Understanding Localness of Knowledge Sharing: A Study of Naver KiN "Here"

Sangkeun Park, Yongsung Kim, Uichin Lee, Mark S. Ackerman†
Department of Knowledge Service Engineering, KAIST
†School of Information and Department of Computer Science and Engineering, University of Michigan
sk.park@kaist.ac.kr, yongsung7@gmail.com, uclee@kaist.ac.kr
†ackerm@umich.edu

ABSTRACT
In location-based social Q&A, the questions related to a local community (e.g. local services and places) are typically answered by local residents (i.e. who have the local knowledge). In this work, we wanted to deepen our understanding of the localness of knowledge sharing through investigating the topical and typological patterns related to the geographic characteristics, geographic locality of user activities, and motivations of local knowledge sharing. To this end, we analyzed a 12-month period Q&A dataset from Naver KiN "Here" and a supplementary survey dataset from 285 mobile users. Our results revealed several unique characteristics of location-based social Q&A. When compared with conventional social Q&A sites, Naver KiN "Here" had very different topical/typological patterns. Naver KiN "Here" exhibited a strong spatial locality where the answerers mostly had 1-3 spatial clusters of contributions, the topical distributions varied widely across different districts, and a typical cluster spanned a few neighboring districts. In addition, we uncovered unique motivators, e.g. ownership of local knowledge and sense of local community. The findings reported in the paper have significant implications for the design of Q&A systems, especially location-based social Q&A systems.

Author Keywords
Location-based Social Q&A; Localness of Knowledge Sharing

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION
Location-based social Q&A allows users to directly tap into the knowledge of local people to obtain information related to a particular geographic location or area (e.g. insider tips about the best places to go). When compared with the traditional social Q&A services such as Yahoo! Answers and Naver KiN, one significant distinction of location-based social Q&A is the localness of knowledge sharing: the questions that are related to a local community (i.e. primarily about local places and services) are answered by those who have the local knowledge (e.g. current local residents). As will be seen from our findings, this leads to very different usage and distinct design implications compared to traditional social Q&A services.

Researchers have strived to build location-based online social platforms that facilitate social interaction and local knowledge sharing; e.g. SocialSearchBrowser [10] and MicroBlog [15]. Also, recently commercial platforms have been designed for location-based social Q&A, such as Locql [4], LocalMind [2], LocalUncle [3], and Naver KiN "Here" [5]. Understanding the localness of knowledge sharing is critical when designing the key features of location-based social Q&A; e.g. categorizing questions/answers, retrieving relevant answers, routing questions, and motivating user contributions. However, our understanding of the localness of knowledge sharing remains limited because prior studies have primarily focused on evaluating prototype systems and drawing design implications with small-scale user studies [10, 11, 15].

In this paper, we deepen the understanding of the localness of knowledge sharing through analyzing a large-scale, longitudinal dataset and surveying current users from Naver KiN "Here", a mobile app for location-based social Q&A in Korea. We consider the following research questions to investigate the key aspects on localness:

- First, are topical/typological patterns related to geographic characteristics, and how?
- Second, how are users geographically focused in their asking/answering activities?
- Third, what are the answer motivations that are unique in location-based social Q&A?

To the best of our knowledge, our work is the first of its kind to investigate the localness of knowledge sharing through analyzing a large-scale real-world dataset.

To answer the research questions, we crawled a 12-month period Q&A dataset from Naver KiN "Here", and we performed supplementary surveys with 285 Naver KiN "Here" users. Using these datasets, we first performed topic analyses using

MobileHCI '14, September 23 - 26 2014, Toronto, ON, Canada

http://dx.doi.org/10.1145/2628363.2628407

different geographic scales (e.g. district/city) and evaluated how the geographic characteristics were reflected in the Q&A activities. We analyzed the district/city-level activity patterns of users in order to gain insights into their geographic focus, and performed spatial clustering analyses to quantify the geographic locality of user contributions. In addition, we performed content analysis of the user survey results to identify the user motives of local knowledge sharing that are unique to location-based social Q&A.

Our primary findings are summarized as follows. First, Naver KiN "Here" had very different typical/typological patterns when compared with traditional social Q&A sites. The factual information seeking was surprisingly high (72.2%), followed by recommendations (17.9%). The overall topic distribution had a significantly higher level of travel and lifestyle information seeking topics, which are often localized in nature. We found that topical distributions varied widely across different districts, as were appropriate for the areas in question. Second, there was strong spatial locality of user contributions where answerers primarily focused on 1-3 spatial clusters. These clusters were closely related to the users' life experiences (e.g. current/former home as well as work and school); many questions required very specific local knowledge that is difficult to answer without local experience. A typical cluster spanned only a few neighboring districts; and its geographic size was much smaller than that found in analyses of Wikipedia knowledge sharing [16, 24]. Third, we found unique motivators for local knowledge sharing, i.e. ownership of local knowledge (competence about local knowledge learned over many years) and sense of community (membership and fulfillment of needs). These are all different from those reported for "normal" social Q&A systems. Finally, we discuss several practical system design implications, such as geographic information retrieval, question routing, and user contribution encouragement.

BACKGROUND AND RELATED WORK

Conventional Social Q&A

Characteristics of social Q&A in general have been heavily studied. For example, Kim et al. [19] classified questions as soliciting facts, opinions, and suggestions. As another example, in their analysis of Yahoo! Answers, Adamic et al. [7] found that user participation varied widely depending on the topic (and are skewed), and that knowledge sharing patterns across different topic categories existed (e.g. experts in different domains that help one another). Similarly, Nam et al. [28] demonstrated that Naver KiN users' level of participation is highly skewed, intermittent, and that the expected level of their expertise is lower than that found in online help forums. Our work differs from the prior studies in that we investigate the localness of knowledge sharing by analyzing the topical/typological patterns related to geographic characteristics, and geographic locality of user activities.

The motivation behind answering questions is largely dependent on a mixture of intrinsic factors (e.g. enjoyment, and feelings of gratitude and respect) and extrinsic factors (e.g. monetary rewards, reputation systems, etc.) [30]. Nam et al. [28] demonstrated that the motivation for answering in Naver KiN results from both intrinsic and extrinsic factors, i.e. altruism is the leading factor, followed by business motives, learning, hobbies, and earning points. In pay-for-answer Q&A sites, researchers have reported that financial incentives and social factors are the key motivators [13, 31, 32, 23]. In this work, we extend the prior studies by identifying unique motivators for local knowledge sharing.

