many answers are still not 'enough' and that, regardless of their number, answer quality is low? This optimization issue is critical when organizations design and plan for the integration of QA services on their sites. The goal of this study is to answer the question: How many answers does it take to provide a good answer on QA sites?

[1] Answer multiplication means that many answers can be posted for a single question.

Content analysis of informational questions (n=174) and their multiple answers (n=1,023) from Yahoo! Answers was performed at two levels of analysis. Findings reveal that answer multiplication significantly improves answer quality and that, in order to provide a reliable answer, seven answers per question are 'enough'.

2. Background

1.1. Question answering sites

There is a growing body of research on QA sites that focuses on information retrieval, information seeking behaviour and use, information intermediation, and the social dynamics of these online communities (e.g., [9, 17, 19-25]). In their respective areas researchers argue that QA sites change information creation, dissemination, intermediation, retrieval, seeking, and use. Most of these studies have focused on Yahoo! Answers; some have examined other QA sites, such as Answerbag, [20, 26-27], Wikipedia Reference Desk [25], and Naver [28], while several have examined and compared multiple QA sites in their studies [9, 17, 24, 29]. One common motivation for research in these domains follows the assumption that there is added value in achieving a better understanding of the question answering process (information intermediation, information reuse, and information retrieval) and outcomes (information quality in terms of answer quality). Information retrieval researchers, for example, assume that the crowd produces information that should be archived and reused because of its quality. This assumption justifies their efforts to identify high quality answers, incorporating social aspects such as user reputation and user ranking of answers.

The popular assumption about the potential benefits of collaborative question answering should not be taken for granted; it has been challenged because empirical findings show that information quality varies not only among answers but also across different QA sites [9, 17, 29]. Despite the fact that all QA sites exploit similar collaborative mechanisms to enable mass user participation, answer quality varies amongst them [9, 17]. Therefore, it is still unclear whether the crowd improves answer quality at all. The present study tries to address this gap, aiming to determine whether answer multiplication improves information quality.

This study tries then to uncover the conditions that can produce good answers, mainly by identifying the optimal number of answers per question and by asking how many answers are needed to yield a reliable answer. This optimization effort is critical for the future design and implementations of next-generation QA systems. It is also useful to examine whether common FOSS laws are applicable to QA sites. Specifically, assuming that bugs resemble questions in that they need to be identified or asked, processed or answered, and solved by the crowd, the study aims specifically to test whether Linus' Law is relevant here. In the context of QA sites, Linus' Law can be stated as follows: 'given enough answers, all questions are answered successfully.'

Posing this statement in the context of QA leads to three main challenges; the meaning of 'all' questions, the meaning of being 'answered', and finally, the meaning of being 'answered successfully'. First, not 'all' questions that are posted on QA sites are answered (e.g., [9, 17]). Response rates range between 16%-96% per QA site (rates of no response ranges between 4%-84%) [9]. Second, different types of questions might call for different answers and might require different evaluation criteria (e.g. [21, 23]); thus considering 'all' questions becomes a complex task. Third, what constitutes an answer is yet another challenge. For example, simply responding to a question with a random statement does not seem to be an answer to the question. Moreover, an answer could be 1) an individual post; 2) all posts for one particular question; 3) an answer that is collaboratively co-authored by more than one user; or 4) a chosen "best answer". Fourth, having an answer does not guarantee that the answer is of high quality (even when it is chosen as "best answer"). Thus, that a question has been successfully answered could mean different things to different scholars and the challenge of determining what makes a good answer becomes apparent. There are multiple points of view as to what constitutes a good answer and how answer quality should be evaluated, which include user rankings of "best answers", user reputation, user satisfaction, and content criteria of answers, such as answer accuracy and completeness [9]. Taking into account these challenges, this paper aims to identify what constitutes 'enough' in the context of question answering.

1.2. Information quality and answer quality

Scholarly publications about information quality are mostly practical and less theoretical [2]. Likewise in reference research, where answer quality has been assessed, "a lack of attention [has been] given to theory" [30, p. 3]. Information quality has attracted much research attention across many scholarly communities; among them are scholars engaged in information systems (IS) research and library and information science (LIS). In IS research for example, information quality is one of the key factors that affect IS success [e.g., 31] and in LIS information quality was examined, for example, through the lenses of information seeking behaviour research [e.g., 32] and reference research [33-34].

Information quality is a multidimensional construct with many different definitions and attributes [35]; it has been the centre of attention well before the introduction of the social web. With the increase interest in the quality of user-generated information, the concept continues to capture scholarly attention. Two different approaches to information quality seem to be prominent [35]. The first is subjective, focusing on users' judgment of information credibility [22-23, 32] or user perceptions of fitness of use [35], and the other focuses on objective measures of an information artefact (a website or an answer), such as accuracy and completeness [25, 33, 36]. The utilization of the second approach to information quality in the study of answer quality on Q&A sites can be useful, but poses certain challenges, as the artefact is dynamic and multifaceted. Under the objective approach, high quality answers were determined based on content analysis of the answers [17, 19, 25-26, 37]. Scholars that analyzed the content of answers have found, for example, that better answers are longer [17, 37-38], or include references to external sources [26]. Interestingly, question category, answer accuracy and completeness, and length of answer are significant predictors of answer quality, whereas asker's and answerer's authority and reputation are not [37].

Prior research on answer quality on QA sites has primarily assessed quality using the subjective approach and was based on user rankings of "best answers". However, user rankings are problematic because they provide a subjective measure of answer

quality. Poston and Speier [39] argue that, "rating validity, [which] describes the degree to which the rating reflects the intrinsic quality of the content … may be low for a variety of reasons … [it is] inherently subjective and voluntarily provided, resulting in mismatch between the true quality of the content and the rating given … [and] those submitting ratings may manipulate ratings…" [39, p. 223]. For example, in 29.8% of cases where users chose "best answers" in Yahoo! Answers, their selections were based on socio-emotional criteria rather than on the content or utility of the answer [22]. Another method to identify answer quality is by tracking user reputation. This method is based on the assumption that certain users are more likely to provide better answers than others [40]. Examples of this approach include the ranking of authoritative responders using link analysis [41-42]. Other ranking methods measure users' reputations based on their activity levels (e.g., [37, 40, 43], their focus on one subject area [44], their credibility (authority), or the number of "best answers" they have previously posted [37, 45]. However, this approach is also problematic because even users with good reputations do not always provide high quality answers.

