Questions in, Knowledge iN? A Study of Naver's Question Answering Community

Kevin K. Nam
School of Information
University of Michigan
Ann Arbor, MI
ksnam@umich.edu

Mark S. Ackerman
School of Information and
Dept. of EECS
University of Michigan
Ann Arbor, MI
ackerm@umich.edu

Lada A. Adamic
School of Information
University of Michigan
Ann Arbor, MI
ladamic@umich.edu

ABSTRACT
Large general-purposed community question-answering sites are becoming popular as a new venue for generating knowledge and helping users in their information needs. In this paper we analyze the characteristics of knowledge generation and user participation behavior in the largest question-answering online community in South Korea, Naver Knowledge–iN. We collected and analyzed over 2.6 million question/answer pairs from fifteen categories between 2002 and 2007, and have interviewed twenty six users to gain insights into their motivations, roles, usage and expertise. We find altruism, learning, and competency are frequent motivations for top answerers to participate, but that participation is often highly intermittent. Using a simple measure of user performance, we find that higher levels of participation correlate with better performance. We also observe that users are motivated in part through a point system to build a comprehensive knowledge database. These and other insights have significant implications for future knowledge generating online communities.

ACM Classification Keywords
H.5.3 Information Interfaces and Presentation (e.g. HCI): Group and Organizational Interfaces; J.0 Computer Applications: General

General Terms
Measurement, Human Factors

Author Keywords
Question-answering, online community

INTRODUCTION
The Web has clearly led to new forms of knowledge production on a scale never seen before. One of the most interesting forms is the question and answer (Q&A) community. Q&A communities serve an important role in informal knowledge production, one focused on users helping one another.

CHI 2009, April 3 - 9, 2009, Boston, MA, USA.

While it is clear that Q&A communities, in general, are pervasive and successful on the Internet, what is less clear are the structural characteristics required for adequate knowledge provision in these worlds. New ultra-large scale Q&A sites, such as Naver Knowledge-iN, Baidu Knows, or Yahoo! Answers, are general-purpose community Q&A sites with millions of users. Preliminary results (e.g., [1]) suggest they have different structural characteristics than smaller sites, and little is known about how people self-organize to provide answers to their millions of questions yearly.

This paper seeks to understand the underlying user participation behavior and strategies in the ultra-large Naver Knowledge-iN (KiN) service. KiN, started in South Korea in 2002 [19], has an average of 110,000 answers to 44,000 questions asked every day and 4.5 million daily visitors [12], and has been a commercial success for Naver. Our goal is to describe how KiN users' expertise is distributed and arranged so that its users' questions can be adequately answered and the community maintained over time. We believe that understanding these issues will lead to new insights for the social facilitation of Internet-based knowledge production.

This paper first discusses related work on online expertise sharing communities. It then describes the features of Knowledge-iN (KiN) and our data collection and analysis methods: a mix of quantatative analysis and qualitative interviews. Following an overall description of the system, we focus on the top answerers on KiN and their characteristics. We then present an analysis of the users' motivations and the observed allocation of expertise. We conclude by discussing design implications for future online Q&A systems.

RELATED WORK
Recently, there has been considerable work on improving the functionality of Q&A sites. One challenge is making the great store of knowledge being generated accessible to subsequent askers. Jeon et al. [10] developed methods to retrieve semantically similar questions that may already contain answers. The validity of such answers is uncertain, a problem that has been addressed by identifying the level of expertise of the answerers [11, 22], and by using community-generated ratings, such as best answer ratings [2, 1, 3]. In the present paper, rather than trying

to identify high quality answers, we instead examine how such Q&A systems can sustain a level of participation to guarantee appropriate answers.

The patterns of participation for online communities in general have been studied in a range of studies. For example, Wenger [17] discussed the importance of different roles in online communities and how they affect community formation and continuation. Nonnecke and Preece [14] studied lurker behavior in different online forums. Holloway et al. [9] and Viegas et al. [16] described different contribution and coordination patterns of Wikipedia authors.

Substantially fewer studies of Q&A communities exist. In general, Welser et al. [18] found that there are certain users who handle a disproportionate share of answering. Zhang et al. [22] examined the Java Forum community and found clusters of users who ask, users who answer, and users who do both. Adamic et al. [1] examined Yahoo! Answers, however, and found far more separation of answerers and askers. These answerers are critical to the success of Q&A sites, since they provide the answers that draw askers. Yet finding enough answerers, especially for ultra-large sites and especially if answerers are a distinct set of users, is likely to be an interesting challenge.

Prior work has uncovered many of the motivations and incentives for answering questions in online communities in general. Yu et al. [21] discuss motivations such as active learning, self-enhancement, reciprocity, reputation, enjoyment of helping others, self-protection, moral obligation and the advancement of the virtual community. Within an organizational context, Constant et al. [6] found altruism to be a strong motivation for answering questions, while strong social ties were not. Butler et al. [5], in an analysis of list-servers, found that users participated to obtain otherwise inaccessible information and visibility in social relationships.

Different incentives and openness of participation also appear to affect the quality of the answers and user behavior. Harper et al. [8] found by comparing different Q&A systems that rewarding answerers with money and increasing the rewards induced higher quality answers. Yang et al. [20] found that offering more money for a solution to a task correlated strongly with attracting more views for the task, and correlated weakly with increased task participation.

Our work complements these findings by examining the highly dynamic participation of answerers in a ultra-large Q&A community. We will see that the structure and arrangement of expertise is different in this community. KiN users get answers to their questions, but the result is a lack of community and intermittency of participation.