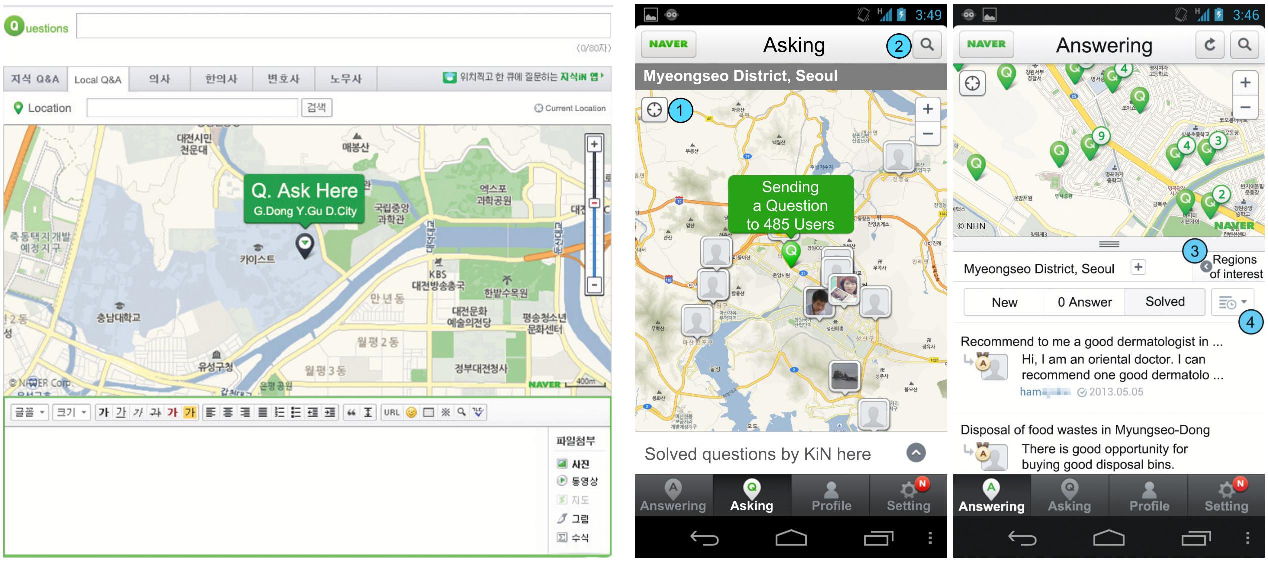
Service	Message Access	Question Delivery	Message Content	Delivery Target	Filtering
GeoNote[14]	Remote *In-situ*	Pull Push (geocast)	Text/MM /Map	Area/ Thing	Topic/ Proximity
SocialSearch Browser[10]	*In-situ*	Pull	Text/Map	Area	Topic/ Proximity
Micro-Blog[15]	Remote *In-situ*	Pull Push (geocast)	Text/MM /Map	Area	Topic/ Proximity
CityFlocks[9]	Remote *In-situ*	Pull Push (unicast)	Text/Map	Area/ Person	Proximity
LocalUncle[3]	Remote *In-situ*	Pull Push (multicast)	Text/Map	POI/City	N/A
Twitter[29] **(TSA Tracker)**	Remote *In-situ*	Pull Push (multicast)	Text/Map	Area/ Person	N/A
Locql[4]	Remote	Pull	Text/Map	POI/City	Interest Areas
Yahoo! Answers[6]	Remote	Pull	Text/Map	City	Topic
Naver KiN "Here"[5]	Remote *In-situ*	Pull Push (geocast)	Text/Map /Map	Area	Interest Areas

Table 1: Classification of location-based social Q&A

Mobile Information Seeking and Local Searches

Recently, mobile information seeking has received significant attention. In their diary study, Sohn et al. found that approximately 38% of mobile information needs have local intent and 72% were prompted by explicit contextual factors, including activity, location, time, and conversation [33]. Similarly, Church et al. demonstrated that over 40% of entries were location-based and 67% of entries were generated when users were away from their familiar contexts [12]. Henrich and Ludecke analyzed the key properties of geographic information needs from the perspective of geographic information retrieval [17], and they found the intention of a geo reference (e.g. to perform activities and obtain facts in a given location), geo-coverage of relevant documents, the shape of geo information needs (point/region, near/within), and the current location of a user.

Search engines are frequently used as information channels for mobile information seeking; this usage is called *local searches*. Researchers have analyzed the search engine log data in order to understand the usage of local search. Gan et al. used an extended version of web query classifications and demonstrated that 12.9% of the overall traffic in an AOL 2006 dataset had local intent. In recent years, the overall portion of geo queries has increased sharply (possibly due to increased smartphone use); Google and Bing reported that geo queries comprise 50% (2012) and 53% (2010), respectively, of their mobile search traffic [34]. Furthermore, Jones et al. [18] analyzed Yahoo! query logs and reported the characteristics of their geo queries: the distance between the home location and queried location (30% were within 100 km), and query reformulation patterns. Teevan et al. [36] conducted a survey of local search usage and found that local searches tended to be highly contextual in that current location and that time constraints and social factors had a significant impact on usage behavior. In this work, we complement earlier studies on



(a) Web (b) App

Figure 1: Naver Local Q&A interface. (Web App). In the case of App, the numbered labels are (1) GPS, (2) search, (3) region of interest, and (4) sort by distance/time. The original text was translated into English for the easier of illustration.

geographic information needs [10, 11, 17, 33] using a real-world dataset; the existing studies examined limited datasets such as web search logs [17] and small-scale user data (e.g. diaries, user study) [10, 11, 33].

Location-based Social Q&A

In Table 1, we summarize the existing location-based social Q&A systems based on their design features, i.e. message access (remote, *in situ*), delivery (push, pull), message content (text, multimedia: MM), user interface (text, map), message destination (location, thing, Point of Interest (POI), city), and message filtering (topic/area subscription, proximity). In the existing systems, message access modes are the key differentiating factor. Remote access means that a user can post and retrieve messages in a remote location; *in situ* access means that a user can retrieve relevant messages only when the user approaches the area of interest. *In situ* access is suitable for facilitating information sharing among local residents, such as in SocialSearchBrowser. Questions can be delivered to target users in various ways: pushing questions to people in a regional area as in Naver KiN "Here" and MicroBlog (geocast), sending questions to subscribed people as in LocalUncle (multicast), and sending questions directly to a person as in CityFlock (unicast). For location-based social Q&A services, text and interactive maps can be used for question/answer navigation [11]. The existing systems support various filtering mechanisms ranging from place/area/topic subscription to proximity sensitive filtering. In this paper, we study localness of local knowledge sharing by studying Naver KiN "Here", which has representative design features.

