

# Analyzing Crowd Workers in Mobile Pay-for-Answer Q&A

Uichin Lee, Jihyoung Kim<sup>†</sup>, Eunhee Yi<sup>‡</sup>, Juyup Sung, Mario Gerla<sup>†</sup>

KAIST Knowledge Service Engineering, <sup>†</sup>UCLA Computer Science, <sup>‡</sup>LG UX R&D Lab  
 {uclee, juyup.sung}@kaist.ac.kr, {jhhkim, gerla}@cs.ucla.edu, eunhee2.yi@lge.com

## ABSTRACT

Despite the popularity of mobile pay-for-answer Q&A services, little is known about the people who answer questions on these services. In this paper we examine 18.8 million question and answer pairs from Jisiklog, the largest mobile pay-for-answer Q&A service in Korea, and the results of a complementary survey study of 245 Jisiklog workers. The data are used to investigate key motivators of participation, working strategies of experienced users, and longitudinal interaction dynamics. We find that answerers are rarely motivated by social factors but are motivated by financial incentives and intrinsic motives. Additionally, although answers are provided quickly, an answerer's topic selection tends to be broad, with experienced workers employing unique strategies to answer questions and judge relevance. Finally, analysis of longitudinal working patterns and community dynamics demonstrate the robustness of mobile pay-for-answer Q&A. These findings have significant implications on the design of mobile pay-for-answer Q&A.

## Author Keywords

Mobile Pay-for-Answer Q&A; User Behavior

## ACM Classification Keywords

H.5.3 Information Interfaces and Presentation (e.g., HCI): Group and Organizational Interfaces

## INTRODUCTION

Mobile devices have significantly changed the way we seek information. People can now search large repositories like the web from anywhere. Additionally, mobile Q&A services like Naver Mobile Q&A, ChaCha, and Jisiklog make it possible to quickly and easily find information by asking questions via instant messaging or SMS. This paper focuses on mobile Q&A, which has experienced significant growth in recent years. For instance, as of December 2012, ChaCha answered more than 4.5 billion questions, far surpassing the number of questions answered in Yahoo! Answers. When compared with traditional web-based Q&A, mobile Q&A users tend to

ask a broad range of questions attributed to everyday life situations [17]. For example, users might use text messaging to check a bus schedule or asking for quick restaurant suggestions. The key drivers of mobile Q&A usage include accessibility, availability, and promptness [17].

Most popular mobile Q&A sites are based on pay-for-answer services where answerers receive monetary incentives either from the askers (e.g., Jisiklog, AQA, KGB Answers) or their employers (e.g., ChaCha). Unlike web-based pay-for-answer Q&A where askers can offer as much as they like to pay (e.g., Google Answers: \$2–\$200), in mobile pay-for-answer Q&A, payment is typically a fixed rate per task. For instance, ChaCha workers are paid \$0.10–\$0.20 for an answer based on answerer level and \$0.01 for vetting answers. Jisiklog workers receive 80 KRW (approximately \$0.08) for answering and 20 KRW (approximately \$0.02) for vetting answers.

Existing mobile pay-for-answer Q&A services are unique in that they provide a structured workflow, breaking question answering into real-time micro-tasks to generate answers, and perform quality control using piece-rate financial incentives. Mobile Q&A services also provide limited opportunity for interaction between answerers and askers, as text input and output is challenging on mobile phones, and messages are often length limited and expensive. Mobile pay-for-answer Q&A services have evolved over many years (with ChaCha and Jisiklog established in 2006) to become real-time knowledge marketplaces for mobile information seekers.

Our goal in this paper is to investigate the behaviors and strategies of crowd workers who answer questions in mobile pay-for-answer Q&A. While this has been investigated extensively in existing web-based Q&A sites, little is known about the mobile pay-for-answer Q&A. We analyze a longitudinal Q&A dataset from Jisiklog spanning over 60 months and the results of a supplementary online survey study of 245 Jisiklog answerers. To the best of our knowledge, this work is the first large scale study on mobile pay-for-answer Q&A.

From the analysis, we find that while previous work on traditional Q&A sites has shown that social factors are important motivators [21, 22, 7], in mobile pay-for-answer Q&A sites answerers are instead mostly motivated by financial and intrinsic factors. Additionally, unlike traditional Q&A sites where users mostly focus on a few topics of interest [19, 3], mobile Q&A answerers tend to respond to a broader range of topics. Response times are fast, with experienced users answering particularly quickly and using unique strategies. Experienced workers also invest a significant amount of time daily, dispersed over multiple work periods. We believe that this understanding of the behavior and strategies of answer-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2013, April 27–May 2, 2013, Paris, France.

Copyright 2013 ACM 978-1-4503-1899-0/13/04...\$15.00.

ers can be used to improve existing mobile Q&A system design, and, more broadly, real-time crowdsourcing systems that leverage the wisdom of crowds.

### RELATED WORK

We begin our discussion of mobile pay-for-answers Q&A by giving an overview of existing related research studying user interactions in social Q&A, pay-for-answer Q&A, and online labor markets.

#### User Interactions in Social Q&A

Social Q&A sites like Yahoo! Answers and Naver KiN use the wisdom of crowds to answer questions. Adamic et al. [3] studied the asker-answerer social network on Yahoo! Answers and users' interests across various topic categories. They found that participation is highly skewed, with answerers tending to focus on just a few categories. Nam et al. [19] found a similar trend in Naver KiN; answerers' participation on Naver KiN is intermittent, and their expertise lower than is often found in specialized online forums [19]. This trend persists in mobile Q&A [17]. The time it takes for questions to be answered on social Q&A sites tends to be fairly long. Hsieh et al. [13] showed that in Microsoft Live Q&A, the average time was almost 3 hours, with 20% of questions never even receiving an answer. The average time in Naver's mobile Q&A was 15.5 minutes, and 34% of the questions were not answered; unlike traditional social Q&A, mobile Q&A users often repeat or refine their information needs over a series of questions as coping strategies [17]. Yang et al. [26] studied two-year-long Q&A datasets from multiple social Q&A sites to understand user lifespan and community stability. The lifespan distribution shows that a large fraction of users leave (30%–70%) after a single post, and only a small fraction of users stay longer than 200 days (3%–10%).

How long people remain active in a social Q&A system varies by their role, with answerers staying longer than askers [26]. People's motivation for answering can be classified into intrinsic and extrinsic factors [20]. Intrinsic motivations include information ownership, perceived benefits (e.g., enjoyment, altruism, learning), and individual attitudes [19, 20, 23]. Extrinsic motivations include social recognition (e.g., points/levels, leaderboards), social norms (e.g., interest votes), and financial incentives. Like existing research in social Q&A, in this paper we look at the motivations of answerers, the topics of questions that a user chooses to answer, answering delay, answering strategies, and community stability. We do so specifically within a system that provides answers in return for financial compensation.