Measuring answer reliability by focusing on answer accuracy and completeness is another common approach in quality assessment on QA sites [9, 17, 25, 37, 46]. Under this objective approach, high quality answers have been determined based on content analysis of the answers [17, 19, 25-26, 37]. Scholars analyzing the content of answers have found, for example, that better answers are longer [17, 37, 39], and include references to external sources [26]. Researchers argue that different questions warrant a different type of answers and that not all measures of quality should apply to all answers [9, 23, 47]. They differentiate between conversational and informational questions [47], subjective and objective questions [48], or navigational, informational, transactional, and social questions [49].

QA sites are socio-technical systems where many different facilitating conditions can affect the quality of answers that can be found on them [18, 20, 24]. This led to the development of theoretical frameworks that integrate both the objective and subjective approach to determine QA sites effectiveness [e.g., 18, 37, 40]; answer quality is an important component in all of these frameworks. The present study examines the relationships between two components in the social reference model [18]: number of users (counting their posts) and answer quality (using reliability measures) under the objective approach to information quality. Given the lack of attention in these frameworks to the issue of optimization, the present study focuses on optimization. It also tests their underlying assumption that the crowd, by providing multiple answers to a given question, answers questions well enough.

3. Method

1.3. Data collection

Data were harvested from Yahoo! Answers, using a Perl program that was set up to collect on July 10th 2008 the most recent question per category over a 24-hour period at a random minute of every hour (24 points of time), and, 24 hours later, to collect all of the relevant answers. Using this method, a random sample of 585 transactions was collected. Yahoo! Answers was chosen because it is the most popular QA site [50]. A transaction includes a question and a whole answer. A whole answer includes any number of answers; most of the time the whole answer includes multiple answers. Transactions that include a "best answer" are resolved transactions. [2]

Figure 1 shows the distribution of answers per question; very few questions received a high number of answers and many of the questions received either very few or no answers at all. The number of answers per question varied between zero and 60 (M=6.12, SD=7.52), and while 70% of the questions received more than one answer, 16% of questions received no answers.

The amount of time that passed between the initial posting of questions and the posting of first answers ranged between 0:01 and 34:33 hours (M=1:01, SD=3:41). The amount of time that passed before last answers (in the data set) were posted ranged between 0:03 and 57:31 hours (M=6:24, SD=9:52).

Because different credibility criteria for informational and conversational questions were reported by users of Yahoo! Answers [23], the aim in this study was to focus further analysis only on informational questions. The questions were categorized, either as conversational or informational, using the following definitions [47, p. 759]:

> *Informational questions* are asked with the intent of getting information that the asker hopes to learn or use via fact- or advice-oriented answers. An example: *What's the difference between Burma and Myanmar?*
>
> *Conversational questions* are asked with the intent of stimulating discussion. They may be aimed at getting opinions, or they may be acts of self-expression. An example: *Do you drink Coke or Pepsi?*

First, the transactions were sorted into one of the two categories by one coder and later, 30% of the data was sorted into categories by a second coder to assure inter-coder reliability and strengthen the validity of the study results [51]. Inter-coder reliability was determined using simple agreement, also called percent agreement, which is based on the percentage of all codes that a pair of coders agreed on [52]. Inter-coder reliability resulted in 90% agreement, which is high; as a rule of thumb, co-efficiency of .90 or greater would be acceptable to all [51-52].

Seventy-three percent of the questions were informational (n=422), and the rest were conversational (n=163). Conversational

[2] A question becomes a Resolved Question when a Best Answer is chosen. After a question becomes Resolved it stays in Yahoo! Answers and is available for searching and browsing. The Best Answer remains open to receive comments and ratings from the community. Retrieved August 28, 2010 from http://help.yahoo.com/l/us/yahoo/answers/vote/vote-702891.html;_ylt=AmzwpYMfAD9jlGSb.RNsoc6hjSN4

transactions had significantly more answers per question (M=9.74, SD=10.73) compared with informational transactions (M=4.92, SD=6.39).

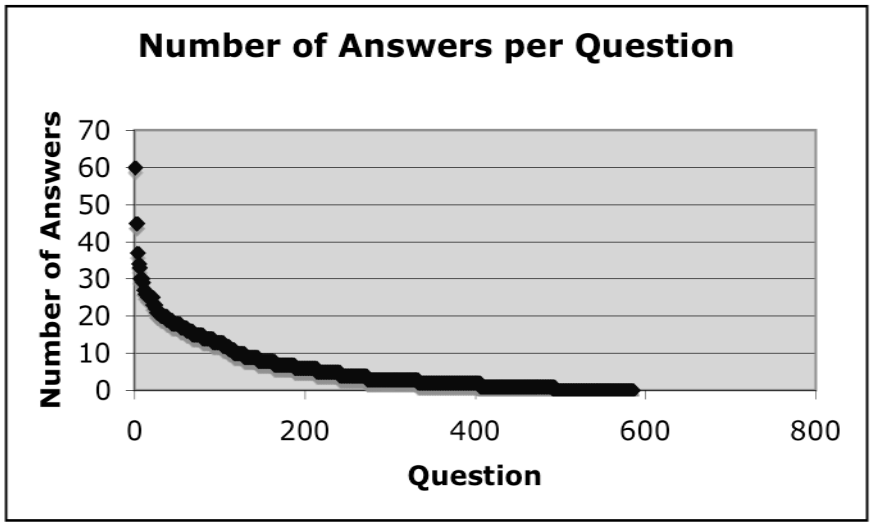

Fig. 1. Number of answers per question

Two samples that complement each other were drawn from the informational questions data set for manual content analysis. The first sample was a purposeful sample of resolved transactions (questions with "best answers") in line with prior research tendencies to include only resolved transactions; it included 74 transactions. However, because the resolved transaction sample included only 17% of the 422 informational transactions and only 12% of the entire data set of 585 transactions, a second with 100 random transactions was collected. The random sample included transactions with questions but no answers (n=19), transactions with answers but no "best answer" (n=65), and resolved transactions with "best answers" (n=16). The average number of answers per question was higher in the resolved transactions sample (M=7.81, SD=8.87) and lower in the random sample (M=4.45, SD=5.54) than it was in the entire informational questions data set (Table 1).

Table 1. Number of answers per question

	Random Sample (n=100)	**Resolved Sample (n=74)**
Mean	4.45	7.81
Std Dev	5.74	8.87
Min	0.00	1.00
Max	30.00	60.00

Data about the users were collected in November 2010, based on user profiles for each of the 100 transactions in the Random sample. At this time only 82 of the transactions were accessible, and 18 of them were not (most likely because these questions had no answers or "best answers" and have not been archived).

1.4. Data analysis

To determine answer reliability level, a content analysis of 174 transactions and 1,197 posts from Yahoo! Answers (174 questions and 1,023 answers in two samples), was conducted [53]. Content analysis of answers is a widely used method to evaluate answer quality on QA sites (e.g., [17, 19, 26, 37]) because it enables the evaluation of answer quality based on quantifying the presence or absence of quality measures (codes) in the answer.