Localness of Information Sharing Behavior

Prior studies have explored the localness of information sharing such as influence of regional characters on social media usage [21], and user contribution behavior of local content in online social systems [16, 24, 38]. Kulshrestha et al. demonstrated that shared national, linguistic, and cultural backgrounds had a significant impact on Twitter usage (e.g. social links and information exchanges) [21]. Yardi and Boyd investigated how local community members used Twitter to share and exchange information about local events [38]. A large percentage of Wikipedia users only edit a few geo-pages (e.g. city, school), and the geo-locations of the edited geo-pages tended to be localized within a 100 km radius [24]. Hecht and Gergle compared the mean contribution distance of Flicker posts and Wikipedia edits (geo-pages) and demonstrated that the Wikipedia edits had a longer contribution distance than Flickr posts [16]. Prior studies have provided a valuable insight into the localness of knowledge sharing. However, knowledge sharing in location-based social Q&A differs significantly from that in Twitter and Wikipedia because it is designed to resolve everyday life geographic information needs with the help of local experts. Our goal is to study the localness of location-based social Q&A through conducting topic analyses and spatial clustering analyses of a large-scale real-world Q&A dataset.

INTERACTION IN LOCAL Q&A

We investigate Naver KiN "Here", which is a location-based social Q&A that was released in December 3, 2012. Naver KiN "Here" extends its existing service called Naver KiN Local to mobile services. For the location-based Q&A, the registered users can interact with the content on either the mobile app or webpages. Both interfaces allow the use of an interactive map to represent the location/region of interest, and the questions are categorized by region names (based on administrative divisions) instead of content topic categories.

The question asking is similar to conventional social Q&A sites such as Yahoo! Answers and Naver KiN. The key difference is that instead of choosing the content topic, users are asked to select the regions using an interactive map in order to categorize the question. In the Naver KiN "Here" app, the user also interacts with a map, which also displays the active app users who have answered more than five questions in a given area. A posted question will be shown on the webpage and app simultaneously. Unlike the web access, the questions posted using the mobile app are pushed to users whose regions of interest match those of the questions. The app users can switch off the push notifications. The registered users can post answers via the app or web interfaces. The front page of the Naver KiN "Here" app (Figure 1) presents an interactive map and a list of questions related to the current location based on GPS; however, the app and web interfaces differ for answering questions. On the web, the questions are ordered based on recency rather than proximity, and users can browse all questions and filter them based on regional hierarchy. In the app, the questions are ordered by proximity to the current location by default. When a user pans or zooms with the map, the list is automatically updated. Depending on the zooming level, the questions in a close geographic area are aggregated into a question bubble with a number inside that indicates the number of questions. Both web and app users can subscribe to the regions of interest for question filtering. For app users, the system automatically pushes new questions based on the subscription information. Naver KiN "Here" users receive points for asking and answering activities. Furthermore, the mobile app users can acquire various types of badges, ranging from app installation and continued use to expert answerer status, which can be obtained through providing fast responses and accurate answers.

Before detailing our results, an illustration of the administrative divisions in Korea is important because it differs to other countries. The administrative divisions have four levels: province ("Do"; there are nine provinces in Korea), city ("Si"; typical size of 100-1000 km^2), sub-city ("Gu"; typical size of 10-100 km^2), and district ("Dong", typical size of 1-10 km^2). The levels are more fine-grained than those of western countries, e.g. the USA, which are typically composed of three levels: state, county/shire, and city/town/village. People typically refer to city/sub-city/district names when searching for or speaking about places in Korea.

METHODOLOGY

Our data collection method was comprised of two complementary parts: automated crawling of publicly accessible local question-answer pairs on Naver KiN "Here" local Q&A and a web-based survey. The data were obtained for the period from December 17, 2012 to December 31, 2013; a total of 508,334 questions and 567,156 answers were obtained. The crawled data enabled comprehensive statistics on user behavior, while the web-based survey aided in obtaining an in-depth understanding of the observed usage behavior.

We supplemented the question and answer data with the web-based surveys of Naver KiN "Here" local Q&A users. In August 2013, we sent a questionnaire link via Naver email to 4,557 users who had asked/answered at least one question via the Naver KiN "Here" app. The link lasted for a week, and the total number of participants was 285 after removing duplicate and erroneous responses. The low response rate may have been caused by the fact that Naver emails may not be

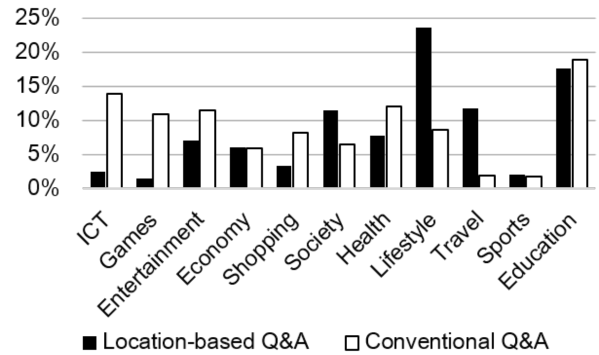


Figure 2: Topic distribution

users' primary email accounts. At the end of the survey, we randomly selected 50 participants and rewarded an online gift voucher worth $10. The survey questions asked about demographics (e.g. age, gender, occupation), app settings, and user motivations. The survey data demonstrated that 59.7% of participants were males and 40.3% were females. Participants were predominantly in their teens (10-19 years; 30%) and in their 20s (43%); the remainder were in their 30s (14.4%), 40s (10.8%), and 50s (1.8%). The occupations were quite diverse: middle/high school students (29.8%), college/graduate students (23.3%), financial/service workers (12.1%), lawyers (0.3%), designers/artists (3%), homemakers (2.3%), architects (2.3%), engineers (7.2%), miscellaneous area workers (11.8%), and unemployed (6.2%).