#### User Interactions in Pay-for-Answer Q&A

Prior work to understand user interaction on pay-for-answer Q&A sites has focused on web-based pay-for-answer Q&A sites such as Mahalo Answers [14], Uclue, and Google Answers (discontinued in December 2006) [7, 22, 20]. Research from psychology and behavioral economics shows that monetary incentives can crowd-out intrinsic and social motivators [10]. In reality, given that individual motivations are diverse, only a subset of the population may be motivated by financial rewards, and the crowding out effect may not be

critical. For instance, Raban et al. [21] analyzed a Google Answers dataset and showed that social features (e.g., social approval and observational cooperation) have a positive impact on the number of answers per user. Even with monetary incentives, researchers found that, similar to free systems, the level of participation in Google Answers followed a power-law, with a small fraction of users answering the majority of questions [21, 7], and most askers posting only a few questions [22]. The median delay for receiving answers on Google Answers was less than three hours which is comparable to online social Q&A; fewer questions were posted during the weekends, and questions posted between 8-10PM received the fastest responses [7]. Since the interaction environments of mobile Q&A are very different from those of existing web-based pay-for-answer Q&A (e.g., real-time, via instant messaging or SMS, piece-rate payment), we are interested in characterizing interaction patterns of answerers and their motivations by studying the Jisiklog dataset and the survey results of Jisiklog workers.

Prior work on the impact of financial incentives on answers has arrived at conflicting conclusions. Harper et al. [9] showed that pay-for-answer sites elicit better answers than free sites (e.g., Google Answers vs. Yahoo! Answers), and that higher quality answers can be acquired by paying more money. In contrast, Chen et al. [6] showed that paying more only elicits longer answers, and not better answers. Hsieh et al. [14] analyzed Mahalo, a pay-for-answer Q&A site, and showed that askers wish to pay when requesting facts and will pay more when asking difficult questions.

#### User Interactions in Online Labor Marketplaces

Pay-for-answer Q&A can be considered as a special form of online labor marketplaces in which people perform a variety of human intelligence tasks, ranging from simple image labeling and transcription to complex information seeking and graphic design. In recent years, several popular online labor marketplaces such as Amazon Mechanical Turk (M-Turk) and Taskcn have been extensively investigated [15, 25, 24]. In their analysis of M-Turk, Ipeirotis [15] found that the number of posted tasks per requester follows a power-law distribution. More than 90% of tasks paid less than \$0.10, and the median delay for task completion was about 16 hours. This level of payment is quite comparable to that in mobile pay-for-answer Q&A services, but, unfortunately, M-Turk does not reveal any information from which we can analyze the longitudinal behavior of workers. Recently, Yang et al. [25, 24] analyzed Taskcn, an online marketplace for specific task categories requiring some level of technical skills (i.e., graphic design, web site design, writing, and programming). Unlike M-Turk, a requester can receive submissions from multiple users and then select the best submission to which the bounty is awarded. In Taskcn, the level of participation is highly skewed such that the majority of users submitted or won a very small number of tasks [25, 24].

In a study of M-Turk, Mason and Watts [18] found that paying more will increase the quantity, but not the quality of work. As the rational choice theory in economics informs, experienced workers in Taskcn often choose to improve their per-

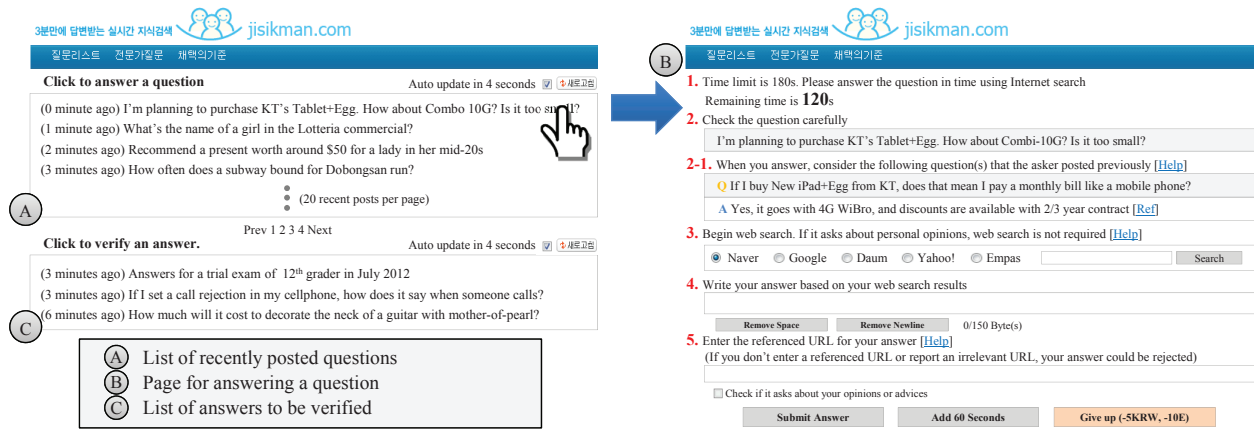


Figure 1. Main page (left) that lists newly posted questions (A) and answers to be verified (C); and an answer page (B) with a five-step guide for answering (URL: <http://answer.jisikman.com>). Here, the text was translated into English for illustration

formance with financial gain (e.g., selecting less competitive tasks) [25]. According to the survey results of M-Turk workers [15], about 40% of US workers expressed that fun and enjoyment is a key motivating factors. It is less clear how the intrinsic and social factors of knowledge sharing (e.g., altruism, learning) co-exist in a knowledge marketplace with piece-rate wage; identifying these factors is one of the primary goals of this work. Since the amount of work and the level of payment in mobile pay-for-answer Q&A are comparable to those in M-Turk, some of our findings may provide valuable insights into understanding the dynamics in online labor marketplaces.

**USER INTERACTIONS IN JISIKLOG**

We study the online labor marketplace of mobile pay-for-answer Q&A by looking at the Jisiklog system. Jisiklog was first launched for SKT mobile subscribers in October 2006, and the service was opened to other mobile service providers (e.g., KT, LG) in April 2007 [1]. The Q&A data are publicly available, and major search engines including Google and Naver have been using the dataset for Q&A search.

Askers post questions by sending SMS messages. As shown in Figure 1, an answerer (also called a *jisikman*) can browse a list of recently posted questions (top-left, A) and the answers that need to be verified (bottom-left, B). There are two answerer classes, namely amateur and expert jisikmen. Experts’ answers are immediately delivered to askers, while amateurs’ answers are first verified by an expert (see a simplified workflow chart in Figure 2). Answers by amateurs are routed to the verification section (located at the left-bottom, C), and experts can either accept or reject these answers after review. An asker pays 200 KRW (approximately \$0.20) for an answer; a jisikman receives 80 KRW (approximately \$0.08) for answering a question and 20 KRW (approximately \$0.02) verifying an answer.