Analysis was conducted at two levels: 1) transaction (n=174) – whole answer; 2) question-answer pair (n=1,023) – first answer and "best answer". Quality rates for the whole answer, the first answer, and the "best answer" were coded. The first answer is the first answer posted in response to a question. The "best answer" is the answer chosen as "best answer" by the asker or by a community vote. The "best answer" encompasses feedback about the fit between question and answer and a selection of one answer as being of good quality.3 Frequencies of reliability codes were aggregated for: whole answer, first answer, and "best answer" using three reliability measures: accuracy, completeness and verifiability. These three measures have been widely used in prior research on QA sites (e.g., [9, 37, 46]), and have been frequently used by Yahoo! Answers' users in their information credibility judgments [23]. Accuracy, completeness, and verifiability are of particular importance in judging the credibility of answers to informational questions [23]:

[3] Coders do not define the "best answer" but determine the quality of the individual answer that was chosen, in some transactions, as the "best answer."

1. **Accuracy** of an answer refers to a correct response.
2. **Completeness** of an answer refers to an answer that is thorough, provides enough information, and answers all parts of a multi-part question.
3. **Verifiability** of an answer refers to an answer that provides a link or a reference to another source where the information can be found.

Using these codes, two coders each coded the entire data set, assigning a value (yes/no) for each code (accuracy, completeness, and verifiability) to the transactions and the question-answer pairs. Coders were graduate students studying library and information science at a Midwestern university. They were instructed to determine the accuracy, completeness and verifiability of the answers "on the surface" [52] and based on their best judgment to verify information with external sources. Inter-coder reliability between the two iterations of coding of all the transactions was determined using simple agreement and Cohen's Kappa. Inter-coder reliability was 92%, which is high [51-52]; Cohen's Kappa was .84, which means that there was almost perfect agreement between the two coders [54].

First, frequency tables were created for each of the two samples (the random sample and the resolved transactions sample), tallying the presence of codes (yes values) for the whole answer, first answer, and "best answer". Then, the percentages of codes per answer were marked and statistical analysis using SPSS 17.0 was done. Later, the location of "best answer" was marked and cumulative quality rates were examined; data about the users, those who asked and answered questions in the random sample, were tallied as well.

1.5. User demographics

The random sample data was examined to collect additional user demographic information in November of 2010. At that time the Yahoo! Answers archive included a total of 322 profiles; 82 user profiles of askers and 450 user profiles of answerers for the archived transactions (Table 2).

Specific attention was given to gender because prior research has found that QA sites are female dominated [9]. As can be seen in Table 2, women, overall, ask more questions than men, answer more questions than men, post first answers more often than men, and provide more "best answers" than men. However, while women's answers are chosen as "best answer" by the asker more often than are men's answers, men's answers are chosen as "best answers" by vote more often than are women's answers.

Table 2. Gender distribution[4]

	Female	Male	Gender unknown
Questions asked (n=82)	39 (47.56%)	22 (26.82%)	22 (26.82%)
Answers posted (n=450)	227 (50.4%)	114 (25.3%)	109 (24.2%)
Best answer posted (n=73)	31 (42.46%)	26 (35.61%)	16 (21.91%)
Best answer, chosen by asker (n=25)	13 (52%)	6 (24%)	6 (24%)
Best answer, chosen by vote (n=48)	17 (35.41%)	21 (43.75%)	10 (20.83%)
First answer posted (n=83)	35 (42%)	22 (27%)	26 (31%)

1.6. Limitations

Because the study uses data only from one QA site, generalizations should be made with great caution, given that reports on variations across sites are significant [9, 24]. The use of data from one QA site only was required to control for site variations in user demographics, policies, and technological infrastructure. Still, Yahoo! Answers is the most popular QA site, where the vast majority of collaborative question answering is taking place, and as such it sets an example for the others to follow. Another limitation is the choice of informational questions, which count for 73% of all questions asked; success rates and evaluation criteria may differ from conversational questions. This type of question was chosen to facilitate the use of evaluation criteria across all the transactions. Nevertheless, these transactions and their success rates characterize the vast majority of transactions on Yahoo! Answers.

4. Findings

The results of the analysis of both samples at two levels of analysis (transaction and question-answer pair) are presented in Table 3. The two samples were compared on all three reliability measures (accuracy, completeness, and verifiability), for all three types of answers (first answer, whole answer, and "best answer"). The differences between the samples were not statistically significant (Table 4); the level of accuracy, completeness and verifiability for "best answer" and first answer did not differ between the samples, but the level of completeness for the whole answer was higher in the resolved transactions. In both samples, the whole answer and the "best answer" are significantly better than the first answer, and the "best answer" shows the highest levels of accuracy and completeness (Tables 3, 5). Verifiability levels are very low for both samples (Table 3). While in both samples there are small differences in verifiability levels between the first answer, "best answer", and whole answer, these differences are not statistically significant (Table 5). Completeness levels in both samples, and accuracy levels in the resolved sample, are significantly different (Table 5).

[4] Gender distribution is based on users' self report on their Yahoo! Answers' user profile as of November 2010.

Table 3. Rates on single variables

		Accurate	Complete	Verifiable
Resolved transactions (n=74)				
Best Answers	%	95%	96%	16%
	#	70	71	12
Whole Answers	%	89%	96%	18%
	#	66	71	13
First Answers	%	68%	62%	9%
	#	50	46	7
Random sample (n=100)				
Best Answers	16 resolved questions	88%	94%	13%
	81 answered questions	17%	18%	2%
	100 posted questions	14%	15%	2%
	#	14	15	2
Whole Answers	81 answered questions	89%	84%	14%
	100 posted questions	72%	68%	11%
	#	72	68	11
First Answers	81 answered questions	78%	57%	11%
	100 posted questions	63%	46%	9%
	#	63	46	9

Table 4. Results of cross tabulation between the two samples

	χ^2	Cramer's v	Df	Sig.
Accuracy – BA [1]	3.15	.12	1	-
Accuracy – WA [3]	0	0	1	-
Accuracy – FA [2]	2.54	.11	1	-
Completeness – BA	.42	.04	1	-
Completeness – WA	8	.2	1	***
Completeness – FA	.52	.05	1	-
Verifiability – BA	.36	.04	1	-
Verifiability – WA	.6	.05	1	-
Verifiability – FA	.22	.03	1	-

***p<.001, **p<.01, *p<.05

[1] BA= Best Answer; [2] FA= First Answer; [3] WA= Whole Answer

The findings indicate that answer multiplication significantly increases answer quality in terms of accuracy and completeness. Information reliability for whole answers is higher than for first answers.