TOPICAL LOCALITY IN KNOWLEDGE SHARING

We performed content analyses in order to characterize the topical patterns (overall distributions and geographic/typological differences). We used the topic categories from Naver KiN: information and communication technology (ICT), games, entertainment/art (e.g. TV, radio, movies), economy (e.g. banking, tax, real estate), shopping, society (e.g. laws, politics, culture, governments), health, lifestyle (e.g. food, pets, cars), travel/transportation, sports, and education. Unlike traditional social Q&A, Naver KiN "Here" does not have a topic field as questions are categorized based on location. We automatically classified the topic categories as follows: for a given question we extracted the key words using a Korean parser called KKMA (Kokoma Korean Morpheme Analysis) [1]. Then, we searched the extracted keywords using Naver KiN in which the question askers manually select the relevant topics when posting questions. From the top 100 search results, the most frequent topic category was selected as the topic category for the question. In order to confirm whether the automatic classification provided accurate results, we manually coded the topics of 100 randomly selected questions, and we measured the inter-rater agreement (between automatic and manual classifications) using Cohen's Kappa statistic. The measured value for the topic classification was $k = 0.87$, which indicates substantial agreement.

In Figure 2, we present the topic distributions of the location-based social Q&A (Naver KiN "Here") as well as conventional social Q&A (Naver KiN). Key differences were found in the ICT, games, lifestyle, and travel/transportation

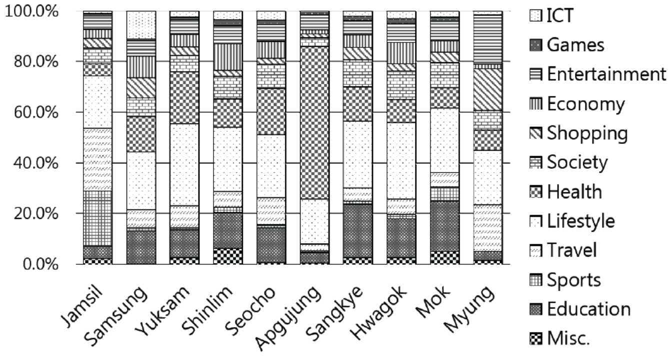


Figure 3: Topic distribution of top 10 districts

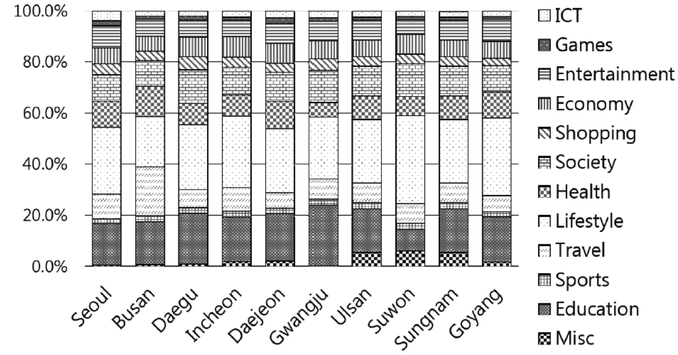


Figure 4: Topic distribution of top 10 cities

topic categories. In the location-based social Q&A, lifestyle (23.7%) and travel (11.7%) were dominant, whereas ICT (2.5%) and games (1.5%) were minimal. Our results confirm that there is a significant difference in the topical distributions between location-based social Q&A and conventional social Q&A.

Next, we investigated whether the characteristics of a region were reflected in the Q&A usage. Identifying such characteristics is beneficial when building geographic information retrieval systems. We found that, in general, the geographic characteristics were well reflected and that some patterns of topical locality existed. In addition, the topical distributions were largely dependent on the size and functional complexity of the region.

In order to understand how regional characters are reflected in the Q&A usage, we first analyzed the topic distributions of the top 10 districts in Seoul (Figure 3). Overall, the regional characteristics were well reflected in the district level questions. For example, the Jamsil district, which is well known for its sports stadium, had a large percentage of sports questions (22.2%). The Apgujeong district had a high percentage of health questions (60.4%) because it is famous for its medical area that is densely populated with plastic surgery clinics. Therefore, we hypothesized that such observations were partly related to the concept of zoning in urban planning; zoning is a method of urban planning that prevents new developments from interfering with the existing residents or businesses and preserves the "character" of a community [8]. This type of planning is widely adopted in most developed nations, and local municipalities in Korea also abide by the national zoning guidelines.

Furthermore, we examined whether there were significant differences in the topical distributions in different geographic scales. In Figure 4, we plotted the topic distributions of the top 10 cities ranked in terms of the number of questions. Unlike the questions in the districts, there were only minor variations in categories across the different cities. Note that there were a few cities that have significant tourist attractions and therefore have distinctive topic distributions with higher percentages of shopping, entertainment, and travel categories, e.g. Jeju Island, which was voted as one of the New7Wonders of Nature in 2011, had significantly high number of questions/answers related to travel.

We also classified the types of questions using previously studied categories [19, 22] where questions were classified into information, suggestion, and opinion questions. Information questions are used to find specific facts; suggestion questions are used to seek recommendations; opinion questions are used to survey other people's thoughts or preferences. In order to analyze these question types, we randomly selected 1000 questions and coded the questions using two external raters. They first coded 200 questions together, and then separately coded the remaining 800 questions (i.e. 400 questions each). These common questions were used to measure the inter-rater agreement using Cohen's Kappa statistic. The measured value for the type classification was $k = 0.84$, which indicates substantial agreement. We found that the information and suggestion question types were dominant, and they comprised 72.2% and 17.9% of the questions, respectively, followed by opinion questions (6.4%) and miscellaneous questions (3.5%). This type distribution differs significantly from other social Q&A sites [19, 22, 27]. In mobile social Q&A, information questions comprised 51.1% of all questions, followed by suggestion (23.4%) and opinion (18.8%) questions [22]. Thus, location-based social Q&A has a significantly higher percentage of information questions and a lower percentage of opinion questions when compared with the conventional social Q&A. Considering the high percentage of factual questions, it would be beneficial to archive them for local searches.