To preserve answer quality (both for amateurs and experts), Jisiklog runs an answer review board. Any members whose level is at least 10 with a weekly acceptance rate of 80% or higher can file an objection to a posted answer. If the objection is sustained (determined by a majority vote by the expert jisikman; 5 KRW per vote), 100 KRW (\$0.10) and 200 points

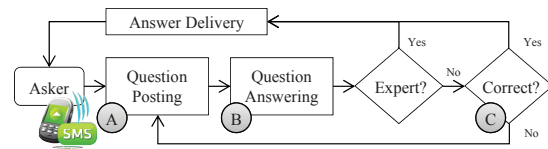


Figure 2. Simplified workflow of Jisiklog (the numbers correspond to the web pages in Figure 1)

deducted from the answerer are awarded to the objector. Conversely, if overruled, 165 KRW (\$0.165) and 330 points deducted from the objector are awarded to the answerer. Note that askers can also file an objection for unsatisfactory answers, and they receive a voucher for a free answer if sustained. Workers can transfer the earned money to their bank account, and a unit of transfer is 10,000 KRW (approximately \$10.00).

Along with monetary incentives, Jisiklog also has a leveling system based on a jisikman’s earned points. Workers earn 100 points for each answer, and there are other activities and events from which they can either earn or lose points [1]. Jisiklog has a conversion table that maps points to a level. The conversion is roughly exponential ( $R^2 = 0.997$ ), with users needing to earn approximately  $\exp(5.25 + 129\ell)$  to increment a level for a given level  $\ell$ . An amateur jisikman whose level is above 20 ( $\geq 18,685$  points) can become an expert jisikman if, in a given week, that jisikman answers at least 10 questions, 90% of their answers are verified, and no objections are received.

The page for creating an answer contains a five-step guide (see Figure 1, B). In Step 1, the answerer is informed that the question needs to be answered in 3 minutes; the remaining time is counted down in the Step 1 slot. Step 2 displays the current question, and the user is asked to consider several recent questions posed by the asker when answering. Step 3 provides a list of popular search engine links, asking an answerer to search the Internet. In Step 4, the answerer fills in an answer (minimum length of 50 bytes, maximum length of 150 bytes), and in Step 5, the answerer enters a referenced URL based on which the answer was made. The worker can skip this step if the asker is seeking an opinion or advice. If

	Fruitful*	\$Major	\$Minor***	KillTime	Help	Fun	Learn	LevelUp	Accept	Expert	Rank	N
Exp.	53.4%	4.2%	77.1%	14.4%	16.9%	26.3%	30.5%	3.4%	3.4%	5.1%	2.5%	118
Novice	40.3%	1.6%	57.3%	16.9%	16.1%	33.9%	21.8%	4.0%	6.5%	3.2%	0.8%	124
Overall	46.7%	2.9%	66.9%	15.7%	16.5%	30.2%	26.0%	3.7%	5.0%	4.1%	1.7%	242

**Table 1. Distribution of motivation items, and distribution based on experience level (\*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001)**

more time is needed, an additional minute can be granted by clicking the “Add 60 seconds” button (bottom-right)—a user is given at most 3+1 minutes in total. The answerer can give up by clicking the “give up” button. No penalty is levied for this before 5 seconds, but after 5 KRW and 5 points are deducted, and after 10 seconds, additional 5 points are deducted.

## METHODOLOGY

To learn more about how people answer questions on Jisiklog, we downloaded the entire Q&A dataset from the period of May 1, 2007 to April 22, 2012 (about 60 months), for a total of 18,869,566 question-answer pairs. For a given question-answer pair, we also collected associated metadata such as the answerer ID and posting timestamps of questions.

We supplemented the crawled data with an online survey of Jisiklog answerers. We posted a link to the survey to the jisikman community website on May 18, 2012, and left it available for a month. After removing duplicate and erroneous responses, we received a total of 245 responses. The survey was composed of four parts: (1) demographics (e.g., age, gender, job, computer usage, income, user ID), (2) motivating factors (e.g., financial, altruistic), (3) answering strategies (e.g., topic focus, information resource usage, answering knowhow), and (4) feelings around financial incentives and quality (e.g., appropriateness, perceived quality).

A participant’s user ID was used to identify the respondent in the crawled dataset, and to classify answerers as experienced and novice based on the number of answered questions. The threshold number was set to 200, corresponding with the number of questions required to be above Level 20, the minimum level to become an expert jisikman (although results are consistent across different thresholds).

Most participants were in their 10s (41.8%), 20s (43.9%), and 30s (11.9%), with only a small number above 40 (2.4%). Participants were mostly single (91.8%), and females were dominant (60.0%). Occupations ranged from junior high school students (10.2%), high school students (29.7%), college students (26.5%), graduate students (2.0%), homemakers (2%), engineering (10.2%), service/finance/education (9.3%), and miscellaneous (5.7%); 3.7% were unemployed. Since a large fraction of participants are students, the distribution of monthly income was highly skewed; < 50K (approximately \$500): 68.2%, 50K–100K (approximately \$500–\$1000): 5.7%, 100K–150K (approximately \$1000–\$1500): 7.8%, 150K–200K (approximately \$1500–\$2000): 6.5%, > 200K (approximately \$2000): 11.8%. The average usage time of computers per day was 4.27 hours (SD: 3.44).

We structured the data analysis along the following aspects of user interaction in mobile pay-for-answer Q&A: (1) key motivators of user participation (including intrinsic and extrinsic factors); (2) working strategies of experienced users (including topic selection, information seeking, and answer assess-

ment); and (3) longitudinal interaction dynamics (including work frequency and duration, contribution, and attrition).

## KEY MOTIVATORS

We begin our analysis of the data collected by investigating how the intrinsic and social factors of knowledge sharing (e.g., altruism, learning) co-exist with monetary incentives in Jisiklog. We find that while previous work on traditional Q&A sites has shown that social factors like levels and ranks are important motivators [21, 22, 7], in mobile pay-for-answer Q&A users are rarely motivated by social factors. Instead, they are motivated by intrinsic factors such as altruism, fun, and learning, in addition to the major motive of financial gain.

To explore motivation, we designed the survey questions related to motivation by building on key motivational factors in online labor marketplaces [15] and including additional motivations related to intrinsic and social motivations of social Q&A [19, 21]. The survey contained 11 potential motivation items, and participants were asked to select a set of items that best described their motivations. The following are the motivation items, shown in Table 1: fruitful way of spending free time (Fruitful); for primary income sources (\$Major), for minor income sources or pocket change (\$Minor); to kill time (KillTime); to help other people (Help); I find the tasks to be fun (Fun); to learn things while answering (Learn); to increase the level (LevelUp); to improve acceptance rate (Accept); to become an expert jisikman (Expert); and to be listed on a leaderboard (Rank).