As can be seen in Table 5, the level of answer reliability in the resolved sample, differs between the first answers, "best answers", and whole answers in terms of accuracy (χ^2=29.91, df=2) and completeness (χ^2=59.36, df=2), but not in terms of verifiability (χ^2=5.82, df=2). Follow-up pair-wise comparisons show that: 1) first answers are significantly less accurate than whole answers (χ^2=13.06, df=1) or "best answers" (χ^2=24.18, df=1); 2) "best answers" and whole answers are equally accurate; 3) first answers are significantly less complete than whole answers (χ^2=34.84, df=1) and "best answers" (χ^2=34.84, df=1); and 4) "best answers" and whole answers are equally complete.

Table 5. Cross-tabulation results for differences across "best answers", first answers and whole answers in both samples

		X^2	Cramer's v	Df	Sig.
Resolved sample					
	Accuracy	29.91	0.31	2	***
	Completeness	59.36	0.44	2	***
	Verifiability	5.89	0.13	2	-
	Follow-up pair-wise comparisons				
	Accuracy – BA[1] & FA[2]	24.18	0.32	1	***
	Accuracy – BA & WA[3]	2.45	0.11	1	-
	Accuracy – FA & WA	13.06	0.25	1	***
	Completeness – BA & FA	34.84	0.41	1	***
	Completeness – BA & WA	0.13	0	1	-
	Completeness – FA & WA	34.84	0.41	1	***
	Verifiability – BA & FA	2.24	0.1	1	-
	Verifiability – BA & WA	0.14	0.02	1	-
	Verifiability – FA & WA	2.49	0.11	1	-
Random sample[4]					
	Accuracy	5.8	1.3	2	-
	Completeness	43.17	0.37	2	***
	Verifiability	0.43	0.03	2	-
	Follow-up pair-wise comparisons				
	Accuracy – BA & FA	3.54	0.13	1	-
	Accuracy – BA & WA	0.05	0.01	1	-
	Accuracy – FA & WA	4.39	0.14	1	-
	Completeness – BA & FA	37.01	0.43	1	***
	Completeness – BA & WA	5.11	0.15	1	*
	Completeness – FA & WA	17.53	0.29	1	***
	Verifiability – BA & FA	0.19	0.03	1	-
	Verifiability – BA & WA	0.04	0.01	1	-
	Verifiability – FA & WA	0.41	0.04	1	-

***p<.001, **p<.01, *p<.05

[1] BA= Best Answer; [2] FA= First Answer; [3] WA= Whole Answer; [4] These calculations are based on 16 resolved questions for BA, and 81 questions with answers for FA and WA.

In the random sample, the level of completeness is significantly different between the first answers, "best answers", and whole answers (χ^2=43.17, df=2). Follow-up pair-wise comparisons show that first answers are significantly less complete than whole answers (χ^2=17.53, df=1) and "best answers" (χ^2=37.01, df=1).

As the results above indicate, answer accuracy and completeness improve for whole answers in comparison with first answers. Still, it is unclear how many answers are required to reach a quality of answer that is good enough. Looking at the average number of answers per transaction can provide one solution to this question (Table 1). Accuracy and completeness levels improve with an average of 7.81 answers per question (resolved sample), and level of accuracy improves with 4.45 answers per question (random sample). In other words, 5 answers are enough for an increase in levels of completeness from first answers to whole answers, while 8 answers are enough for an increase in levels of completeness and accuracy. However, because the quality of "best answers" is equal to that of whole answers on all measures in both samples, it is possible that, if a "best answer" has been selected, fewer answers than the average are enough. It is likewise possible that the quality can improve beyond the levels of whole answers or "best answers" for resolved transactions with more answers than the average. Further analysis moving beyond the simple averaging of numbers was done next, looking into the optimization challenge.

First, using data from the resolved sample, the cumulative percentage and frequency of "best answer" location have been noted (Table 6); Figure 2 illustrates the distribution of "best answer" location in the resolved transactions sample. In most of the transactions the "best answer" was one of the first three answers (59.45%), and in many cases the "best answer" was the first answer (28%). Only by the seventh answer did 80% of the resolved transactions have a "best answer" (Table 6).

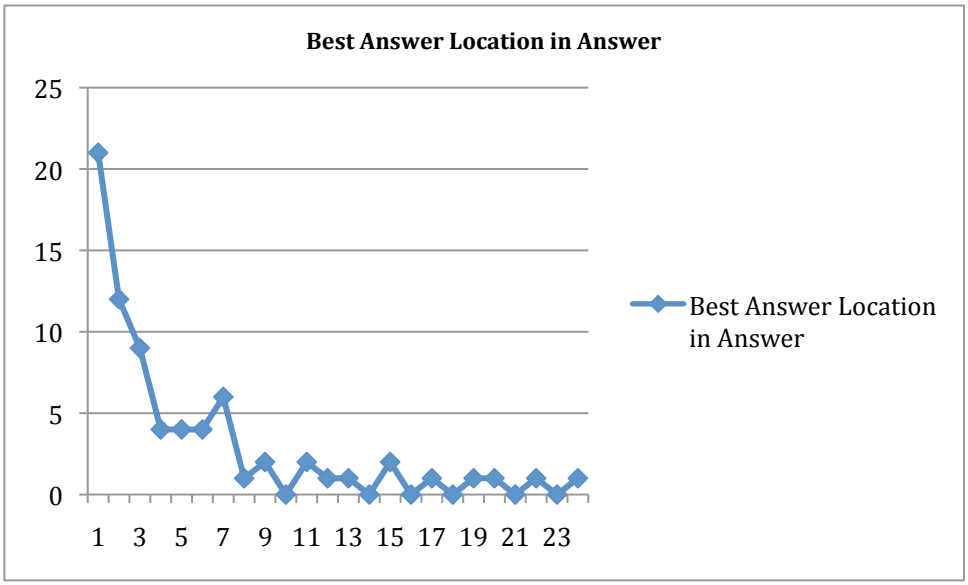

Fig. 2. Location of "Best Answer" (Resolved Sample)

Table 6. "Best answer" location in transaction

"Best Answer" Location in Answer	**First**	**Second**	**Third**	**Fourth**	**Fifth**	**Sixth**	**Seventh**
	Resolved Sample						
Number of Answers (n=74)	21	12	9	4	4	4	6
Cumulative Percent of Resolved Transactions (n=74)	28%	44%	56%	62%	67%	73%	81%
	Random Sample (2010 data)						
"Best Answer" Location in Answer (By Vote; n=48)	26	11	2	3	2	1	3
Cumulative Percent of Resolved Transactions (By Vote; n=48)	54%	77%	81%	88%	92%	94%	100%
"Best Answer" Location in Answer (By Asker; n=25)	10	3	3	1	1	1	1
Cumulative Percent of Resolved Transactions (By Asker; n=25)	40%	52%	64%	68%	72%	76%	80%

Next, the random sample was revisited in November 2010, when a higher percentage of the transactions had been resolved (n=73 in 2010 compared with n=17 at the time of original data collection); these "best answers" were chosen either by the asker (n=25) or by a community vote (n=48). The cumulative percentage and frequency showing the location of "best answers" have been noted (Table 6) for these transactions. All "best answers" chosen by the asker, and 80% of "best answers" chosen by a community vote were selected from the first 7 answers posted.