SPATIAL LOCALITY OF USER ACTIVITIES

City/District-level User Activities

We analyzed the dataset in order to understand the geographic focus of the users' activities (asking/answering). In order to capture the degree of geographic focus by an asker/answerer, we used an entropy measure, i.e. the lower the entropy, the higher the level of focus on certain regions. Considering an answerer i who made p_k percentage of answers for region k, the answerer's entropy is given as $-\sum_k p_k \log_2 p_k$. If a user only answered for a single region, e.g. region j (i.e., $p_j = 1$ and $p_i = 0$ for all i other than j), the entropy value is zero. The entropy is maximized when any region is equally likely to be asked by a user (i.e. a uniform distribution). We only considered users who asked/answered more than ten questions/answers for the entropy calculation. In addition, we used the two different region levels of district and city. Note that due to privacy concerns, Naver only revealed three characters

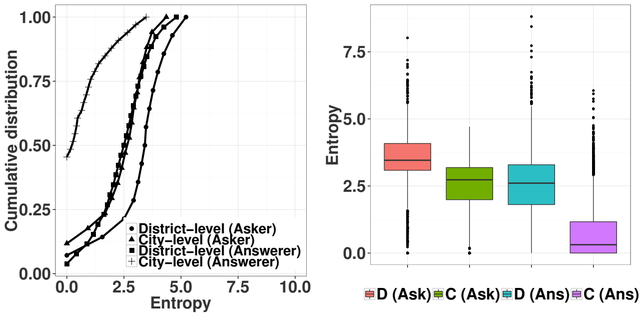


Figure 5: CDF and Boxplot of entropy (C:City, D:District)

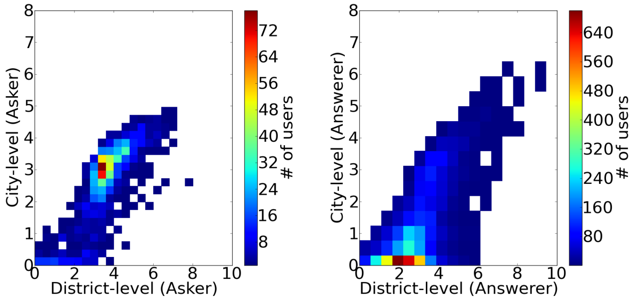


Figure 6: Heat map of entropy

of the asker's ID. Nonetheless, we could uniquely identify a user because the ID was accompanied with two other statistics, the question closing rate and answer acceptance rate. Those who answered more than ten questions tended to have diverse numbers.

In Figure 5, we present the cumulative distributions and corresponding boxplots in parallel. The results demonstrated that, in general, the asking activities had higher entropy values than the answering activities: asking at the district/city levels had means of 3.5/2.5 and answering at the district/city levels had means of 2.6/0.8. In order to better understand the geographic containment (e.g. many districts but in the same city) for each user, we arranged a pair of entropy values, i.e. (district, city), and drew a heat map for both activities as seen in Figure 6. This heat map allows better visualization of the geographic focus and containment. For the asking activities, the figure demonstrates that the center area has bright colors. This shape indicates that the askers' activities typically span multiple districts and cities. Furthermore, the core of the shape had an entropy value of 3 at the city level, which represents eight cities, if a uniform distribution is assumed. In practice, the distribution was usually skewed, and the actual number was significantly larger. For the answering activities, we found that the overall shape and color differed significantly from that of the asking activities. In particular, the cores were located at the bottom in the city level, whereas the district levels were widely scattered, which indicates the localness of answering activities.

In order to understand localness, we asked the survey participants to report 1) the number of their selected regions of interests in the app, and 2) a list of those names as well as the reason for each choice. The number of selected regions of interests is presented in Figure 7. The mean number of

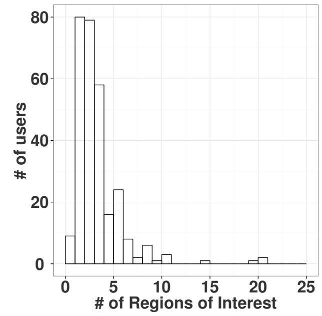


Figure 7: Number of regions of interest per user

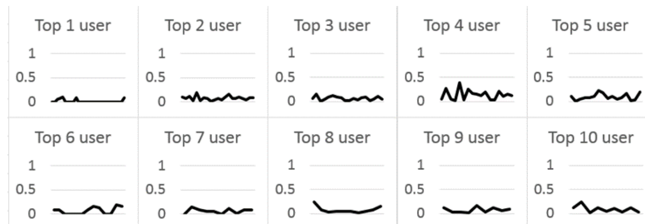


Figure 8: City-level entropy changes of top 10 answerers

districts was 2.9 (SD = 3.2) and the maximum number was 35. Regarding the second part, there were 142 valid answers with a detailed list of interest regions and the reasons for their choices, and this led to 317 annotated regions of interest. The major categories of areas included home, work/school, and downtown areas. The manual classification results indicated the following: home (93.7%), nearby home (16.2%), previous home (14.8%), school (23.9%), nearby school (0.7%), previous school (0.7%), work (28.9%), nearby work (2.1%), previous work (3.5%), and downtown (24.6%). When those who had their home/work in the same region, they were counted twice. Miscellaneous regions of interest (25 regions) included churches, hobby places, parents' houses, relatives' houses, and tutoring institutes. The results aided in understanding the relationship between a user's local connection and knowledge. However, despite local familiarity (home, work, school), the survey participants reported that perceived percentage of answerable questions was 37% on average (SD = 24%). As shown later, many of the questions required very specific knowledge based on local experiences, e.g. *"In Daejeon, are there any places that I can buy big dumplings after midnight?"* and *"In Wonju, please let me know where I can buy less expensive medium and large size vases."*

In Figure 8, we plotted the city-level entropy changes of the top 10 answerers. For a series of answers by a user, we calculated the entropy in each block of 50 consecutive questions. Then, we plotted the magnitude of the entropy differences between two consecutive blocks. The figures clearly demonstrate that the changes in the entropy values were quite small. This indicates that the city selection strategies did not significantly change over time: those who had low/high entropy values would continue to have low/high values.

We divided the users into two groups based on the city-level entropy values. If the city-level entropy was greater than 2, we assumed that the user's activities were geographically scat-

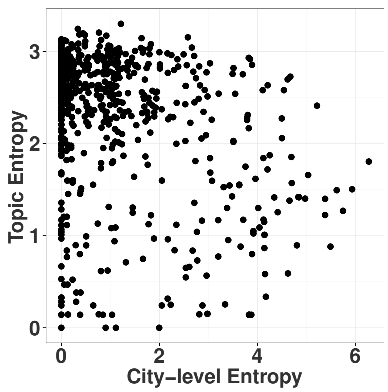


Figure 9: City-level entropy vs. topic entropy of a user

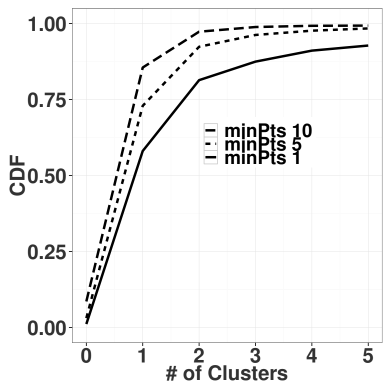


Figure 10: Number of clusters per user

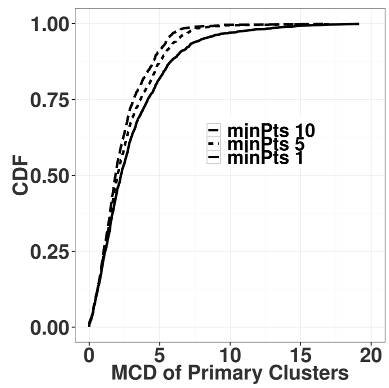


Figure 11: Distribution of MCD values

tered (GS group); otherwise, we assumed that the user's activities were geographically focused (GF group). From this categorization, 81% of heavy users had city-level entropy values of less than two (mean = 0.48), and the actual number of contributed cities was typically less than four.