Table 1 shows the distribution of responses. As expected, the strongest motivating factor is monetary incentives (total 69.8%). However, 31.0% of users were not motivated by monetary incentives. Unlike the results of Google Answers by Raban et al. [21], intrinsic motivators were considered more important than social motivators. Intrinsic motivators such as altruism (16.5%), fun (30.2%), and learning (26.0%) were marked by a considerable number of participants. These factors did not show a significant correlation between age groups (teenagers vs. adults). In contrast, social motivators like leveling-up (3.7%), becoming an expert jisikman (4.1%), and making the leaderboard (1.7%) were regarded as less important.

We hypothesize that a combination of several unique characteristics of mobile Q&A such as real-time constraint, workflow design, and piece-rate incentives, influences users to pay more attention to intrinsic factors than social factors. For example, the lack of social motivation may derive from the limited forms of user interactions that are possible within the system (e.g., only one answer, verification by professional jisikmen). Participants expressed a fairly low sense of belonging to the community, asked using a 5-point Likert scale (average 2.08, SD: 0.98).

We further divided participants into experienced and novice answerers. A Chi-square test with Yates' continuity correction revealed that the distribution of motivation is significantly differentiated by experience ( $\chi^2 = 238.5$ ,  $p < 0.001$ , Cramer's  $V = 0.67$ ). The difference comes from two sources, namely fruitful use of free time ( $p < 0.05$ ), and income sources or pocket change ( $p < 0.001$ ). It appears that while experienced users were more motivated by monetary incentives (77.1% vs. 66.1%), these monetary incentives did not crowd-out the intrinsic and social motivation of experienced users. Our findings suggest that even with two market norms, intrinsic and social motivators can co-exist with monetary incentives and play an important role in long-term participation. Moreover, the fraction of participants who did not select monetary incentives as a motivation factor is not statistically different across different groups (28.8% and 33.06% in experienced and novice groups, respectively).

### USER WORKING STRATEGIES

We next look at answerer behavior as it relates to topic selection, answering strategies, and quality assessment. Unlike traditional Q&A sites where users mostly focus on a few topics of interest (e.g., Naver KiN [19], Yahoo! Answers [3]), we see that Jisiklog answerers tended to respond to a much broader range of topics, possibly to increase earnings or learn more. Experienced users tended to answer faster, but there was significant variation in answer speed partly due to diverse skill levels and individual attitudes (e.g., valuing fast responses vs. witty or accurate answers). Overall, the time to response was very fast, which suggests an efficient real-time knowledge marketplace. Experienced answerers were likely to refer to diverse web information sources and to assess answer quality mainly based on accuracy, source quality, and specificity.

### Topic Selection

To understand the variation in topic of the questions answerers selected to answer, we asked answerers in the survey whether they tended to focus on specific topics of interest when selecting questions to answer on a 5-point Likert scale (1: never to 5: always). Participants expressed moderate level of focus on topic selection (average: 2.67, SD: 1.06). For example, one person responded, "*I usually select questions that fit into my major.*" There was no significant difference in topic selection between experienced and novice answerers ( $U = 7360.5$ ,  $p\text{-value} = 0.8043$ ).

Since participants reported a moderate level of topic preference, we decided to quantify the spectrum of their topic preferences in the crawled data. Unlike traditional social Q&A, Jisiklog askers do not specify a topic when they text questions via SMS. To automatically classify topic categories, for a given question we extracted the key words using a Korean parser called KKMA (Kokoma Korean Morpheme Analysis) [2]. We then searched the extracted keywords using Naver KiN, a popular social Q&A service in which askers manually select the relevant topics when posting questions. From the top 10 search results, the most frequent topic category was selected as the topic category for the question; Naver KiN has 13 major topic categories.

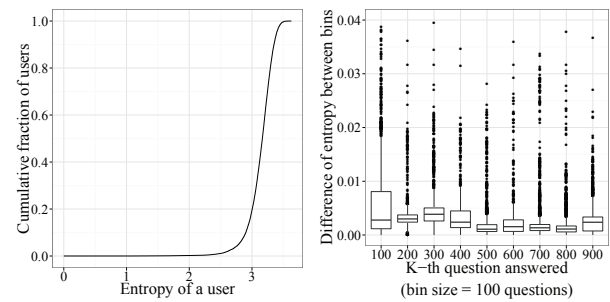


Figure 3. Topic entropy of users

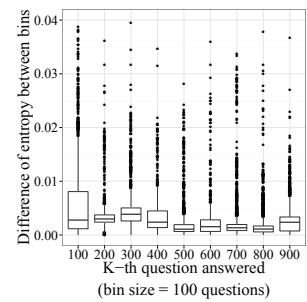


Figure 4. Entropy changes

The classified results are given as follows: computer (2.2%), games (4.6%), entertainment/art (19.3%), economy (3.7%), shopping (5.7%), society/politics (9.2%), health (7%), life (22.7%), travel (2.4%), sports (2.7%), education (10.2%), regional (5.0%), and counseling (6.0%).

To capture the degree of focus on particular topics by an answerer, we use an entropy measure—the lower the entropy, the higher the level of focus on certain topics. Consider an answerer  $i$  who made  $p_k$  fraction of answers on topic  $k$  (for all  $k = 1, \dots, 13$ ). An answerer  $i$ 's entropy is  $-\sum_k p_k \log p_k$ . If a user answered only a single category (say topic  $j$ ,  $p_j = 1$  and  $p_i = 0$  for all  $i = 1, \dots, j - 1, j + 1, \dots, 13$ ), the entropy is zero. Entropy is maximized when any topic is equally likely to be selected by a user (i.e., a uniform distribution). The cumulative distribution of topical entropy for all answerers who posted more than 100 answers can be seen in Figure 3. The shape of the distribution is very narrow and highly skewed, with average entropy of 3.14. To place this value in context, we calculate the entropy of a hypothetical random answerer who randomly selects questions from the list, and find that it is very close to the observed data (3.32). This result is very different from what has been observed in traditional social Q&A, where answerers tend to focus on several topic categories of interest [3, 19].

We also analyze whether experienced answerers' topic entropy changed over time, and found considerable consistency. For those who answered more than 1000 questions, we calculate the entropy difference between two consecutive bins (e.g., 1–100th questions vs. 101–200th questions) and plot the results in Figure 4. The figure clearly shows that a user's topic entropy rarely changes over time (within 0.04).

### Answering Strategies

To better understand how answerers create their responses, we asked participants to report: (1) what effective answering strategies they used (free-text); (2) what kinds of information resources they used when answering (e.g., web search, Jisiklog search, dedicated web sites like maps and weather, personal experience), and how frequently each was used (on a 5-point Likert scale); (3) whether their strategies changed over time (free-text); and (4) whether they found that 3+1 minutes are enough time to answer (on a 5-point Likert scale). Additionally, we analyzed the Q&A dataset to measure the average time it took to answer a question.