The findings from both samples indicate that 7 answers are enough to achieve an accuracy rate of 95% (Table 6). Further, while the findings indicate that it takes at least seven answers to achieve high quality information in the form of a "best answer" (Table 6), high quality answers appear also before and after the "best answer". The number of accurate and correct answers that are posted before the "best answer" strongly correlates with the "best answer" location (r=.86) and with the number of answers (r=.90) (Table 7). In fact, the total number of accurate and complete answers correlates with the total number of answers (r=.87) and there is a strong Pearson product-moment correlation coefficient between the number of answers and the location of the "best answer" in the transactions (r=.93) (Table 7). Moreover, it was evident that only two answers were needed to achieve accuracy in more than 80% of the resolved transactions (Table 8).

Table 7. Accurate and complete answers per transaction (resolved sample)

Transaction	Number of Answers	"Best Answer" Position	Number of Accurate & Complete Answers before "Best Answer"	Total Percent and Number of Accurate Answers	Total Percent and Number of Complete Answers	Total Percent and Number of Accurate & Complete Answers
1	3	3	0*	33% (1)	33% (1)	33% (1)
2	11	9	5	55% (6)	55% (6)	55% (6)
3	6	6	0	50% (3)	17% (1)	17% (1)
4	2	2	1	100% (2)	100% (2)	100% (2)
5	11	11	5	73% (8)	55% (6)	55% (6)
6	25	23	11	64% (16)	60% (15)	56% (14)
7	1	1	N/A	100% (1)	100% (1)	100% (1)
8	14	5	4	93% (13)	93% (13)	93% (13)
9	2	2	0*	50% (1)	100% (2)	50% (1)
10	14	4	1	43% (6)	43% (6)	43% (6)
11	20	15	6	55% (11)	50% (10)	50% (10)
12	17	17	12	100% (17)	76% (13)	76% (13)
13	3	1	N/A	100% (3)	100% (3)	100% (3)
14	12	11	2	25% (3)	25% (3)	25% (3)
15	6	2	0*	50% (3)	67% (4)	50% (3)
16	8	7	5	88% (7)	100% (8)	88% (7)
17	2	2	0*	50% (1)	50% (1)	50% (1)
18	1	1	N/A	100% (1)	100% (1)	100% (1)
19	2	1	N/A	100% (1)	100% (1)	100% (1)
20	14	2	0*	71% (10)	71% (10)	71% (10)
21	5	2	0*	80% (4)	60% (3)	60% (3)
22	14	5	2	86% (12)	86% (12)	86% (12)
23	6	2	0*	67% (4)	50% (3)	50% (3)
24	10	8	7	100% (1)	100% (1)	100% (1)
25	1	1	N/A	100% (1)	100% (1)	100% (1)
26	1	1	N/A	100% (1)	100% (1)	100% (1)
27	1	1	N/A	100% (1)	100% (1)	100% (1)
28	1	1	N/A	100% (1)	100% (1)	100% (1)
29	1	1	N/A	100% (1)	100% (1)	100% (1)
30	5	4	2	80% (4)	60% (3)	60% (3)
31	2	2	0*	100% (2)	50% (1)	50% (1)
32	3	3	2	100% (3)	100% (3)	100% (3)
33	10	4	2	70% (7)	70% (7)	70% (7)
34	15	15	10	73% (11)	73% (11)	73% (11)

35	2	2	0*	50% (1)	50% (1)	50% (1)
36	9	6	2	33% (3)	78% (7)	33% (3)
37	9	9	7	89% (8)	89% (8)	89% (8)
38	3	3	1	100% (3)	67% (2)	67% (2)
39	2	1	N/A	100% (2)	100% (2)	100% (2)
40	1	1	N/A	100% (1)	100% (1)	100% (1)
41	3	3	1	67% (2)	67% (2)	67% (2)
42	30	30	8	33% (10)	47% (14)	27% (8)
43	2	2	0*	0% (0)	100% (2)	0% (0)
44	2	2	1	100% (2)	100% (2)	100% (2)
45	2	1	N/A	100% (2)	100% (2)	100% (2)
46	1	1	N/A	100% (1)	100% (1)	100% (1)
47	4	1	N/A	75% (3)	75% (3)	75% (3)
48	1	1	N/A	100% (1)	100% (1)	100% (1)
49	25	22	21	100% (25)	100% (25)	100% (25)
50	2	1	N/A	0% (0)	0% (0)	0% (0)
51	14	6	4	79% (11)	79% (11)	79% (11)
52	9	1	N/A	78% (7)	67% (6)	56% (5)
53	5	5	2	100% (5)	60% (3)	60% (3)
54	4	3	1	75% (3)	75% (3)	75% (3)
55	15	13	5	60% (9)	53% (8)	53% (8)
56	15	7	6	100% (15)	100% (15)	100% (15)
57	14	12	8	79% (11)	71% (10)	71% (10)
58	4	3	2	100% (4)	100% (4)	100% (4)
59	7	6	3	57% (4)	43% (3)	43% (3)
60	10	7	4	80% (8)	80% (8)	80% (8)
61	15	7	1	47% (7)	40% (6)	27% (4)
62	4	4	2	75% (3)	75% (3)	75% (3)
63	7	7	0*	57% (4)	14% (1)	14% (1)
64	9	1	N/A	100% (9)	89% (8)	89% (8)
65	1	1	N/A	100% (1)	100% (1)	100% (1)
66	6	1	N/A	50% (3)	50% (3)	33% (2)
67	3	3	1	100% (3)	67% (2)	67% (2)
68	7	7	2	43% (3)	43% (3)	43% (3)
69	4	2	1	100% (4)	100% (4)	100% (4)
70	1	1	N/A	100% (1)	100% (1)	100% (1)
71	5	5	1	40% (2)	100% (5)	40% (2)
72	9	3	0*	44% (4)	33% (3)	33% (3)
73	3	3	0*	67% (2)	33% (1)	33% (1)
74	60**	48	26	76% (38)	60% (30)	58% (29)

0* There are no accurate & complete answers before “best answer” appears.