INTERACTION IN KNOWLEDGE IN

Naver (http://www.naver.com) is the most popular portal site and search engine in South Korea, set up by NHN Corporation in 1999. At the time, most of the documents on the Web were still written in English, which made the development of a search engine in Korean very difficult. NHN started its KiN service in 2002 to overcome the lack of Web documents in Korean [19]. By having users ask and answer many questions, Naver created a searchable database helping the portal to be the number one search engine in Korea.

The main activities on KiN are searching, asking, and answering. The KiN main page aims to capture users' attention while at the same time satisfying information needs and acknowledging the most active users. The page lists some of the most "popular" questions (questions that people voted as useful), questions that need an answer, and questions for which an answer needs to be selected by the askers. The site also presents its visitors with the list of top categories where people can post a question and answer, and provides a list of real time search words that reflect the current interest of the users. Some of the top answerers in the site are also posted on the main page.

The format of KiN constrains how a question-answer exchange proceeds. An entry is created by a user's posting of a question, and subsequent answers to the question follow in the same entry (see Figure 1). Since a user can only post one answer to a specific question, it makes Usenet or forum-type discussions difficult. While users may leave a one line "comment" to the question or any of the answers, and an answerer can rebut to an objection, the structure is flat and not threaded as often seen in a forum like Slashdot. A user may also directly email or send a message to another user to ask a question.

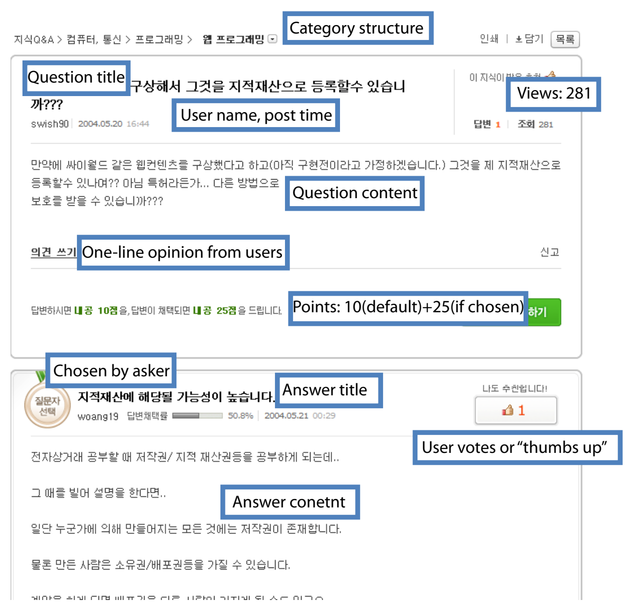

Figure 1. An example of a question-answer exchange

KiN has accumulated over 60 million user-generated posts in 14 top categories and over 3,500 sub-categories. As of our data collection, the three largest top categories were 'Computer / Telecommunication' (4.58 million entries), 'Entertainment and Arts' (2.58 million entries) and 'Education' (2.53 million entries), reflecting the interest of Korean Internet users. Other top categories include 'Games', 'Business / Economy', 'Shopping', 'Society/Culture', 'Health/Medicine', 'Family/Life', 'Travel/Leisure', 'Sports', 'Local Q&A', 'Advice Q&A', and 'Juniver KiN' (for people under the age of 14).

There is a huge range of question types, from factual questions (e.g., "Who was the first president of Korea?"), to procedural questions (e.g., "How do you build a computer?"), opinion oriented questions (e.g., "Who is the better singer, A or B?"), task oriented question (e.g., "Can you write a program in C to do X?"), and advice seeking questions (e.g., "I broke up with my girlfriend. What should I do?"). The prevalence of question type differs by category. The Medical and Finance categories are full of advice seeking questions. Categories such as Java or C/C++ have many task-oriented questions, and an example is shown in Figure 2. The question received three similarly helpful answers and one "junk" post about a current event.

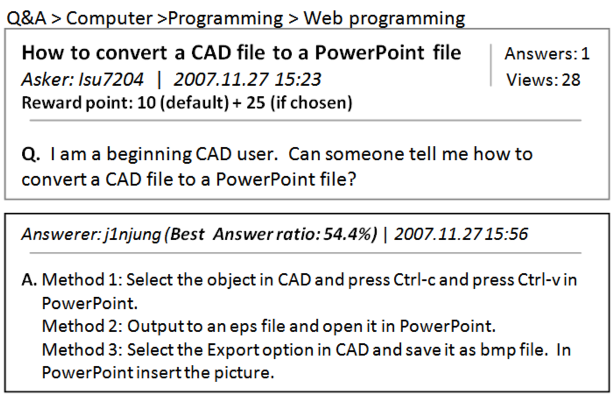

Figure 2. A typical question in Software

In the Movie and Singer categories, many of the questions are opinion-garnering ones that often turn into a flame war. The following question in the Singer category garnered an explosive 277 answers, the one chosen as best being humorous rather than informative:

Q. Do you think the new singer Shiny will be a hit?

A. What? How come someone younger than me can be a singer? They are all fake. Who's Blackbeat? Park Hyun-Bin is the best.
Other answers: 276

Many of the other 276 answers are similarly opinionated and flippant, and some of them have more than 100 one-line "comments" that a user can leave regarding the answer. While questions that generate this many answers are relatively rare, they are more likely to occur in discussion categories such as Entertainment. In contrast, in our data of 2.6 million questions spanning 15 categories, the average number of answers per question was 1.7 and about 59% of the questions received only one answer.

Askers in KiN can ask just about any question they can come up with. Naver assists them by providing an automatic category suggestion feature that can guess relevant categories based on the keywords in the question. KiN rewards answer with a point system that allows users to advance from a status of "lowlife" to "hero" or "god". As of the data collection, an answerer gets 10 points for providing an answer, and 25 additional points for being selected as the best answer. The asker can post up to 100 extra points in hopes of receiving faster and better answers. Users receive points not only for answering questions, but also for various activities in KiN. For example, logging in everyday is 3 points and voting for an answer is 1 point, and so on. Users with a large amount of points are ranked and listed on the site. The top 1,000 users from the entire site are listed on the KiN People page, and each sub-category lists its own top 10 users on its page.