To analyze the free-text answering strategies reported by our participants, we performed open coding and identified com-

mon themes. The four main strategies that emerged include: (1) choosing the type of question to answer (e.g., based on expertise or difficulty); (2) effectively using various information resources (e.g., web search, Jisiklog, dedicated web services) based on question content; (3) developing thoughtful and sincere answers; and (4) actively maintaining question-answering resources (e.g., bookmarks, summaries of frequently asked questions and frequently used sentences). Our manual classification with two coders shows that the participants' strategies can be reliably classified (Cohen's kappa: 0.8). The four categories have the following distribution: (1) 15.9%, (2) 67.7%, (3) 16.9%, and (4) 10.0%.

Participants reported being selective about the type of questions they chose to answer 15.9% of the time, with one participant saying, "I wait until there are questions that I would like to answer." As discussed in the previous section, participants sometimes looked for topics of interest or topical diversity. For example, one reported, "Since I'm majoring in chemistry in college, I prefer answering science questions. I usually refer to my textbook and several online forums to answer questions." This may be because these questions are easier for the individual to answer, and ease was a common theme (e.g., "I prefer answering easy questions like TV schedules, restaurant recommendation, SAT answers, etc."). One participant reported actively seeking questions that could be addressed via a particular resource ("I select a topic category that I can answer using Naver web search.").

Regardless of the question selected, most questions asked on Jikilog require outside resources to answer, and resource selection was the most common aspect of question answering mentioned by participants (arising in 67.7% of responses). For example, one participant said, "For study and college entrance related questions, I use search results from Naver KiN and add some personal comments. For direction and health questions, I use Naver Map and Encyclopedia. For computer related questions, I use Naver KiN and Jisiklog. I have a list of dedicated web sites for various kinds of questions." Another said, "For questions on last bus schedules or directions, I usually use Naver Map or Naver Public Transit Info."

When asked directly about what information resources they used, participants answered that web search was the most frequently used (average 4.54, SD: 0.82) with other resources rating below 3.3 on average (SD < 1.25). To check whether the usage of information resources differed on the basis of experience level (experienced vs. novice), we ran a Mann-Whitney's U test. We found that experts were more likely to search using web search (Exp. 4.66 vs. Novice 4.43;  $U = 8538$ ,  $p\text{-value} < 0.05$ ) and Jisiklog search (Exp. 3.37 vs. Novice 2.74;  $U = 9607.5$ ,  $p\text{-value} < 0.001$ ), and were more likely to use dedicated web sites (Exp. 3.54 vs. Novice 3.07;  $U = 9096$ ,  $p\text{-value} < 0.01$ ).

The third common theme emerged among answering strategies related to developing thoughtful and sincere responses, occurring 16.9% of the time. For example, one participant stated, "Instead of simply copying and pasting the answers from a web search, I try my best to explain the answer in detail such that if I were the askers, I could also easily understand

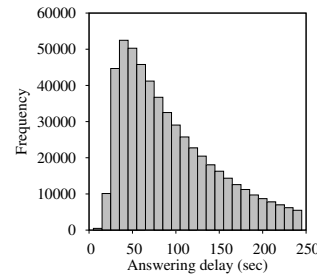


Figure 5. Answer delay dist.

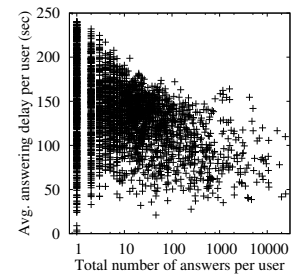


Figure 6. Answer delay per user

the answer." Another said, "For sports related questions, I try to provide as much detail as possible; like the bottom of 8th inning, 1 out, bases loaded, the next batter is Daeho Lee. For counseling questions, I try to analyze the asker's characteristics and give personalized answers as much as possible."

To support question answering 10.0% of participants reported active maintenance of question answering resources, ranging from bookmarks ("I came up with question categories and arranged the bookmarks of frequently visited web sites based on those categories.") to common questions and answers ("I maintain frequently asked questions or expected answers in my memo pad.").

When asked how their answering strategies developed over time, participants gave examples of how they arrived at the above strategies, from question selection ("In the beginning, I had quite a few cases of having difficulties in finding answers in three minutes. As time passed, I gained a sense of what questions I can answer and learned how to utilize various dedicated web sites.") to resource selection ("I used to search mainly web portals, but now I use various web sites simultaneously to answer questions.") to more answer generation ("I changed my strategy from simply copying/pasting answers to providing more reliable/accurate answers."). We also observed that experienced users were more likely than novice users to maintain their own answer resources, as such require time and experience to construct. Overall, as time passes, it appears that users become more skillful at answering questions from the web.

In general, the labor market literature shows that this kind of on-the-job learning plays a significant role in developing worker skills and facilitating worker productivity [16]. Consistent with this, we found that experienced users in Jisiklog felt less time pressure when compared with novice users. When asked whether the time limit for answering (3+1 minutes) was appropriate, participants generally felt that they need more time (average 2.5, SD: 0.94). However, when experienced and novice users were compared, the novice group (average 2.38, SD: 0.91) reported feeling more time pressure than the experienced group (average 2.65, SD: 0.95), as confirmed by a Mann-Whitney's U test ( $U = 9035$ ,  $p\text{-value} < 0.01$ ).

We also looked at the amount of time that answerers invested in answering questions. The crawled dataset only contains the timestamp of question pickup from the main page in Figure 1 and does not have an actual timestamp of answer sub-

mission. However, since newly posted answers are immediately displayed on the peer review board (in which users can file an objection), we can use this information to measure the time it takes to answer a question. To this end, we monitored the board for 7 months, from November 1, 2011 to April 31, 2012. The total number of downloaded questions was 529,581, and the average answering delay was 88.6 seconds (SD: 54.5). The delay histogram is depicted in Figure 5. The shape of the distribution is quite smooth with a near linear tail (cut-off value of 4 minutes).

We also looked at whether the time it took to generate an answer was dependent on the total number of answers given by a user, and found experienced answerers required less time. Figure 6 shows a scatter plot of all the users. The overall delay dispersion of novice users (less than 200 questions) varies widely, whereas the average delay of experienced users is mostly below 150 seconds. Nonetheless, the average answering delay of experienced individuals still sometimes differed significantly. As shown in the figure, the following are the average delays of top seven users (located at the bottom right): 45.0 (SD: 27.2), 109.9 (SD: 50.4), 41.6 (SD: 30.3), 78.0 (SD: 45.7), 116.6 (SD: 50.3), 52.7 (SD: 33.5), 105.6 (SD: 51.7).