However, the remaining 19% of the heavy users had activities that were scattered. The GS group's city-level entropy (mean = 3.67, SD = 1.11) was significantly higher than the GF group's (mean = 0.57, SD = 0.57). In Figure 9, we present a scatter plot: each dot represents a user's city-level entropy and topic entropy. We manually investigated the answerers in the GS group in order to understand what types of local knowledge they were providing. Those who had high topic entropy values were mostly top-ranked answerers in Naver KiN. The manual investigations revealed two types of answerers: active web searchers and province-level local experts. The active web searchers typically provided local answers that were easily searchable on the web (mostly factual) such as transportation costs, traffic status, and local facilities. The province-level experts were actively contributing to a number of cities within a province; however, detailed local answers were more skewed to a few familiar cities. Interestingly, those who had low topic entropy values mostly promoted their business (e.g. clinics, lawyers, online shopping malls) or online communities (e.g. volunteering). It appears that these users tended to copy and paste similar, but lengthy, general answers about specific topics. For example, a medical doctor answered local questions about recommending local clinics for a specific cosmetic surgery, but the lengthy answers only provided general information about that surgery.

Spatial Cluster Analysis

The city/district-level analyses assisted in gaining a good insight into spatial locality, but they had several limitations: only coarse-grained spatial locality could be found and they did not provide a user's locus of contributions. Therefore, we extended our locality analyses through performing spatial cluster analyses. We ran a density-based clustering algorithm (called DBSCAN) where the radius (known as eps) was set to 10 km in order to merge nearby districts, and clusters were repeatedly expanded as long as the number of minimum points (minPts) is greater than thresholds (e.g. minPts = 1, 5, 10). When we gave minPts threshold of 1, we repeatedly expanded a cluster whenever there are at least two questions within 10 km radius. Due to the requirements of the minimum points for the spatial cluster analysis, we only considered heavy users whose number of answers was equal to or greater than 30; this resulted in 1492 users. Note that the city/district-level entropy distributions with these users were consistent with the earlier results.

Figure 10 presents a Cumulative Distribution Function (CDF) of the number of clusters with different minPts thresholds. A majority of users had a single cluster; in particular when minPts = 1, more than 50% of users had a single cluster, and more than 75% of users had less than two clusters. The mean values of the minPts = 1, 5, and 10 were 2.3, 1.5, and 1.2, respectively. The maximum values of the minPts = 1, 5, and 10 were 101, 61, and 40, respectively; indeed, there were some users whose contributions were widely scattered across many cities.

We used the mean contribution distance (MCD) to measure the localness of the user contributions in each cluster [16]. For a given cluster of a user, the cluster's MCD was defined as $\sum_{i=1}^{n} d(C, c_i)/n$, where C is the centroid of the cluster, c_i is the location of an answered question, $d(C, c_i)$ is the Euclidean distance between two points (C and c_i), and n is the total number of answers in a cluster. In Figure 11, we plot the CDF of the MCD values from the users' primary clusters (i.e. the cluster with the largest number of answers). The mean MCD values for minPts = 1, 5, and 10 were 3.1, 2.6, and 2.3 km, respectively. This indicates that the primary clusters' MCD values covered a few nearby districts, e.g. home and nearby home, and work and nearby work.

In Figure 12, we plot the cluster-level entropy values. Because the number of clusters was very small, a majority of users had a zero entropy value. Then, we compared the city-level entropy and cluster-level entropy values; Figure 13 presents the scatter plots with minPts = 1. The figure demonstrates that there is a linear relationship between these two variables. The cluster-level entropy is smaller than the city-level entropy because the DBSCAN removed the noisy clusters that had points less than or equal to the minPts. Interestingly, those users whose cluster entropy values were zero, but had high city-level entropy values, were mostly business promoters.

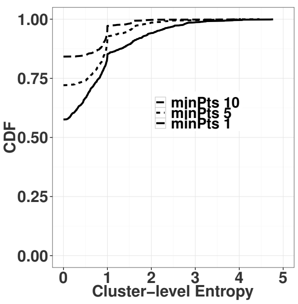


Figure 12: CDF of cluster-level entropy values

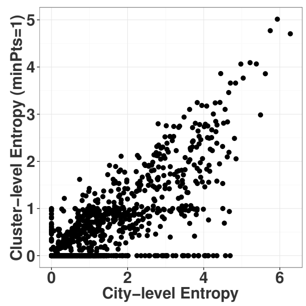


Figure 13: City-and cluster-level entropy (minPts=1)

ANSWER MOTIVATIONS

In order to understand the answer motivation, we asked the participants to write in detail why they answered questions posted in Naver KiN's Local Q&A using an open-ended question. From the answers, we extracted all motivators and performed affinity diagramming. The following major themes were derived: knowledge exchange (24.9%), altruism (18.2%), ownership of local knowledge (10.1%), points (9.8%), pastime (9.2%), and sense of community (7.0%). Miscellaneous themes included business promotion and learning. Unlike the existing results on social Q&A that reported intrinsic (altruism, enjoyment), external (points), and social (knowledge exchange) motivators [28, 30], we found two unique motivators in the location-based social Q&A: ownership of local knowledge (competence about local knowledge learned over many years) and sense of community (serving the information needs of other community members).