N/A -- The “best answer” is the first answer so no other answers can appear before it.

**File included only the first 50 answers. All calculations were figured using the 50 answers.

Table 8. Location of first accurate answer (resolved sample)

Answer location in transaction	**First**	**Second**	**Third**	**Fourth**	**Fifth**	**Sixth**	**Seventh**	**Eighth**	**Ninth**
First accurate answer	50	13	5	3	0	0	0	0	1
%	67.57	17.57	6.76	4.05	0	0	0	0	1.35
Cumulative %	67.57	85.14	91.9	95.95	95.95	95.95	95.95	95.95	97.3

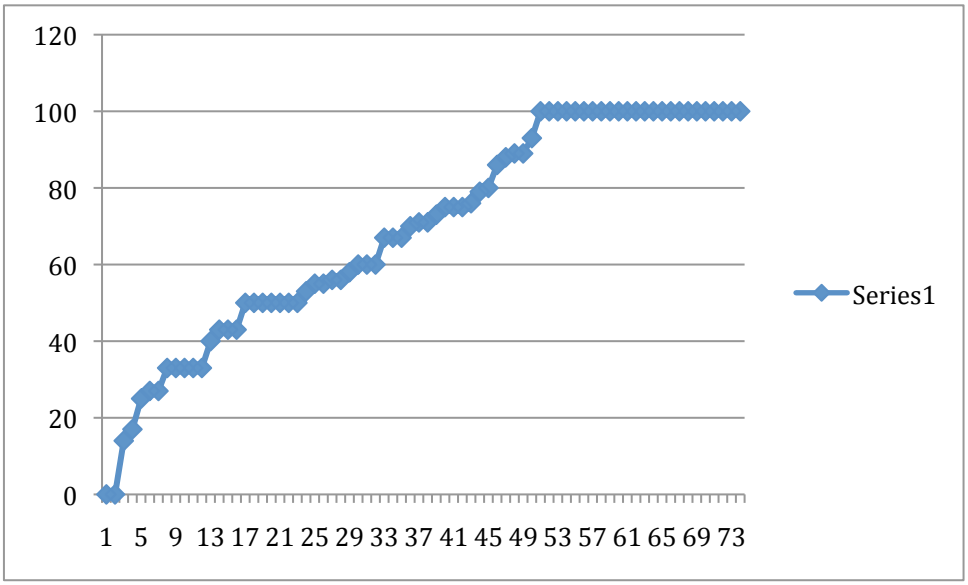

Figure 3. Percent of accurate and complete answers per transaction (resolved sample)

Next, the number and percentage of accurate answers and complete answers per transaction were marked, based on analysis of all question-answer pairs (Table 7). One hundred percent means that all the answers in a transaction are accurate and complete. Figure 3 illustrates the percentage distribution of accurate and complete answers in the resolved sample. It is interesting to note that it is most frequently the case that, for each transaction in the resolved sample, 100% of answers are accurate and complete. This occurs in 30% of the transactions. Although another 30% of the transactions have 50% or lower rates of accurate and complete answers per transaction, it remains that half of the transactions have 70% or more accurate and complete answers per transaction.

Thus, the findings indicate that seven answers would yield good answers, but that most of the time two answers are good enough.

To sum up, the findings indicate that: 1) answer multiplication, with or without the "best answer" feature, results in higher answer quality than the first answer; 2) for over 80% of the transactions, seven answers are enough to get good answers.

5. Discussion

Questioning the almost unquestioned belief that "given enough eyeballs, all bugs are shallow," could be done through multiple lenses; philosophically, statistically, or empirically. This study treats empirically a variation of this belief and postulates that 'given enough answers, all questions are answered successfully'. It defines success as measured by answer accuracy and completeness. While probabilistically the argument that 'given enough answers, all questions are answered successfully' is a sound argument, in reality it can take an endless number of users (or answers) and a long time. The study determines that the number of answers that are needed for 80% of the questions to be answered successfully is the optimum.[5] Under these conditions, the findings show that seven answers are enough to yield good answers for over 80% the questions, that answer quality improved with additional answers, and that there was no evidence of a number of answers, after which additional answers reduce quality.

As such, the findings do not provide evidence to support Brooks' Law, which is a competing theory used in the context of FOSS. Brooks' Law argues that there might be an optimum number of people involved in one successful project, but that adding more participants after reaching that point may hinder performance [14]. According to Brooks' Law, Linus' Law may not hold up ad infinitum. In the context of QA sites, it could mean that answers produced by large groups may be of a lower quality than those of smaller groups. In fact, resembling the inverse relationship between incentives to contribute and group size [55], despite the fact the Yahoo! Answers is the most popular QA site (it has more users and questions than other QA sites), and that a question on Yahoo! Answer gets more answers on average than the Wikipedia Reference Desk, for example, answer quality on Yahoo! Answers was lower than that of the Wikipedia Reference Desk [9]. The findings of the present study, focusing only on Yahoo! Answers, show that additional answers only improved answer quality. Thus it can be concluded that the findings are in alignment with prior FOSS research that found evidence in support of Linus's Law, rather than Brooks' Law [4]. Schweik et al. [4] found that adding more developers improves the chances that the FOSS project will be successful, but they caution that the correlation between size and success does not necessarily mean that bigger groups produce better software, and that it is likewise possible that successful projects attract more contributors. Schweik et al. [4] claimed that size is only one factor that may contribute to the success of the FOSS project and, because they did not conduct multivariate analysis, it is possible that other factors could well serve as competing explanations. Similarly, Meneely and Williams [48] found empirical support for Linus' Law, but they also found some support for

[5] Following the Pareto principal for optimization, eighty percent is a normative benchmark.

Brooks' Law and argue that their findings "do not necessarily negate Linus' Law … [but that] they are a legitimate opposing force" [56, p. 460].

In addition to the support for Linus' Law the findings of the study also show that answer multiplication leads to quality improvement (better accuracy and completeness of answers). There was no evidence in prior research that supports this assumption, yet, under the assumption that the crowd can produce good answers, scholars have made efforts to describe and understand the process of social question-answering [18, 20, 24]. In order to provide empirical support for this assumption, data was analyzed and compared at two levels of analysis; this comparison is essential when looking into the benefits of answer multiplication. While quality improvement was evident for all three variables (verifiability, accuracy, and completeness), it was only statistically significant for two of them, accuracy and completeness. Answer verifiability was very low at both levels of analysis, for whole answers, "best answers", and first answers, echoing prior research. For example, only 8% of answers on Yahoo! Answers include details of the source from which the information was taken [57] and only one out of ten messages on the Wikipedia Reference Desk includes references [25]. This dimension of answer quality is perceived to be very important by Yahoo! Answers' users [23]. Librarians, when answering user's questions in email and chat reach a higher level of verifiability (53%) [58]. Low verifiability levels not only correspond with, but also give rise to, concerns about information quality and the lack of authority on the social Web.