DATA COLLECTION

Our data collection was comprised of two complementary parts: automated crawling of publicly accessible question-answer pairs on the Knowledge-IN website (www.kin.naver.com) and phone interviews. The crawl of millions of questions and their answers allowed us to gather comprehensive statistics on user behavior, while the phone interviews with the most active users yielded insights into the driving motivation behind the observed behavior.

The Web interface of Knowledge-iN allows only viewing of up to a certain number of questions (1,500 as of the data collection) per topic category, which can be as little as a few days' worth of data. To overcome this limitation, we devised a search method using frequently used terms in questions in each category that we believe retrieves over 99% of Knowledge-iN's questions with answers. We collected approximately 2.6 million questions and 4.6 million answers from 15 categories from a five year period between 9/1/02 and 12/31/07. This excludes questions without answers, which are removed from the system after a period of time.

We supplemented the question and answer data with telephone based interviews of 26 KiN users. These users were a mix of heavy and moderate answerers, along with two askers. We found these users by asking people who had recently posted a question or answer (so as to form critical incident interviews), and by asking people who were in the top 1,000 answerers. We had a very low response rate (less than 2%), and as a result this group cannot be considered representative. Nonetheless, their responses were extremely useful in understanding the site, and if taken appropriately, may lead to insights about those who answer on KiN.

The interviews were conducted by the first author (who speaks Korean) and were semi-structured, meaning that the interviewer asks specific questions but was then free to follow-up as appropriate. The interviews lasted between 20 and 45 minutes. They were recorded, and then translated and transcribed. We used standard qualitative coding techniques, based in Miles and Huberman [13], to code and analyze the transcripts. These techniques allow for the appropriate categories of analysis to emerge from the data; as such, our results should be considered generative rather than confirmatory. We then compared with the crawled data which revealed mutually supporting evidence for the role of the point system, users' strategy and desire to answer unanswered questions, and the behavior of correcting existing answers. As the reader will see below, we believe that the combination of the two was analytically fruitful.

PATTERNS IN PARTICIPATION

Those who ask don't answer

In KiN, the users are largely divided into askers and answerers, with only 5.4% both asking and answering in the same category. The Naming category (for suggesting baby names) had the smallest overlap (1.9%) and the Singer category had the largest (10.4%). As a result of the separation of the asker and answerer role, there is very little within-category reciprocity in KiN. This lack of reciprocity is reflected in the absence of reported community by interviewees. As we will see, answerers tend not to know other participants, and rarely exchange personal email.

Table 1. Number of askers and answerers from the 13 categories. Most categories have more askers than answerers, except for the Fashion, Movie, Naming, and Singer categories.

Category	Num Asker	Num Answerer	Ans. by top1%	Avg. ans. by top1%
C/C++	37145	15960	33.5%	223.53
car repair	26618	11735	41.4%	234.90
fashion	117934	142482	22.1%	48.67
finance	148346	72251	52.8%	344.13
java	10437	3986	31.4%	194.28
laws	152221	107832	44.4%	152.45
linux	6723	3390	36.3%	159.09
medical	417243	381210	35.1%	123.17
movies	75082	93794	27.4%	61.49
MS windows	105442	59378	35.6%	138.43
naming	35048	46808	29.7%	55.89
singers	134002	203726	23.4%	63.02
stocks	24352	11982	43.6%	226.97
translations	313637	155686	37.7%	207.35
web design	49816	33052	34.9%	110.32

We also note an overwhelming tendency of users to specialize. Even though Naver contains over 3,500 different subcategories, many users focus on one or a mere handful. Among the top 1,000 repliers, over 52% the users had more than half of their answers in a single category. Among those, 30% had more than three quarters of their answers in a single category. It is possible to maintain this focus either by consistently browsing questions in just a single category, or by specifying a particular category as an "interest" and being shown questions specifically in that category.

Top answerers

Because there is such a strong separation of roles between askers and repliers, we decided to focus on the answerers only. It is among this group of users that we find the heavy tail in activity: a number of users have answered tens of thousands of questions, but it is rare for an asker to ask more than a couple of hundred questions. It is therefore unlikely that a highly active asker is individually significantly contributing to the activity of the site, while a "top answerer" may very well be helping hundreds to thousands of others. Given that the point incentives in Naver are to answer as many questions as possible (with points awarded and no cost, save for effort expended, in providing a poor answer), we wanted to develop a measure that would accurately reflect the quality, and not just the quantity of a person's answer. We use a γ or *guru* score, described in the appendix, to measure the answerers' performance against the random chance of their answers being selected as best.

We compared the most active 1% of the users (by the number of the answers provided) with the rest of the users. Their contributions account for 22% to 52% of all answers ever given (Table 1). The top answerers not only produce many answers, but the quality of the answers, measured by the guru score is better than the average (Table 2).

There are important differences in categories that may affect top answerers' behavior. First, there is little competition for the best answer within a question. It varies among the categories: C/C++ and Java are the least competitive with the average of 1.42 and 1.30 answers per question respectively, and Movies and Singers are the most competitive with 2.23 and 2.57 answers per question. This may be due to the type of expertise sought in each category. In categories where many questions require hierarchical expertise (i.e., to know an advanced topic such as thread programming, one would be sufficiently experienced in basic programming) may have fewer qualified answerers than categories that require "flat" knowledge (i.e., one does not need to know about all singers to answer about a particular singer.). Additionally, one "correct" answer to a programming question may be sufficient, but several opinions may be welcome about a singer.