Researchers have demonstrated that when people believed they owned information, they were more likely to engage in knowledge sharing [37]. This result can be attributed to individuals' internal satisfaction derived from sharing their knowledge with others. As they have lived in that location for a long time, we expected that the local residents gained a strong ownership of local knowledge, ranging from specific goods/services to distinctions between places and place popularity [25]. In our survey, the ownership of local knowledge was clearly noted as follows. One user stated, *"Because I know everything about my town as I have lived in my town for a long time"*; another user stated his experience by saying, *"I have been living here for 20 years and went to schools in this region. With this knowledge, I'm sure I can help answer other people's questions. That's why I started answering here."*

Another theme was the sense of community. McMillan and Chavis [26] defined the sense of community as "a feeling that members have of belonging, a feeling that members matter to one another and to the group, and a shared faith that members' needs will be met through their commitment to be together". Some key factors in the sense of community include *membership* and *fulfillment* of needs. Note that the sense of community in location-based Q&A differs significantly from that in social network services (and the Q&A therein) because the interpersonal relationships among answerers are weak. In our survey, we found that many users stated "my regions", which is a good indicator that they regard themselves as members of their local community. Furthermore, for those who have a strong membership attachment, they are eager to invest their time in the local community. One user stated, *"I have been answering questions in the Laws section. But after I noticed that the areas that I live in and know well did not receive answers, I decided to answer questions. I subscribed to question delivery from those areas and started answering questions."* Another user commented, *"I think kind and sincere answering is one of the representative images of the area, and I want to build a good image of my area."*

The second important element of the sense of community is the fulfillment of needs. Our participants concurred that they wanted to serve the information needs within their communities (for both members and visitors). A user expressed their feeling as follows: *"Knowing that I can help other local community members makes me feel really great."* In addition, responsibility as a community member was expressed: *"I was born here, and I know [this region] very well. I feel fresh because people from other areas often ask questions. I feel like I should take care of them just like taking care of a baby."* Some users fulfilled the needs in the expectation of receiving useful help in return later, e.g. *"By exchanging questions, I can receive help from other people when I visit other areas; I can also give help to other people when they visit ours."*

DISCUSSION

We present practical design implications based on our findings and discuss the limitations of this work.

Leveraging Topical/Typological Patterns

Our topical and typological analyses results provide insights into the design dimensions of location-based social Q&A, e.g. category refinement, question routing, and related answer searches. When compared with conventional social Q&A sites [7, 23], the overall topic distribution of location-based social Q&A had a higher level of travel and lifestyle information seeking, but a lower level of computer related questions. Our manual investigation of sample questions revealed that local questions were primarily related to local services (e.g. housing, entertainment, shopping, eating/drinking, education, transportation) and local attributes (e.g. culture, history, geography).

Recall that KiN "Here" users are asked to subscribe to regions of interest. The users will be automatically notified of questions sent to these regions via push messages. Because our participants reported that 37% of the questions were answerable on average, it would be beneficial to include an additional option for selecting topical categories that are customized for location-based social Q&A. This will assist in filtering the push notifications based on topics of interest to lower interruption overhead.

Our typological analyses demonstrated that factual information seeking was high (72.2%), followed by recommendations (17.9%); it also demonstrated that the topic distributions varied widely across different districts. This observation implies that a location-based social Q&A dataset could be effectively archived for local searches. For some areas, similar questions were posted repeatedly; the localized content analyses and ranking may aid in improving the quality of the retrieved answers. Furthermore, topic clustering algorithms could be applied such that the regional topic characteristics could automatically be extracted and utilized for local search optimization. In some cases, a user would prefer to post a local question via traditional social Q&A services; in this case, it would be beneficial to recognize that it is a local question, and the system could automatically recommend a candidate place for the question routing.

Leveraging the Spatio-Temporal Activity Analyses

Our spatial locality analyses revealed that there was strong spatial locality of user contributions. The answerers primarily focused on 1-3 spatial clusters that were closely related to their life experiences (e.g. current/former home, work, and school), and a cluster typically spanned a few neighboring districts. The mean contribution distance in Naver KiN "Here" was approximately 2-3 kilometers on average, whereas those in Wikipedia were much greater in scale (hundreds of kilometers) [16, 24]—most Wikipedia geo-pages contain "general" information about regions and famous POIs. Our results imply that while a user's subscribed districts are currently considered for push notifications, the radius of question geocasting could be extended slightly to a few neighboring districts. Another implication is that whenever a user subscribes to a new district, we could recommend that the user also subscribe to neighboring districts. Alternatively, during the service subscription process, users could be asked to enter district and city names related to familiar places and to automatically recommend corresponding districts for subscription.

We found that many questions required very specific local knowledge, which were difficult to answer without local experience. Most heavy users were geographically focused. Our spatial-topical entropy analyses revealed that some heavy users who showed high topical diversity were the web searchers (who heavily used geographic information systems such as transportation schedule/status pages), province-level experts (covering nearby local cities), and business promoters (who tended to copy and paste similar content). Providing local search tools that are specialized for popular topics of geographic information seeking will help improve user participation and answer quality. Furthermore, it is possible to identify local business promoters by applying machine learning techniques; several useful features would be the cluster and city-level entropy values, content similarity of posted answers, and the best answer selection rates.

Motivating User Contributions

We found additional motivators that are unique in location-based social Q&A, i.e. ownership of local knowledge and sense of community. These motivators could be leveraged to encourage user contributions and increase user commitment. One immediate method is to use community-level symbols. A key element of membership in a sense of community [26] is a common symbol system. Considering that the existing location-based services often employ badges, we could create community badges and award them to those who are actively participating in that community, which could reinforce their motivation [20].

We could use to leverage local membership and attachment. First, when a question is pushed, framing the request in a way that aligns with the motivators of ownership of local knowledge and sense of community could increase contributions. Second, sending out hyperlocal news about recent regional activities could increase regional members' awareness of the service and may encourage contributions. In addition, visualizing popular/trending topics in location-based social Q&A would help people understand local characteristics (e.g. knowing "hot topics" in my area). This would not only provide significant insights for the local residents about their regions, but it would also assist in people from other regions to understand topical characteristics of the region. Third, introducing regional competition could elicit more contributions; for example, the system could display scoreboards (e.g. answer and selection rates of nearby regions) at different geographic scales (e.g. district- and city-level). Fourth, when a new user starts contributing to the location-based social Q&A, we can automatically classify the user's topical interests and the level of expertise on that topics as in conventional social Q&A [35], and this along with location information can be leveraged to encourage user participation.

Limitations

As with any qualitative or single-site work, the generalizability of this work is limited such that additional work on similar location-based social Q&A services, such as Locql, LocalUncle, and Yahoo! Answers' Local Business, is necessary, which will be part of our future work. In addition, because social networking services can be used to ask questions [26], it would be interesting to see if existing location-based social Q&A services such as Foursquare are used for social Q&A activities.