Two of the top seven users participated in the survey, and provided comments about their answering strategies. One reported, “*Speed is the most important thing; first pickup a question and answer as soon as possible.*” The other said, “*For direction questions I usually get directions from maps, but I would prefer giving more detailed answers based on my experience. I’m quite knowledgeable with directions. People frequently ask about song titles, but to improve satisfaction of askers I try to include witty remarks.*” These individuals have different perspectives on giving answers, with one valuing speed, and the other satisfaction. This in turn leads to very different answering delay (41.6 seconds versus 105.6 seconds). These findings are consistent with Wang et al.’s [23] observation that individual attitudes are important factors of information sharing.

### Assessing Answer Quality

There are two ways answer quality is assessed on Jisiklog; pending answers are verified by an expert jisikman, and objections can be filed to delivered answers. In the survey, we asked open-ended questions on the criteria participants used to assess answers for these tasks. We used criteria proposed by Barry and Schamber [4], including: (1) depth, scope, and specificity, (2) accuracy and validity, (3) clarity, (4) currency (up-to-date), (5) tangibility (proven with real data or numbers), (6) source quality (reputable, trusted, accurate, expert), (7) accessibility (costs of access), and (8) availability. For verification (total 87) and objection (total 67), two raters classified the responses. The reliability of the ratings in each category was checked with Cohen’s kappa; most of our ratings had a kappa value greater than 0.7, meaning substantial agreements. For verification, the primary factors that participants consider important include accuracy and validity (59.8%), source quality (47.1%), and depth, scope, and specificity (36.8%). For example, one participant responded, “*1. Correctly answered, 2. Used correct reference, 3. Correctly*

*understood the user’s needs, 4. Sincerely gave answers. Also, check whether the answer is too short or content in the reference page is well reflected.*” Another said, “*I consider trustworthiness (accuracy) of an answer most important. Askers are paying money for answers. If accuracy is low, it is hard to differentiate Jisiklog from other free services on major portals like Naver KiN.*” The same criteria were important when filing an objection.

The remaining criteria for assessment were rarely mentioned, including tangibility (3.4%), clarity (2.3%), currency (2.3%), availability (2.3%), and accessibility (0.0%). While currency is mentioned by few participants, it is closely related to accuracy and validity, and its importance may have discounted. In fact, quite a few questions in Jisiklog are time sensitive, including questions for sports scores, TV schedules, and traffic information. This suggests that currency may in reality be an important factor. One of our participants states the following reason for filing currency-related objections: “*Traffic related information changes quite often and updates may be slow, and I found many cases of wrong answers. I have been mainly filing objections on direction questions. Since I joined this community, I filed more than 6000 cases. I think that this is the largest number of objections in Jisiklog.*”

### INTERACTION DYNAMICS

We investigate longitudinal user interactions by corroborating the analysis results of five-year-long Q&A dataset with the survey results. In particular, we study longitudinal working patterns of answerers (i.e., working hours, frequency) and community dynamics (i.e., churning and dropouts). We find that experienced answerers invest a significant amount of time daily (on average 8 hours), and their daily working hours are likely dispersed over multiple work periods. There is a tendency that the longer the daily working hours, the larger is the number of work periods (e.g., heavy users work more hours more frequently). Further, while answerers’ monthly churning rate is fairly high, the contribution of top- $k$ % answerers is quite consistent over time, showing the robustness of Jisiklog.

### Longitudinal Working Patterns

We begin by looking at peoples working patterns on Jisiklog, including the hours spent working and the frequency of work blocks. We tracked the activity of workers who answered more than 100 questions over the entire period. For a given worker, we define a work block as a series of answers where the inter-answer time of any pair of consecutive answers is no greater than 30 minutes. The assumption is that the worker is likely to do other things besides working in Jisiklog if the break time is longer than 30 minutes. To measure how frequently an individual worked for a given day, we use a metric of the number of work blocks per day. The duration of a work block (in hours) is the difference of posting times between the first and last questions in a work block. If a work block contains only one question, we treat it as a singular work block and fix the duration as three minutes.

We present the cumulative distribution of work hours per day in Figure 7. The shape of the distribution is quite narrow and skewed (average: 8.55 hours). We also present the boxplot

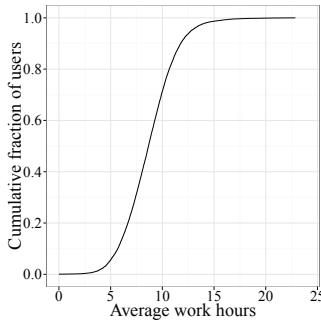


Figure 7. Daily work hours per user

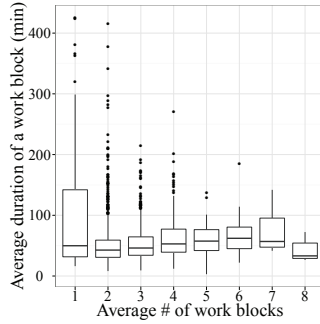


Figure 8. Num. work blocks vs. block duration

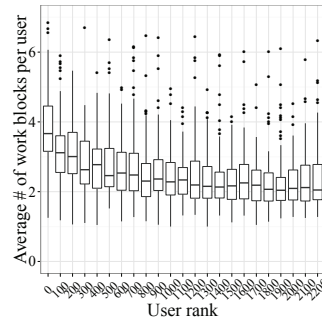


Figure 9. Rank vs. num. work blocks

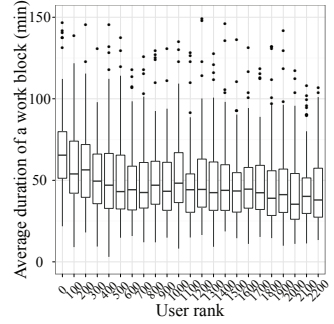


Figure 10. Rank vs. block duration

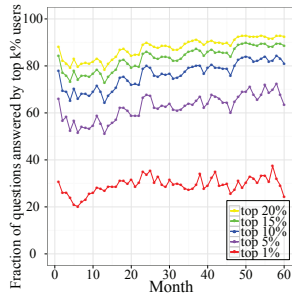


Figure 11. Top- $k$ % contribution

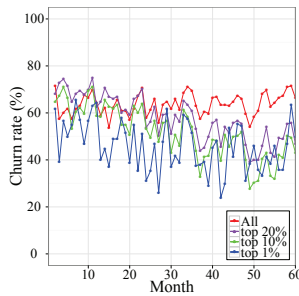


Figure 12. Top- $k$ % churning

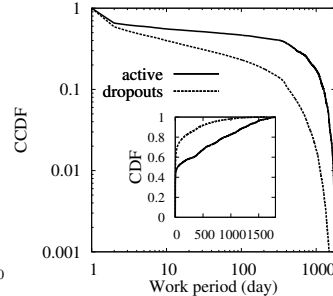


Figure 13. Work periods

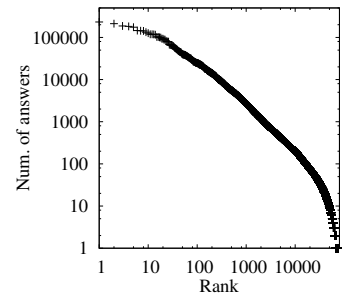


Figure 14. Rank vs. contribution

of the average duration of a work block per user as a function of the average number of work blocks per user in Figure 8. This figure clearly shows that many answerers attempt to work over the course of a day as time permits, consistent with our survey results—one of the leading motivators was to make fruitful use of spare time. For the workers who work only once per day (the number of blocks = 1), the dispersion of work block duration is much wider than that of the other workers. For the other workers who work more than once, there is a tendency of working more hours and more frequently; as the number of work blocks increases, the median work duration also increases. To explore whether this pattern appears in the experienced worker group (top 2000 users), we also plot the number of work blocks (Figure 9), and the average work block duration (Figure 10) as a function of worker rank (with a bin of 100 workers). The graphs clearly show that the higher the rank, the longer the work duration and the larger the number of blocks.