Table 2. Guru scores for top answerers compared to the rest. All differences significant at $p < 0.001$

category	guru top 1%	guru rest
C/C++	0.0884	-0.172
car repair	0.0230	-0.132
fashion	0.0465	-0.0451
finance	-0.0454	-0.144
java	0.0484	-0.117
law	0.151	-0.176
linux	0.172	-0.0950
medical	0.0573	-0.0859
movies	0.0738	-0.147
MS Windows	0.0428	-0.157
naming	0.0278	-0.116
singers	0.0765	-0.0732
stocks	0.0902	-0.154
translations	0.102	-0.199
web design	0.103	-0.164

MOTIVATION FOR PARTICIPATION

What is most essential to the success of sites such as KiN is the participation of its users, but this is also one of the biggest puzzles. What motivates users to essentially provide knowledge services for others, when there is very little in the way of explicit rewards?

Most of our understanding of why people participate in KiN comes from our interviews. In the interviews, users gave a number of reasons. By far, the most often stated reasons were the wish to help others, to learn and review material, or to participate as a hobby. Several people also pointed to participating for business reasons. These are not mutually exclusive motives: Most interviewees gave multiple reasons.

These motivations echo prior literature (e.g., [6], [5]), as might be expected. However, two of the motivations, participating as a hobby and learning, have implications for the way participation by answerers was intermittent. Furthermore, motivations were often co-mingled with interviewees' discussions of KiN's game-like points and their quest for those points. We will discuss each in turn below, and relate it our analysis of the question-answer dataset as appropriate.

Altruism and helping others

When asked why they participated, many users said to help others, by providing knowledge that others did not have. Altruism, then, was the most oft provided answer. The type of help was dependent on the answerer's expertise. One doctor stated:

> Since I was a doctor, I was browsing the medical directories [in KiN]. I found a lot of wrong answers and information, and was afraid they would cause problems. So I thought I'd contribute in fixing it hoping that it'd be good for the society. [Sangmin]

(All responses have been anonymized.) Another expert stated:

> Many people in Korea have incorrect information about social security, and I was a bit frustrated because I work in the area. So I started out to explain it to people. Sometimes people sent me a thank-you email message for answering, and it motivated me more. [Youngsoo]

Another answerer, who participated in the Translation category, was not as expert. She instead stated:

> I try to answer so that regular people can share knowledge, rather than technical knowledge. ...Someone needs it, and I have the ability to do it, and it'll be a service to society. [Mirae]

Altruism was a very common response for our interviewees. It may seem, at first glance, as though these claims of altruism are merely "public" statements, that is, socially-sanctioned responses that people give as a matter of routine to those they do not know well. There is evidence in our data that suggests that we saw some "backstage" responses (i.e., not merely "public" responses); we will discuss this below further. However, this is deeply resonate with Korean culture, as one of our interviewees insightfully noted:

> I think there's something about Korean people that they take time to write about something when there's no real benefit to them. If you take the GRE for example, one doesn't really need to share information, but on Hackers [a famous Korean site for sharing standardized exam information], people share a lot of information. Some people make their own report and study material [available] without anything in return. [Garam]

Business motives

Yet another interesting motivation also discussed in interviews was promoting a user's business through answers. This goes on in many online worlds - not just KiN. For example, it is a recognized problem in review and recommendation sites [7].

In a more implicit manner, this seemed quite acceptable to our interviewees. One person reported wanting to be considered an expert for the status in his online e-commerce community. Another user reported getting offers to publish books or other tour guides based on his activity in the Tourism category of KiN. At its most extreme, two interviewees reported getting solicited by other online communities attempting to garner online participation.

A more explicit manner of promoting a business was reported in the Medical and Finance categories. (It may exist elsewhere as well; this is a limitation of our interviewing.) Said one interviewee:

> I've been working as an insurance agent for 9 years. I started answering in Knowledge-iN as part of my business activity. In the evening, I answered questions to solicit potential clients.... So when I'd leave an answer, I'd say I would meet with you face-to-face to talk about more details and give you advice. [Taein]

Two interviewees stated that they had originally started on Naver to gain clients, but they found it to be less valuable than they had hoped. Instead, they stayed as a hobby and for altruistic reasons.

Learning

Many interviewees reported wanting to gain further understanding or to maintain their current understanding of a topic. This included reviewing what they knew before or extending their knowledge by explaining it to others. One interviewee said:

> My first intention [in answering] was to organize and review my knowledge and practice it by explaining it to others. [Taein]

Others reported learning through practice:

> Answering questions helps me study. I can learn from answering [in Translation]. I get to review what I used to know such as vocabularies and idioms. [Minhyuk]

Still others reported that explaining a topic to others maintained and perhaps extended their understanding. In a few cases, interviewees reported active learning. Two answerers, both in the C/C++ category, reported taking programming questions as practice problems to learn more about the language.

Review may strike some as less important than active learning. On the other hand, review of material is an essential part of Korean educational processes (much more so than American or most European systems), and so this motivation is heavily resonate with Korean culture.

Hobby and personal competence

Many interviewees reported that they viewed answering questions as a hobby - something to do when they had spare time, as this user did:

But in the evening after work, my kids are asleep, and I don't just want to stare at the TV. I go to Naver, and it's fun to answer. It's interesting to know what questions are coming up. I keep repeating [going there]. [Kisoo]

This implies less of a sense of obligation on the part of the participants and more of a casual interest. One interviewee, however, did note that his involvement was heavier:

Yes [I answer everyday]. I am addicted (laughs). [Nami]

Several other interviewees, as one might expect with any community, spent from morning to night online during some period. For these people, their involvement might be obsessive; others reported continued involvement not from compulsiveness, but from a continued sense of mastery and competency. When interviewees talked about their trajectory of participation over time, many people talked about a strong desire to earn points and advance in level, to promote their businesses and professional lives, and to help others early in their KiN "careers." After the initial period, it appeared that, while the altruism stayed or grew stronger, many participated more as a hobby (which reflected their casualness of participation). As we will see later in the paper, this resulted in often intermittent participation over time.

