# Questions are Content: A Taxonomy of Questions in a Microblogging Environment

**Miles Efron**
Graduate School of Library and Information Science
University of Illinois, Urbana-Champaign
mefron@illinois.edu

**Megan Winget**
School of Information
University of Texas, Austin
megan@ischool.utexas.edu

## ABSTRACT

Microblogging services such as `twitter.com` have become popular venues for informal information interactions. An important aspect of these interaction is question asking. In this paper we report results from an analysis of a large sample of data from Twitter. Our analysis focused on the characteristics and strategies that people bring to asking questions in microblogs. In particular, based on our analysis, we propose a taxonomy of questions asked in microblogs. We find that microblog authors express questions to accomplish a wide variety of social and informational tasks. Some microblog questions seek immediate answers, while others accrue information over time. Our overarching finding is that question asking in microblogs is strongly tied to peoples' naturalistic interactions, and that the act of asking questions in Twitter is not analogous to information seeking in more traditional information retrieval environments.

## Keywords

Twitter, information seeking, social networks, taxonomy, microblogging.

## INTRODUCTION

Although microblogging services are becoming more and more popular among the general population, research into microbloggers' motivations, habits and strategies is in its infancy and our understanding of peoples' information behavior with respect to microblogs remains murky. In this paper we treat a particular aspect of this silence, asking: what role do questions play in the information ecology of Twitter?

We focus on questions for a variety of reasons. First, as we note below, previous research has found that people use Twitter as a *de facto* social search system (Evans et al.,

2008; Hearst et al., 2008). In this context an analysis of question asking will help us understand how Twitter users express information needs. Second, we argue that asking questions is an important, and very common, strategy for generating social interaction via microblogs. Questions, we argue, are a particularly vital form of content on Twitter.

Our contribution is twofold. First, based on a naturalistic analysis of Twitter data we offer a taxonomy of questions in microblogs. Second, we contextualize these articulations through analysis of a large body of tweets. In this analysis we found that statements of information need in microblogs fulfill two purposes: first, as in traditional information retrieval systems, these statements solicit information. Posing questions, in addition to indicating a specific information need, represents an especially valuable type of content for Twitter users.

We argue that inquiry constitutes a particularly effective method of maintaining forward momentum in Twitter conversations. Thus we hypothesize that a strong factor in Twitter's success as a social service lies in the fact that querying in microblogs has particular social value. Asking questions is a common and effective means of maintaining social ties and of articulating personal interests in microblog ecologies. When people ask questions on Twitter they act on complex agendas, searching for information, expressing ideas worth further consideration, and inviting commentary. These functions lend vitality to users' social networks, which, we argue, is instrumental in maintaining Twitter's success and is instructive for those of us interested in supporting social search.

## RESEARCH OBJECTIVES

Our goal in conducting this research and in presenting the resultant findings is to address two overarching questions:

1. Do the questions that people pose to their social networks on Twitter adhere to an identifiable taxonomy? What are the *types* of questions that people ask on Twitter?

2. What rhetorical and social role does question asking play in people's interaction with Twitter?

With respect two our first research question, if questions on Twitter do fall into discrete categories, system designers

may capitalize on taxonomic information to improve information retrieval (IR) over microblog data. For instance, in the context of social search, perhaps some questions could find a wide audience while others should remain within a close social sphere.

Regarding question number two, ad hoc information retrieval has traditionally been predicated on the presence of a clearly articulated query. Recognizing this as a simplistic (albeit useful) model of interaction is not new. What is new, however, in the context of Twitter is the possibility of identifying information needs in a naturalistic, yet tractable way. The brevity of tweets makes them an appealing source of data about users' information needs.

However, it is not the case that tweets and queries are interchangeable. To capitalize on Twitter streams as a source of data for IR (e.g. for question answering or personalized IR), we must understand the way in which information needs find a voice on Twitter. Understanding the relationship between information needs and the rhetoric of tweets would admit a more nuanced strategy for IR systems where search is embedded in users' ongoing information interactions.

At a high level, our goals are two-fold. First, we aim to articulate findings that will improve the utility of information retrieval systems in the context of Twitter. Secondly, articulating a taxonomy of questions on twitter will enrich scholars' understanding of how this relatively new medium figures into people's experience with information.

## REVIEW OF THE LITERATURE

Although the literature on microblogs is young, a number of studies have informed the analysis presented here. In this section we situate our own work in the context of prior research on microblogging and several cognate areas.

### Overview of Microblogging

Like other social network services, microblogging platforms facilitate interaction among people. As boyd and Ellison (2007) note, these interactions often entail reinforcements and maintenance of social ties that were created in more traditional venues. Additionally, boyd and Ellison point out that social networking services are notable not because they allow people to meet strangers, but because they enable people to formally identify and characterize their social networks. This explicit identification of previously implicit relationships might have some impact on use of the service.