It is important to note that significant improvement in answer quality, in terms of accuracy and completeness, was also associated with "best answers." While "best answers" are individual answers, their quality was equal to that of whole answers. This may be due to the selection process of "best answers", which involves an additional step of information processing that includes feedback about the quality of the answer in light of the question. In fact, in two third of the transactions with "best answers" the choice of a "best answer" was a result of a community vote (48 "best answers" have been chosen by the community and 25 have been chosen by the asker).

6. Conclusion

This study shows that answer multiplication and user's choice of best answers on QA sites significantly improves answer quality in terms of accuracy and completeness when compared with an individual (first) answer. However, the collaborative process did not produce a significant change in the (low) levels of answer verifiability. In support of the argument that given enough answers all questions are answered successfully, the findings reveal that it takes seven answers to achieve a 95% accuracy level for resolved transactions and, on average, seven answers to achieve an 89% accuracy level for transactions that have not yet been resolved; in more than 85% of transactions, one of the first two answers provide an accurate response to the question.

The present study addresses a gap in existing frameworks of answer quality, by focusing on optimization. It also links answer quality to number of users and tests the underlying assumption of these frameworks, that multiple answers improve answer quality. Organizations that consider the implementation of web 2.0 tools, such as QA systems to handle costumer support, can potentially increase their effectiveness while reducing costs. Systems should be designed to enable multiple answer, to allow answer rating, and for quality assurance to generate reports of questions that have not been answered, those that received less than seven answers, or that their answers have not been rated, to be then processed by expert. That way, cost reduction while maintaining high quality can be achieved.

7. References

1. Briggs, R.O., Nunamaker, J., and Sprague, R. Introduction to the Special Section: Social Aspects of Sociotechnical Systems. *Journal of Management Information Systems* 2010; 27, 1: 13-16.
2. Ballou, D., Madnick, S., and Wang, R. Special Section: Assuring Information Quality. *Journal of Management Information & Systems* 2003-4; 20, 3: 9-11.
3. Nelson, R.R., Todd, P.A., and Wixom, B.H. Antecedents of Information and System Quality: aAn Empirical Examination within the Context of Data Warehousing. *Journal of Management Information Systems* 2005; 21, 4: 199-235.
4. Schweik, C.M., English, R.C., Kisting, M., and Haire, S. Brooks' versus Linus' Law: An Empirical Test of Open sSource Projects. In *Proceedings of the 2008 International Conference on Digital Government Research*. Montreal, Canada: Digital Government Society of North America, ACM, 2008, pp. 423-424.
5. Howe, J. The Rise of Crowdsourcing. *Wired* 2006; 14, 6: n.p. (available at http://www.wired.com/wired/archive/14.06/crowds.html).
6. Howe, J. *Crowdsourcing*. New York: Crown Publishing Group, 2008.
7. Leimeister, J.M., Huber, M., Bretschneider, U., and Krcmar, H. Leveraging Crowdsourcing: Activation-Supporting Components for IT-Based Ideas Competition. *Journal of Management Information Systems,* 2009; 26, 1: 197-224.
8. Giles, J. Internet Encyclopedias Go Head to Head. *Nature* 2005; 438: 900-901 (available at http://www.nature.com/news/2005/051212/full/438900a.html).
9. Fichman, P. A Comparative Assessment of Answer Quality on Four Question Answering Sites. Journal of *Information Science* in press.
10. Keen, E. *The Cult of the Amateur: How Today's Internet is Killing Our Culture.* New York: Doubleday/Currency, 2008.
11. Weinberger, D. *Everything is Miscellaneous: The Power of the New Digital Disorder*. New York: Henry Holt & Co., 2007.
12. Surowiecki, J. *The Wisdom of Crowds*. New York: Anchor Books, 2004.
13. Raymond, E. The cathedral and the bazaar. *Knowledge, Technology & Policy* 1999; 12, 3: 23-49.
14. Brooks, F.P., Jr. *The Mythical Man-Month: Essays on Software Engineering.* Reading, MA: Addison-Wesley Publishing Company, 1975.