**Community Dynamics**

We investigated the community dynamics by examining the temporal changes of contribution and churning rates of top- $k$ % answerers. We also studied the work periods of individual workers (active vs. dropout) and identified the major reasons why some participants temporarily stopped from work.

We first present the contribution of top- $k$ % answerers over the five-year period ( $k = 1-20\%$ ) in Figure 11. For a given month, we rank all the workers, and the contribution of top- $k$ % answerers is plotted. The average contributions of top-1%, top-5%, top-10%, top-15%, and top-20% answerers are 29.3%, 62.4%, 76.1%, 83.2%, and 87.6%, respectively. We also analyze the monthly churn rate (or attrition), which measures

the fraction of workers who are active in one month becomes inactive in the following month. For example, assume that 4 users (A, B, C, D) are active in the first month, and 4 users (C, D, E, F) are active in the second month. Since two users (A, B) are left, the churn rate of active workers in the first month is 0.5. The churn rates of the top- $k$ % active answerers in each month ( $k = 1-20, 100$ ) are plotted in Figure 12. The average churn rate of top 1%, top 10%, top 20%, and 100% is 45.7%, 51.6%, 57.8%, and 63.5%, respectively. Even with high churning in each month, the contribution of top- $k$ % answerers does not dramatically change over time, which partly demonstrates the strength of power-law networks [8, 20].

Since a large fraction of survey participants are students, we examined whether school schedules such as back-to-school affected the churn rate, but we do not find any significant correlation. Instead, we find that some of the promotion events by Jisiklog have great impact on churning rates. For example, when the piece-rate was increased in the 12th month as seen in Figure 12, the churn rate of top 1% and top 10% decreased from 64.5% and 58.9% to 40% and 58.3% in the following month, respectively. Also, when Jisiklog announced that they would give more experience points, which helped users raise their levels easily in the 14th month, the churn rate of top 1% and top 10% was decreased from 44.7% and 65.4% to 37.1% and 62.8%, respectively.

To understand how long workers stay in the system, we calculate the work period of individual workers. We divided workers into two groups, namely active and dropout. If a worker is not active in the last 6 months, we assume that the worker has dropped out of the system. Note that there are some users who are not active for more than 6 months and become active again, but such cases are quite rare in our dataset. We



plot the cumulative distribution of the work periods for both groups in Figure 13. The distribution for dropouts approximately follows a power-law distribution with an exponential cutoff due to the 54 month limit. This means that quite a large fraction of people leave quite early (e.g., about 41% in 1 day, 80% in 154 days, and 90% left in 400 days), whereas once they are committed, they tend to stay longer and consistently contribute as shown in Figure 14. This phenomenon is also observed in traditional social Q&A sites such as Yahoo! Answers and Naver KiN [26]; it appears that monetary incentives do not significantly increase user activity lifespan as opposed to social motivators.

In our survey, we asked the participants whether they had experiences of stopping from work (for example, for a week to a few months) and if so, what the reasons were. 77.1% of the participants said they had such experiences. The following are the major reasons that emerged from the responses: (1) too busy with primary work (46.8%); (2) lack of motivation due to low financial incentives (11.2%); (3) tiresome to answer questions (8.1%); (4) too competitive to pick up questions (7.5%); (4) low acceptance rate and frequent objections (7.0%); and (5) difficult to answer questions or/and to provide good references (4.3%). These reasons partly explain the findings about community dynamics. The entry barrier of novice users is fairly high; a large portion of users tend to drop out of the systems at an early stage since the work environment is quite competitive and demanding and the compensation is low. Once users are committed to working (e.g., successfully surviving the on-the-job training period, and having become an expert jisikman), they tend to stay longer. However, their contributions could be intermittent, as a large portion of users are doing this work during their spare time; Jisiklog nonetheless has enough committed workers who can sustain the overall community.

## DISCUSSION

We discuss in this section several design implications on the basis of our findings, namely designing assistive tools for answerers and employing mechanisms for improving intrinsic/extrinsic motivations. We also discuss the limitation of this work and possible directions of future work.

There are several ways of designing effective assistive tools for answerers. First, we found that the answerers in the current system lack a sense of belonging to the community, and thus they are not inclined to share knowledge (e.g., tips, bookmarks) with other answerers. Introducing sharing mechanisms as in social bookmarking (e.g., del.icio.us, Digg) would not only improve the efficiency of workers but also stimulate social interactions, thus enhancing the sense of belonging. In the case of bookmarks, it is possible to use machine learning algorithms for automatic web link recommendation. Second, since a majority of answerers are part-time workers, and their interaction patterns are quite intermittent, another effective assistive tool would be supporting work scheduling functions on the basis of detailed analysis of working patterns; e.g., efficiently reminding users of the importance of their contributions. Finally, helping users to track work related activities would be very useful to improve work efficiency and

participation (e.g., validation/objection event tracking, extra audio/visual cues on remaining time).

We showed that both intrinsic motivators (e.g., enjoyment, learning, altruism) and monetary incentives play important roles, but the social motivation is quite weak in Jisiklog. We suggest introducing mechanisms that can improve perceived values of learning and altruism and also facilitate social interactions. One way of increasing a user's perception of learning would be to permit the user to browse a list of answers or visualizing tags/keywords based on interaction history (or topics). We could enhance a user's feeling of altruism by enabling a donation feature; e.g., directly transferring earnings to a charity foundation—in this case, answerers are helping both the askers and other people in need. A community-wide donation campaign would improve the sense of belonging to the community. The current system lacks features that can facilitate social interactions; some well-known features can be integrated (e.g., sharing awareness information of other workers, displaying performance in leaderboards, etc.). With piece-rate wages, goal setting happens quite naturally [12]. We can elicit such behavior by introducing related features, and designing more effective financial incentive mechanisms for mobile pay-for-answer Q&A is part of our future work.