Points

The point system in Naver is another source of motivation for users. It also interacts sharply with the other forms of motivation. As mentioned, when an asker posts a question, he or she may give up to 100 additional points hoping to get a faster and better answer. Since some of the top answerers enjoy high visibility and sometimes a celebrity-like reputation on the site, points can be a driving reason for participation.

Interviewees often dismissed the points, as this user did:

I don't really care about the points. Points don't affect me in looking at questions or leaving an answer. [Mirae]

Most interviewees echoed this. But she goes on to say:

Although points do not affect whether I answer a specific question or not, it's fun to see them accumulate.

This pattern was also typical in our data. For example, another interviewee stated:

I don't care about the points. [but] It's fun to see points accumulate and my character level up [increase to the next level]. [Jeyeon]

This interacted with helping others. Interviewees felt that others would trust them more or consider them greater experts if they were higher levels, such as this interviewee:

I felt I needed to be a high rank so people would trust my answers more. That way more people would look at my answers. [Sangmin]

Overall, interviewees claimed they did not care about the points, but points did seem to have an effect, often weak, on their activities:

I'd be lying if I said [points] had no effect. Points are really nothing if you think about it, though. But you care about it. [Sunhan]

That users are weakly motivated by points is also evident in the number of answers a given point reward attracts. First, we describe the distribution of points offered across all questions in our dataset. Over 60% of all questions across the categories have only the minimal reward points. The number of questions with a given number of points decreases gradually as the number of points increases with the exception of bumps at 50 and 100 points (Figure 3). These two bumps, along with small peaks at increments of 10, may be indicative of askers thinking in terms of coarse granularity of points (i.e., small, medium, and large rewards).

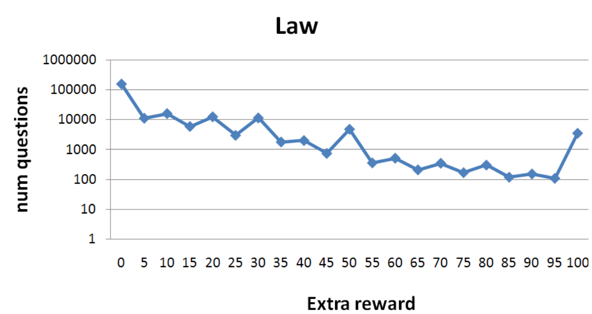

Figure 3. Typical number of questions per expected rewards. Other categories exhibit a similar distribution.

By observing the number of replies at each point level, we find that users are slightly more motivated to answer questions with higher point awards. In most categories, the average number of answers to a question gradually increases as the expected reward goes up (Figure 4). The additional motivation to answer questions with higher awarded points was mentioned by several users, though primarily as a secondary motivation following answering unanswered questions. Many also mentioned weighing the amount of time a question would take to answer against the number of points offered.

While users may be motivated by higher numbers of points, the points offered are also used an indication of how motivated the asker is to obtain an answer. One interviewee noted:

Usually questions with points do not seem frivolous. I feel like answering questions with points, not because of the points, but because those questions are more detailed and seek realistic help. [Taein]

Therefore the attraction of answerers to questions with higher awards may be in part due to the perceived need of the asker, in addition to the interestingness of the question. The points offered are a surrogate for the perceived need. A related observation was made in an analysis of Yahoo! Answers that showed the good answers tended to correspond to good questions [2].

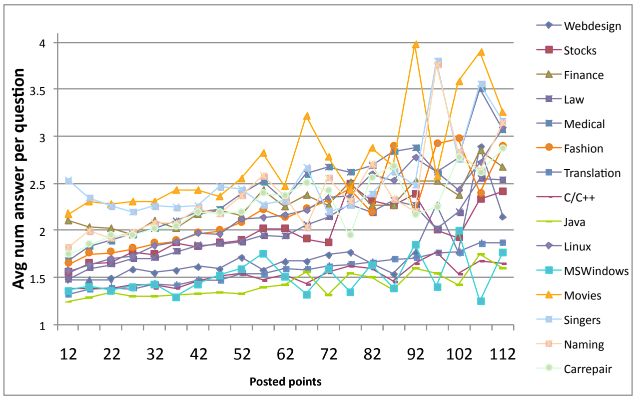

Figure 4. Average number of answers for a question offering a given number of points as reward.

ALLOCATION OF EXPERTISE

As mentioned, KiN is a highly popular site whose community has collaboratively generated over 60 million questions and answers, with approximately two thirds of the questions receiving an answer. But the success of such a system hinges not just on the volume of questions answered. In this section, we examine three critical characteristics for any Q&A site - what knowledge level users can expect, whether problematic or erroneous responses are corrected, and how well the necessary expertise is covered across time and categories. We cover each in turn.

Knowledge level and quality

There appears to be a level of knowledge that users can expect in answers on KiN. All of the interviewees stated that the information in KiN was useful for getting information on commonsense knowledge, current events, basic domain knowledge, advice and recommendations from people, and diverse opinions. Our examination of the site confirms this view. As one interviewee stated:

> It's very useful for getting everyday information. The most useful information is recipe or directions. [Sunhan]

Some interviewees were able to point to certain types of information that could not be found in KiN:

> It's hard to get professional knowledge. [Jinoh]

> Knowledge-iN is not for very domain specific, technical information. [Hyeil]

Many of the heavy answerers in our interview pool stated that online cafés were a better source of detailed information. A café is a type of popular online community in South Korea where people with similar interests can join and perform various activities such as posting new information, uploading and downloading materials, and discussing related topics. Sometimes the members have offline meetings.