15.Noguchi, Y. Web Searches Go Low-Tech: You Ask, a Person Answers. *Washington Post* 2006: p. A01. (available at http://www.washingtonpost.com/wp-dyn/content/article/2006/08/15/AR2006081501142.htm).
16.Yahoo Answers Hits 200 Million Visitors Worldwide! *Yahoo Answers Blog.* Yahoo, 2009 (available at http://yanswersblog.com/index.php/archives/2009/12/14/yahoo-answers-hits-200-million-visitors-worldwide/).
17.Harper, F.M., Raban, D., Rafaeli, S., and Konstan, J.A. Predictors of Answer Quality in Online Q&A Sites. In *Proceedings of the Conference on Human Factors in Computing Systems*. Florence, Italy: ACM, 2008, pp. 865-874.
18.Shachaf, P. Social Reference: A Unifying Theory. *Library & Information Science Research* 2010; 32, 1: 66-76.
19.Agichtein, E., Castillo, C., Donato, D., Gionides, A., and Mishne, G. Finding High-Quality Content in Social Media. In *Proceedings of the International Conference on Web Search and Web Data Mining*. Palo Alto: ACM, 2008, pp. 183-194.
20.Gazan, R. Microcollaborations in a Social Q&A Community. *Information Processing & Management* 2010; 46, 6: 693-702.
21.Harper, F.M., Weinberg, J., Logie, J., and Konstan, J.A. Question Types in Social Q&A Sites. *First Monday* 2010:15, 7: n.p. (available at http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/2913/2571).
22.Kim, S. and Oh, S. Users' Relevance Criteria for Evaluating Answers in Social Q&A Site. *Journal of the American Society for Information Science and Technology* 2009; 60, 4: 716-727.
23.Kim, S. Questioners' Credibility Judgments of Answers in a Social Question and Answer Site. *Information Research* 2010; 15, 2: paper 432 (available at http://InformationR.net/ir/15-2/paper432.html).
24.Rosenbaum, H., and Shachaf, P. A Structuration Approach to Online Communities of Practice: The Case of Q&A Communities. *Journal of the American Society for Information Science and Technology* 2010; 61, 9: 1933-1944.
25.Shachaf, P. The Paradox of Expertise: Is the Wikipedia Reference Desk as Good as your Library? *Journal of Documentation* 2009; 65, 6: 977-963.
26.Gazan, R. Specialists and Synthesists in a Question Answering Community. In *Proceedings of the American Society for Information Science & Technology Annual Meeting.* Austin: ASIST, 2006, pp. 1-10.
27.Gazan, R. Seekers, Sloths and Social Reference: Homework Questions Submitted to a Question-Answering Community. *New Review of Hypermedia & Multimedia* 2007; 13, 2: 239-248.
28.Nam, K.K., Ackerman, M.S., and Adamic, L.A. Questions in, Knowledge in?: A Study of Naver's Question Answering Community. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems*. Boston: ACM, 2009, pp. 779-788.
29.O'Neill, N. Chacha, Yahoo!, and Amazon. *Searcher* 2007; 15, 4: 7-11.
30.Saxton, M.L., and Richardson, J.V. *Understanding Reference Transactions: Transforming an Art into a Science*. San Diego: Academic Press, 2002.
31. DeLone, W.H., and McLean, E.R. The DeLone and McLean Model of Information Systems Success: A Ten-Year Update, *Journal of Management Information Systems* 2003; 19(4): 9-30.
32. Rieh, S.Y. Judgment of Information Quality and Cognitive Authority in the Web, *Journal of the American Society for Information Science and Technology* 2002; 53(2): 145–161.
33. Fallis, D. On Verifying the Accuracy of Information: Philosophical Perspectives, *Library Trends* 2004; 52(3): 463–487.
34. Frické, M., and Fallis, D. Indicators of Accuracy for Answers to Ready Reference Questions on the Internet, *Journal of the American Society for Information Science and Technology* 2004; 55(3): 238–245.
35. Arazy, O., Nov, O., Patterson, R., and Yeo, L. Information Quality in Wikipedia: The Effects of Group Composition and Task Conflict, *Journal of Management Information Systems* 2011; 27(4): 71–98.
36. Stvilia, B., Twidale, M.D., Smith, L.C., and Gasser, L. Information Quality Work Organization in Wikipedia, *Journal of the American Society for Information Science and Technology* 2008; 59(6): 983–1001.
37. Blooma, J.M., Chua, A.Y.K., and Goh, D.H. A Predictive Framework for Retrieving the Best Answer. In: *Proceedings of the 2008 ACM Symposium on Applied Computing*, (Fortaleza, Ceara, Brazil, ACM, 2008).
38. Adamic, L.A., Zhang, J., Bakshy, E. and Ackerman, M.S. Knowledge Sharing and Yahoo! Answers: Everyone Knows Something. In: *Proceedings of the International World Wide Web Conference*, (Beijing, ACM, 2008).
39.Poston, R., and Speier, C. Effective Use of Knowledge Management Systems: A Process Model of Content Ratings and Credibility Indicators. *MIS Quarterly* 2005; 29, 2: 221-244.
40.Bouguessa, M., Dumoulin, B., and Wang, S. Identifying Authoritative Actors in Question-Answering Forums: The Case of Yahoo! Answers. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Las Vegas: ACM, 2009, pp. 866-874.
41.Jurczyk, P., and Agichtein, E. Discovering Authorities in Question Answer Communities by Using Link Analysis. In *Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management*. New York: ACM, 2007a, pp. 919-922.
42.Jurczyk, P., and Agichtein, E. Hits on Question Answer Portals: Exploration of Link Analysis for Author Ranking. In *Annual ACM Conference on Research and Development in Information Retrieval*. Amsterdam, ACM, 2007b, pp. 845-846.
43.Chen, W., Zeng, Q., and Wenyin, L. A User Reputation Model for a User-Interactive Question Answering System. In *Proceedings of the Second International Conference on Semantics, Knowledge, and Grid*. Washington D.C.: IEEE Computer Society, 2006, pp. 40-45.
44. Adamic, L.A., Wei, X., et al. Individual Focus and Knowledge Contribution, *First Monday* 2010; 5(3).

45.Dom, B., and Paranjpe, D. A Bayesian Technique for Estimating the Credibility of Question Answerers. In *Proceedings of the Society for Industrial and Applied Mathematics (SIAM).* Atlanta: SIAM, 2008, pp. 399-409 (available at http://www.siam.org/proceedings/datamining/2008/dm08_36_Dom.pdf).
46.Ong, C., Day, M., and Hsu, M. The Measurement of User Satisfaction with Question Answering Systems. *Information & Management* 2009; 46, 7: 397-403.
47.Harper, F.M., Moy, D., and Konstan, J.A. Facts or Friends?: Distinguishing Informational and Conversational Questions in Social Q&A Sites. In *Conference on Human Factors in Computing Systems*. Boston: ACM, 2009, pp. 759-768.
48.Li, B., Liu, Y., Ram, A., Garcia, E.V., and Agichtein, E. Exploring Question Subjectivity Prediction in Community QA. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. Singapore: ACM, 2009, pp. 735-736.
49. Liu, Y., Li, S., Cao, Y., et al. Understanding and Summarizing Answers in Community-Based Question Answering Services. In *Proceedings of the 22nd International Conference on Computational Linguistics.* Manchester, UK: ACL, 2008, pp. 497-504.
50. Hitwise. U.S. Visits to Question and Answer Websites Increased 118 Percent Year-Over-Year. New York: Hitwise, March 19, 2008, n.p. (available at http://www.hitwise.com/news/us200803.html).
51.Neuendorf, K.A. *The Content Analysis Guidebook*. Thousand Oaks, CA: Sage, 2002.
52.Lombard, M., Snyder-Duch, J., and Bracken, C.C. Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability. *Human Communication Research* 200; 28, 4: 587–604.
53.Krippendorff, K. *Content Analysis: An Introduction to its Methodology.* 2nd ed. Thousand Oaks, CA: Sage, 2004.
54.Landis, J.R., and Koch, G.G. An Application of Hierarchical Kappa-Type Statistics in the Assessment of Majority Agreement among Multiple Observers. *Biometrics* 1977; 33, 2: 363-374.
55.Zhang, X., and Feng, Z. Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia. *American Economic Review* 2011; 101, 4: 1601–1615
56.Meneely, A., and Williams, L. Secure Open Source Collaboration: An Empirical Study of Linus' Law. In *Proceedings of the 16th ACM Conference on Computer and Communications Security.* New York: ACM, 2009, pp. 453-462.
57.Oh, S., Oh, J.S., and Shah, C. The Use of Information Sources by Internet Users in Answering Questions. In *Proceedings of the Annual meeting of the American Society for Information Science and Technology.* Columbus, OH: ASIST, 2008, pp. 1-13.
58.Shachaf, P., and Shaw, D. Bibliometric Analysis of Virtual Reference Transaction Sources. *Library & Information Science Research* 2008; 30, 4: 291-297.