CONCLUSION

We investigated the localness of knowledge sharing through analyzing the topical and typological patterns related to geographic characteristics, geographic locality of user activities, and motivations of local knowledge sharing. We collected a large-scale real-world dataset from Naver KiN "Here" and conducted a complementary survey of 285 mobile app users. From the analyses, we found that, compared with conventional social Q&A sites, the location-based social Q&A has unique topical and typological patterns that vary widely across different districts. A strong spatial locality of contributions exists around a few spatial clusters related to users' life experiences, which span a few neighboring districts. Localness is well reflected in the motives for local knowledge sharing, i.e. ownership of local knowledge and a sense of community. Finally, we discussed several practical system design implications such as leveraging topical and spatio-temporal activity patterns (e.g. question routing, local searches) and motivating user contributions (e.g. badges, request framing, hyperlocal news).

ACKNOWLEDGMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2012R1A1A1008858).

REFERENCES

1. KKMA. http://kkma.snu.ac.kr.
2. Localmind. http://www.localmind.com.
3. Localuncle. http://www.localuncle.com.
4. LOCQL. http://www.locql.com.
5. Naver Local Q&A. http://kin.naver.com/qna/list.nhn?dirId=12.
6. Yahoo! Answers. https://answers.yahoo.com/.
7. Adamic, L. A., Zhang, J., Bakshy, E., and Ackerman, M. S. Knowledge Sharing and Yahoo Answers: Everyone Knows Something. In *WWW* (2008).
8. Berke, P., and Kaiser, E. J. *Urban Land Use Planning*. University of Illinois Press, 2006.
9. Bilandzic, M., Foth, M., and De Luca, A. CityFlocks: Designing Social Navigation for Urban Mobile Information Systems. In *DIS* (2008).
10. Church, K., Neumann, J., Cherubini, M., and Oliver, N. SocialSearchBrowser: A Novel Mobile Search and Information Discovery Tool. In *IUI* (2010).
11. Church, K., Neumann, J., Cherubini, M., and Oliver, N. The ”Map Trap”?: An Evaluation of Map Versus Text-based Interfaces for Location-based Mobile Search Services. In *WWW* (2010).
12. Church, K., and Smyth, B. Understanding the Intent Behind Mobile Information Needs. In *IUI* (2009).
13. Edelman, B. Earnings and Ratings at Google Answers. *Economic Inquiry* (2012).
14. Espinoza, F., Persson, P., Sandin, A., Nyström, H., Cacciatore, E., and Bylund, M. Geonotes: Social and Navigational Aspects of Location-Based Information Systems. In *Ubicomp* (2001).
15. Gaonkar, S., Li, J., Choudhury, R. R., Cox, L., and Schmidt, A. Micro-Blog: Sharing and Querying Content Through Mobile Phones and Social Participation. In *MobiSys* (2008).
16. Hecht, B. J., and Gergle, D. On the ”Localness” of User-generated Content. In *CSCW* (2010).
17. Henrich, A., and Luedecke, V. Characteristics of Geographic Information Needs. In *GIR* (2007).
18. Jones, R., Zhang, W. V., Rey, B., Jhala, P., and Stipp, E. Geographic Intention and Modification in Web Search. *International Journal of Geographical Information Science* (2008).
19. Kim, S., Oh, J. S., and Oh, S. Best-Answer Selection Criteria in a Social Q&A Site from the User-Oriented Relevance Perspective. *Proceedings of the American Society for Information Science and Technology* (2007).
20. Kraut, R. E., and Resnick, P. Encouraging Contribution to Online Communities. *Building Successful Online Communities: Evidence-based Social Design* (2011).
21. Kulshrestha, J., Kooti, F., Nikravesh, A., and Gummadi, P. K. Geographic Dissection of the Twitter Network. In *ICWSM* (2012).
22. Lee, U., Kang, H., Yi, E., Yi, M., and Kantola, J. Understanding Mobile Q&A Usage: An Exploratory Study. In *CHI* (2012).
23. Lee, U., Kim, J., Yi, E., Sung, J., and Gerla, M. Analyzing Crowd Workers in Mobile Pay-for-Answer Q&A. In *CHI* (2013).
24. Lieberman, M. D., and Lin, J. You Are Where You Edit: Locating Wikipedia Contributors through Edit Histories. In *ICWSM* (2009).
25. Ludford, P. J., Priedhorsky, R., Reily, K., and Terveen, L. Capturing, Sharing, and Using Local Place Information. In *CHI* (2007).
26. McMillan, D. W., and Chavis, D. M. Sense of Community: A Definition and Theory. *Journal of community psychology* (1986).
27. Morris, M. R., Teevan, J., and Panovich, K. What Do People Ask Their Social Networks, and Why?: A Survey Study of Status Message Q&a Behavior. In *CHI* (2010).
28. 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).
29. Nichols, J., and Kang, J.-H. Asking Questions of Targeted Strangers on Social Networks. In *CSCW* (2012).
30. Raban, D., and Harper, F. Motivations for Answering Questions Online. *New media and innovative technologies* (2008).
31. Raban, D. R. The Incentive Structure in an Online Information Market. *Journal of the American Society for Information Science and Technology* (2008).
32. Shah, C., Oh, J. S., and Oh, S. Exploring Characteristics and Effects of User Participation in Online Social Q&A Sites. *First Monday* (2008).
33. Sohn, T., Li, K. A., Griswold, W. G., and Hollan, J. D. A Diary Study of Mobile Information Needs. In *CHI* (2008).
34. Sterling, G. Study: 43 Percent Of Total Google Search Queries Are Local. Search Engine Land, 2012.
35. Sung, J., Lee, J.-G., and Lee, U. Booming Up the Long Tails: Discovering Potentially Contributive Users in Community-based Question Answering Services. In *ICWSM* (2013).
36. Teevan, J., Karlson, A., Amini, S., Brush, A. J. B., and Krumm, J. Understanding the Importance of Location, Time, and People in Mobile Local Search Behavior. In *MobileHCI* (2011).
37. Wang, S., and Noe, R. A. Knowledge Sharing: A Review and Directions for Future Research. *Human Resource Management Review* (2010).
38. Yardi, S., and Boyd, D. Tweeting from the Town Square: Measuring Geographic Local Networks. In *ICWSM* (2010).