In mobile pay-for-answer Q&A, most activities are time critical (e.g., question pick-up/answering). While this level of interactivity would provide a feeling of engagement, it would be cognitively demanding and could result in reduced long-term participation. Gamifying some of the activities or introducing social awareness (e.g., online chatting) could lower this burden. Our analysis showed that labor supply and demand dynamically change over time in Jisiklog. Further investigation of the dataset is needed to better understand the ecology of an online labor marketplace using economic theories [11]. For instance, labor oversupply creates a competitive work environment (leading to high entry barriers to newcomers and high cognitive loads to workers that are due to the attrition of workers). On the other hand, because of the real-time nature of question asking [17], labor undersupply may lower service quality (e.g., large response delay), thus causing asker attrition.

The main findings on Jisiklog can be used to understand other mobile pay-for-answer Q&A sites that share the same design characteristics. For instance, ChaCha has a very similar workflow/incentive structure as that of Jisiklog. More broadly, mobile pay-for-answer Q&A can be thought of as a real-time knowledge marketplace for mobile information seekers. Our findings will thus provide valuable insights into the design of real-time crowdsourcing systems that exploit the wisdom of crowds such as real-time visual Q&A like VizWiz [5] (e.g., designing a workflow for real-time quality control, assistive tools for creating efficient work environments, and mechanisms of accommodating diverse motives).

However, as with any qualitative or single-site work, the generalizability of this work is limited such that additional research on similar mobile Q&A services is necessary. For this reason, we have performed a preliminary study on the user behavior of ChaCha based on a trace dataset for almost a one

year period. Due to limited availability of public information, thus far we have only been able to confirm that user contribution and topic selection patterns are similar to those of Jisiklog. Further investigation is needed to compare the findings with those related to ChaCha and other sites. Another aspect worth exploring is the investigation of the cultural implications on the online labor marketplaces; identifying key differences between different cultures will help in generalizing the results.

## CONCLUSION

We studied the behaviors and strategies of crowd workers in mobile pay-for-answer Q&A using the longitudinal Q&A dataset from Jisiklog and the results of a complementary survey study of Jisiklog workers. The following findings were derived from our analysis structured along the key aspects of user interactions, namely major motivators of user participation, working strategies of experienced users, and longitudinal interaction dynamics. First, unlike traditional Q&A sites, we found that users are rarely motivated by social factors, but are more motivated by financial incentives and intrinsic factors (e.g., altruism, fun, and learning). Second, experienced users have fairly unique working strategies on topic selection, answer search, and answer quality assessment. Unlike traditional Q&A sites, an answerer's topic selection tends to be broad, and yet, answers are provided quickly (taking less than 90 seconds on average). Experienced users have several common answering strategies (e.g., referring to diverse web resources), and the primary factors of quality assessments include accuracy/validity, source quality, and specificity. Third, we found that even with high churning in each month, the contribution of top- $k$ % answerers is quite consistent over time, showing the robustness of mobile pay-for-answer Q&A.

## ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (MEST) (NRF-2012R1A1A1008858). We would like to thank Jisiklog for providing the dataset, Dr. Jaime Teevan for shepherding the manuscript through the revision process, and the anonymous reviewers for their helpful comments.

## REFERENCES

- Jisiklog Service Guide.  
[http://cs.jisikman.com/cs/h\\_info/info.jsp](http://cs.jisikman.com/cs/h_info/info.jsp).
- Kokoma Korean Morpheme Analysis.  
<http://kkma.snu.ac.kr/>.
- Adamic, L. A., Zhang, J., Bakshy, E., and Ackerman, M. S. Knowledge Sharing and Yahoo Answers: Everyone Knows Something. In *WWW* (2008).
- Barry, C. L., and Schamber, L. Users' Criteria for Relevance Evaluation: A Cross-situational Comparison. *Information Processing & Management* (1998).
- Bigham, J. P., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R. C., Miller, R., Tatarowicz, A., White, B., White, S., and Yeh, T. VizWiz: Nearly Real-time Answers to Visual Questions. In *UIST* (2010).
- Chen, Y., Ho, T.-H., and Kim, Y.-M. Knowledge Market Design: A Field Experiment at Google Answers. *Journal of Public Economics Theory* (2010).
- Edelman, B. Earnings and Ratings at Google Answers. *Economic Inquiry* (2012).
- Granovetter, M. The Strength of Weak Ties: A Network Theory Revisited. *Sociological Theory* (1983).
- Harper, F. M., Raban, D., Rafaeli, S., and Konstan, J. A. Predictors of Answer Quality in Online Q&A Sites. In *CHI* (2008).
- Heyman, J., and Ariely, D. Effort for Payment: A Tale of Two Markets. *Psychological Science* (2004).
- Horton, J. J. Online Labor Markets. In *WINE* (2010).
- Horton, J. J., and Chilton, L. B. The Labor Economics of Paid Crowdsourcing. In *EC* (2010).
- Hsieh, G., and Counts, S. mimir: A Market-Based Real-Time Question and Answer Service. In *CHI* (2009).
- Hsieh, G., Kraut, R. E., and Hudson, S. E. Why Pay?: Exploring How Financial Incentives are Used for Question & Answer. In *CHI* (2010).
- Ipeirotis, P. G. Analyzing the Amazon Mechanical Turk Marketplace. *Journal XRDS: Crossroads* (2010).
- Jovanovic, B., and Nyarko, Y. Learning by Doing and the Choice of Technology. *Econometrica* (1996).
- Lee, U., Kang, H., Yi, E., Yi, M. Y., and Kantola, J. Understanding Mobile Q&A Usage: An Exploratory Study. In *CHI* (2012).
- Mason, W., and Watts, D. J. Financial Incentives and the "Performance of Crowds". In *KDD-HCOMP* (2009).
- Nam, K. K., Ackerman, M. S., and Adamic, L. A. Questions in, Knowledge in?: A Study of Naver's Question Answering Community. In *CHI* (2009).
- Raban, D., and Harper, F. M. Motivations for Answering Questions Online. In *New Media and Innovative Technologies* (2008).
- Raban, D. R. The Incentive Structure in an Online Information Market. *JASIST* (2008).
- Shah, C., Oh, J. S., and Oh, S. Exploring Characteristics and Effects of User Participation in Online Social Q&A Sites. *First Monday* (2008).
- Wang, S., and Noe, R. A. Knowledge Sharing: A Review and Directions for Future Research. *Human Resource Management Review* (2010).
- Yang, J., Adamic, L. A., and Ackerman, M. S. Competing to Share Expertise: the Taskcn Knowledge Sharing Community. In *ICWSM* (2008).
- Yang, J., Adamic, L. A., and Ackerman, M. S. Crowdsourcing and Knowledge Sharing: Strategic User Behavior on Taskcn. In *EC* (2008).
- Yang, J., Wei, X., Ackerman, M., and Adamic, L. Activity Lifespan: An Analysis of User Survival Patterns in Online Knowledge Sharing Communities. In *ICWSM* (2010).