This inclination towards a level of answer appears to be heavily affected by two factors in how answerers pick candidate questions. First, people tend to want to answer quickly. Some of this results from the point system. Said two users:

> I look at high points first. But I answer all questions. [Eunjin]

and

> To higher point ones I provided more detailed answers (laughs). [Sanggyu]

However, our answerers are also pressed for time. Some answer when out of work or when their job responsibilities are light, but others try to cover many questions quickly:

> If I'm answering when I am taking a break at work, or I am answering at home in the evening after work, I don't have much time. A longer answer takes more time. [Sunhan]

Time pressure, of course, interacts with the desire to obtain points.

Second, our interviewees tended to answer questions for which either they already knew the answer or they had to look up only minor additional information:

> I only try to answer with what I know. If I know somewhat, but not entirely, then if I answer it may be incorrect. It might hurt the asker. [Minjae]

Others were willing to answer questions slightly beyond their expertise:

> If I am not sure about the answer, I skip [it]. Sometimes I study further and provide an answer. [Manki]

While our expert professionals mentioned answering in depth, they also added two additional constraints. Answering in depth required substantial time, which they seldom had. As well, askers did not provide the relevant detailed information to make recommendations or suggest diagnoses. These kinds of interactions were, in their opinion, more prevalent in the more intimate cafés mentioned above.

This constant churn of mid-level questions and answers is reflected in our interviewees' view that KiN lacks a sense of community, again in contrast with the close-knit atmosphere of the cafes. All of the interviewees mentioned that they did not interact with anyone in KiN other than through asking and answering, and that there was neither a sense of community in existence nor the possibility of forming a community in the future. No one mentioned detailed interactions with other users.

Correcting inaccurate information

If there are many erroneous answers, as interviewees believe, then it is important to a well functioning site that those answers be corrected. Indeed, interviewees reported often that they supplemented another person's answer when it was incorrect.

We see evidence of this in the reply patterns. In the Java and C++ categories questions that receive between two and five replies (we chose to omit long threads that may represent discussions), the next-to- last question is chosen as best answer 51% of the time (in C++, 50% in Java), and the last question 69% (66%) of the time ($\chi^2 = 974$, $p < 10^{-16}$). Note that more than one answer can be selected as best. Even in the Singer category, where arguably people are frequently expressing opinions, the last answer is selected as best 50% of the time, compared to 40% for the next to last answer. This indicates that later replies, especially the last reply, are likely a correction or improvement upon earlier answers.

Further evidence that users evaluate the quality of the previous answer before posting their own is found in the fact that answers posted by users who have a history of good answers are less likely to prompt additional answers. Specifically, if the first answer is given by a user with a guru score, the number of subsequent replies is reduced ($\rho = -0.11$, $p < 10^{-16}$). This suggests that other users, upon seeing a good answer, are less likely to submit another. The above two observations are consistent with a strategy where users elect to answer questions where their expertise is needed, but tend to avoid wasting effort on already adequately answered questions.

Intermittent participation

In addition to providing answers at an expected level of knowledge and quality, the necessary expertise has to be at hand on the site. This can occur only if there are sufficient answerers to handle the workload. However, we find that *individual* users are highly intermittent in their participation.

One user commented:

> ...since the last quarter of 2007, I became the head of the team and the range of my work changed, so it was difficult to find time. [Mike]

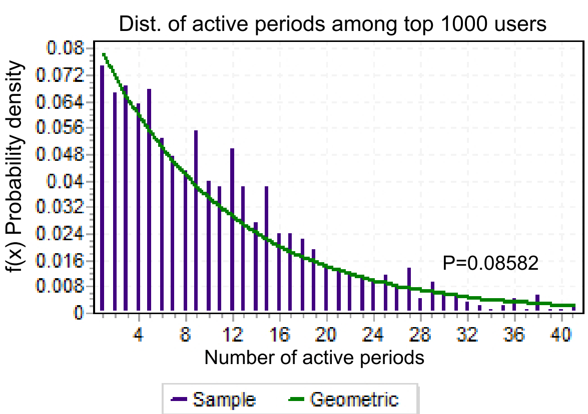

Figure 5. Distribution of active periods among top 1000 users.

Another user, whose activity patterns over the span of a year are shown in Figure 6 mentioned new familial obligations, and loss of access to a computer at home (having to pay for Internet access at Internet cafés instead) as reasons for weeks of inactivity. When we consider the average number of answers posted per week for all users who had posted at least 100 answers in the Java and other categories, we see a steady decay in activity from the very first week (see Figure 7). This is due to many users starting out with a burst of activity, but ceasing all or most activity within a matter of weeks.

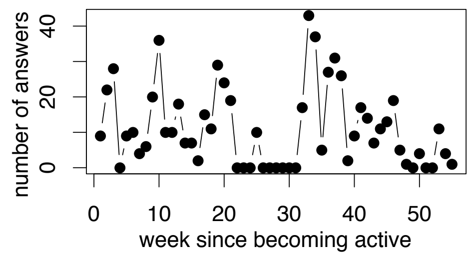

Figure 6. Weekly activity levels of a user.

One might expect that users who remain on the site and answer thousands of questions do so by participating regularly. The surprising result was that even among the most active users, intermittency was the norm. We defined an inactive period to be one where the user posts no answers for one or more weeks. Figure 5 shows the distribution in the number of active periods, separated by inactive periods. Although users typically have a few active periods, quite a number of users have dozens of active and inactive periods. Note that an active period may encompass a single week up to years. A user with a single active period may have only used the system briefly, or has been using the system so regularly, that he/she never became inactive.

That many users are leaving after a short period, or attending to questions only intermittently, appears to have relatively little impact on the rate of questions being answered and the quality of those answers. The number of answers per question remains steady, even as many users join and leave. For example, in the medical category, the number of questions per week fluctuates by 25%, due to seasonality in the data, but the weekly ratio of answers to questions varies only 8%. Nevertheless, the users do not always meet the demand for answers with a proportional supply. An increase in the number of questions is correlated with fewer answers per question ($\rho = -0.29, p < 0.001$).

Looking at 1,000 users who answered the most in the 15 categories we crawled, those who were active a greater number of weeks also answered more questions ($\rho = 0.44, p < 0.001$), but the time span from initial to final post was less correlated with volume ($\rho = 0.12, p < 0.001$). This again indicates that many users' activity need not be contiguous. Indeed, as we learned from the interviews, it often occurs as a hobby, in front of the television, when one has few job responsibilities, or otherwise has free time. From a regression, shown in Table 3, we can see that being active for a greater number of weeks has a weak positive correlation with quality, while being more intermittent has a weak negative impact. This implies that more committed and consistent users are more likely to provide good answers.

DESIGN IMPLICATIONS AND CONCLUSION

Naver's Knowledge-iN, an Internet-scale Q&A site, with its 60 million questions and answers, is arguably one of the most successful sites of its kind in the world. The success of KiN and other Q&A forums masks to what extent providing answers by ordinary users about all kinds of questions on such a scale is a difficult task. Expertise must be organized and allocated, which is especially difficult when it is

Table 3. Regression model of users' guru score based on temporal activity. * indicates $p < 0.001$**

Variable	estimate (se) * 100	
# answers	-0.001	(0.002)
# active weeks	0.170***	(0.043)
# active periods	-0.863***	(0.205)
time span (weeks)	0.047.	(0.025)

$R^2 = 0.04$

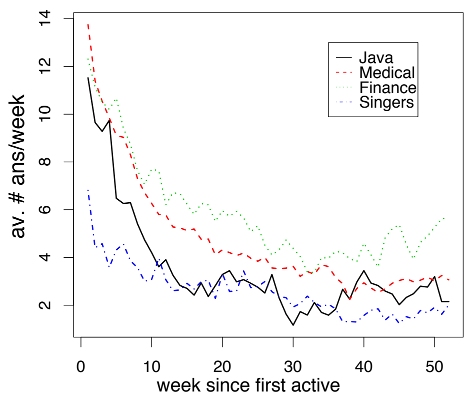

Figure 7. Weekly contributions averaged over users who posted > 100 answers and became active more than a year prior to the crawl.

discretionary and unpaid. In this paper, we analyzed approximately 2.6 million questions and 4.6 million answers, and interviewed 26 participants. We found that answerers' motivations were a combination of altruism, learning, participating in a hobby that brought a sense of competance, business motives, and questing for points. We also saw, through a quantitative analysis, how these motivations led to critical features of KiN's question answering - a range of knowledge that was displayed and sought, the ability to correct answers, and coverage.

KiN is a commercial, production-level system that has made many changes over the years. These work very well, and they carry their own immediate design implications for people running or constructing online forums :

- KiN has a view where answerers can quickly see what questions have not been answered. This view can be filtered for only the topics of interest to a user – an important feature, since as we have seen, top answerers tend to stay within their specific topics.
- The point system works well in motivating answerers, which is critical. In particular, allowing askers to give more points for high-urgency questions channels answerers to those questions.
- Providing Q&A results in Naver's search engine allows other users to see what KiN does and, perhaps more importantly, gives them a sense of what is possible and what is allowable in KiN before they ever post a question.

On the whole, Naver appears to work well. While the percentage of questions answered is not as high as it could be, askers do get answers, people continue to come to the system, and an enormous corpus of Q&A continues to be built. On the other hand, some of the questions that are off-topic, too complex, or unanswerable may have come from new users who do not know how to phrase a suitable question. It may be valuable to provide them with tutorials or other help even within the search function.

Above, we showed that the point system motivates some users. However, the point system doesn't do enough to *de*mot-ivate users whose answers are sub-par. Incorporating our guru score in the point system might deter some of the users we currently observe in the system who have accumulated many points but provided few best answers.

As well, customizing interfaces to fit users' answering strategies could increase contribution and more effectively allocate expertise [15]. A user may choose to answer many easy questions quickly while another user may want to accumulate points by answering more difficult but highly rewarding questions. An expert may only want to answer questions that would reflect his or her expertise, thus establishing an online status in the community. A user may be less interested in answering a question that has already been answered by an expert answerer, or they may be interested in precisely these questions to enhance their learning. Supporting these different strategies through user interface or other mechanisms may provide more motivation to answerers.

Finally, although Naver seems to work well in part due to these few very active top answerers, it also heavily relies on the influx of new users every week. Users leave online communities for several reasons [4]. We've seen that even the most active users are likely to become inactive, either intermittently, or permanently. It is therefore also important to garner new users. KiN gets many new answerers because Naver search returns data from it. Users who originally did not consider participating sometimes do (as indicated in the interviews) when they first use KiN after viewing a search result. Considering Q&A as part of a larger online venue may be critical.

Considerable work remains to be done on Q&A forums in general and KiN specifically. The level of questions in KiN and Yahoo! Answers appears to be relatively low; not much expertise is required to answer them. Finding technical or social mechanisms that encourage people to bring more difficult questions might keep some answerers motivated; finding mechanisms that encourage people with expertise to come to the site may bring more difficult questions. It may be possible, for example, to provide an incentive-based design to re-allocate answerers or to foster more difficult questions. Alternatively, users may use KiN within an ecology of help. We have evidence of that in our interviews, where interviewees go to more specialized forums within Naver to ask difficult questions.

The above design implications all point to the need for further inquiry into how users perceive help facilities and online Q&A forums. This paper presented a fruitful approach involving both quantitative analysis of a large set of user generated data and in-depth interviews of top forum participants. The two approaches gave mutually supporting evidence for the role of the point system as an incentive, the strategy and desire to answer unanswered questions, and also the drive to correct incorrect answers. In future work, we will extend our approach in order to understand more fully why questions go unanswered, and what can be done to influence the rate and quality of answering.

As with any qualitative or single-site work, generalizability is limited; additional work, including collecting attitudinal data, will be required to further understand the role of these motivations and the point system in similar ultra-large Q&A communities.

ACKNOWLEDGMENTS

We would like to thank Jun Zhang for valuable advice on the study. Rayoung Yang and Taeho Ko helped with our understanding of Knowledge-iN. We would also like to thank Minhyung Kang at MIT for insightful discussion and the anonymous reviewers for their helpful suggestions. This work was supported, in part, by the Army Research Institute, the Air Force Office of Scientific Research (MURI award FA9550-08-1-0265), and by the National Science Foundation (0325347).

REFERENCES

1. L. A. Adamic, J. Zhang, E. Bakshy, and M. S. Ackerman. Knowledge Sharing and Yahoo Answers: Everyone Knows Something. In *WWW '08*, pages 665–674.
2. E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne. Finding High-Quality Content in Social Media. In *WSDM '08*, pages 183–194, 2008.
3. J. Bian, Y. Liu, E. Agichtein, and H. Zha. Finding the Right Facts in the Crowd: Factoid Question Answering over Social Media. In *WWW '08*, pages 467–476, 2008.
4. P. B. Brandtzæg and J. Heim. User loyalty and online communities: why members of online communities are not faithful. In *INTETAIN '08*, pages 1–10, 2007.
5. B. Butler, L. Sproull, S. Kiesler, and R. Kraut. Community effort in online groups: Who does the work and why? *S. Weisband. (Ed.), Leadership at a distance.*, pages 171–194, 2007.
6. D. Constant, L. Sproull, and S. Kiesler. The Kindness of Strangers: The Usefulness of Electronic Weak Ties for Technical Advice. *Organization Science*, 7(2):119, 1996.
7. C. Dellarocas. The Digitization of Word-of-Mouth: Promise and Challenges of Online Reputation Mechanisms. *Management Science*, 49(10):1407–1424, 2003.
8. F. M. Harper, D. Raban, S. Rafaeli, and J. A. Konstan. Predictors of Answer Quality in Online Q&A Sites. In *CHI '08*, pages 865–874, New York, NY, USA, 2008.
9. T. Holloway, M. Bozicevic, and K. Börner. Analyzing and Visualizing the Semantic Coverage of Wikipedia and its Authors. *Complexity*, 12(3), 2007.
10. J. Jeon, W. B. Croft, and J. H. Lee. Finding Similar Questions in Large Question and Answer Archives. In *CIKM '05*, pages 84–90, 2005.
11. P. Jurczyk and E. Agichtein. Discovering authorities in question answer communities by using link analysis. pages 919–922, 2007.
12. G. T. Kim. KOREA: Naver's Knowledge Search Database Tops 60 Million . *Asia Media Media News Daily.*, March 2007. Retrieved June 2, 2008.
13. M. Miles and H. A. M. *Qualitative Data Analysis*. Sage Publications Inc, 1994.
14. J. Preece, B. Nonnecke, and D. Andrews. The Top Five Reasons for Lurking: Improving Community Experiences for Everyone. *Computers in Human Behavior*, 20(2):201–223, 2004.
15. A. M. Rashid, K. Ling, R. D. Tassone, P. Resnick, R. Kraut, and J. Riedl. Motivating participation by displaying the value of contribution. In *CHI '06*, pages 955–958, New York, NY, USA, 2006.
16. F. Viégas, M. Wattenberg, and K. Dave. Studying Cooperation and Conflict Between Authors with History Flow Visualizations. *CHI2004*, 2004.
17. E. Wegner. *Communities of Practice: Learning, Meaning, and Identity*. New York: Cambridge University Press, 1998.
18. H. T. Welser, E. Gleave, D. Fisher, and M. Smith. Visualizing the Signatures of Social Roles in Online Discussion Groups. *Journal of Social Structure*, 8(2), 2007.
19. Wikipedia. Knowledge Search — Wikipedia, The Free Encyclopedia, 2004. [Online; accessed 1-June-2008].
20. J. Yang, L. A. Adamic, and M. S. Ackerman. Crowdsourcing and Knowledge Sharing: Strategic User Behavior. In *EC'08*, pages 246–255, 2008.
21. J. Yu, Z. Jiang, and H. C. Chan. Knowledge contribution in problem solving virtual communities: the mediating role of individual motivations. In *SIGMIS-CPR '07*, pages 144–152, 2007.
22. J. Zhang, M. Ackerman, and L. Adamic. Expertise Networks in Online Communities: Structure and Algorithms. *WWW2007*, pages 221–230, 2007.

APPENDIX: THE GURU MEASURE

Let $b_i = 1$ if the user provided the best answer to question i and 0 otherwise, and n_i be the number of answer to question i. We exclude those questions where $n_i = 1$ because there is no point of comparison about the quality of the answer. A user providing m answers would be expected to give x best answers, where $x = \sum_{i=1}^{m} \frac{1}{n_i}$. This probability takes into account the number of other users answering each question. The guru measure $\gamma = \frac{(\sum_{i=1}^{m} b_i) - x}{x}$, then indicates whether a user's performance is better or worse than chance.