Social networking sites range from services like Facebook, (`Facebook.com`) which allows users to communicate with friends using embedded email services, photos, and an extensive commenting service; to Yelp (`yelp.com`), which provides timely, socially focused recommendations. Microblogging, another subset of social networking, is best exemplified by the popular website `twitter.com`,

which, as of April 2010, has over one hundred million users (Watters, 2010). Alternately characterized as a "status updating service," (Lenhart & Fox, 2009) or "social awareness streams" (Naaman et al., 2010); Twitter allows users to post brief textual status updates, which are limited to 140 characters per update. These status updates are often called *tweets* within the vernacular of Twitter users. When a user $U$ posts a tweet to his or her account, it is broadcast to $U$'s "followers," other users who have chosen to read his or her status updates. Likewise, when $U$ logs onto Twitter he or she is presented with the updates written by his or her "friends," people whom $U$ has chosen to follow. Tweets are typically displayed in reverse chronological order.

Because of Twitter's public API[1], there are numerous methods by which users can interact with the service. Twitter users can post messages via SMS (i.e., texting), by using web services such as the Twitter interface, or through a large selection of third-party applications. Variability of interaction allows users to update their streams in real time, independent of what technology limitations they might have. Users only need an Internet connection to post, do not need any specific application, nor do users need to learn any standardized interaction technique. This variability in interaction models allows users to personalize Twitter to suit their own needs, and results in a diverse user base using the service for heterogeneous ends.

The Twitter community has devised several strategies for enhancing the expressiveness of their tweets. For instance, if user U wants a follower F to pay special attention to a particular tweet, U might include the character string @F in that tweet. The @ sign signals that a tweet is directed "at" a particular user. Honeycutt and Herring (2009) found that although the @ symbol is not a native feature in Twitter it is used extensively. Another user-generated strategy is the use of the hash tag # to denote something like subject access (for example the hash tag #ASIST10 has been defined by the conference organizers to group together tweets that refer to the conference) (Efron, 2010). If a user is describing a specific incident, they will either make up their own keyword, or will follow naming conventions conceived by someone else, and spread through the Twitter community, usually by tweeting (Efron, 2010).

The brief, informal, and timely nature of tweeting allows users to keep up to date with people whom they don't encounter on a daily basis. Zhao and Rosson (2009) studied the dynamics of social interaction on Twitter, finding that in addition to helping friends and co-workers keep abreast of each other's lives, Twitter acts as a vehicle for "opportunistic conversations" where users are able to "converse" with people who may or may not be close ties, and who may or may not be in close geographical proximity. Such social functions are common on Twitter. According to Java et al. (2007), about 12% of tweets

---

[1] `http://api.twitter.com`

contain an @ symbol, and approximately 13% of tweets point readers to an external URL. This suggests that Twitter is widely used not only to "describe," but also to recommend and share information (Jansen et al., 2009); Wagner & Strohmaier, 2010).

## Asking Questions in Online Social Environments

Morris et al. (2010a) report a detailed survey of people with respect to question-asking and question-answering behavior on Twitter. Their analysis suggests that in many cases people turn to their Twitter networks to help them resolve information needs. In these situations users rely on Twitter as an informal social search service (Morris et al., 2010b).

The term social search refers to the process of finding information only with the assistance of social resources (cf. inter alia Cross et al. 2001; Borogatti and Cross, 2003; Cross and Sprouli, 2004). This typically includes enlisting the help of others, such as friends or other peers (as opposed to gatekeepers such as reference librarians) in the search process and is thus related, though not identical to collaborative search (Pickens et al. 2008). Evans and Chi (2008) emphasize that the social interactions in this economy may be implicit or explicit, and they may take place synchronously or asynchronously over varying geographical distributions. Although Twidale et al. (2007) describe the collaborative nature of traditional search, social search stands in contrast to traditional search in that in social search the choice of collaborators is more focused, and the nature of the questions are more contextualized.

Evans and Chi (2008) analyze social search interactions under the lens of Broder's (2002) taxonomy of search: *transactional, navigational,* and *informational.* They note that social interactions entail an especially promising tool for searchers with *informational* needs—i.e. people trying to gather information, as opposed to people trying to accomplish a particular task (transaction) or find (navigate to) a particular Web resource.

Of particular interest to the present study is the dissection of social search tactics outlined in (Evans et al., 2009). This paper makes a distinction between *targeted asking* and *public asking.* Targeted asking includes modalities such as email, where a searcher directs a question to a particular individual or delineated group. On the other hand, *public asking* involves broadcasting a question to a wide audience, either through posting a question to a public feed on Twitter, or by enlisting a search service such as Aardvark (vark.com). Targeted asking in Twitter can be accomplished by the use of the @ symbol (described above) or through the use of a direct message (DM), which creates a private conversation between the sender and receiver.

When people ask questions on Twitter they typically do so in a fashion that lies somewhere between targeted and public asking. Excluding direct messages, which are private, and will not be covered in this paper, questions on Twitter are posted to all of a user's followers, and therefore have a significant public component. On the other hand, questions are only available to a user's self-selected followers, thus limiting the scope of the question's audience. Directing a question to a particular follower via an @ mention signals the user's intention that his or her question has a narrow target, but its presence on the public feed (rather than a private direct message) means that the question is serving another purpose within the ongoing exchange between user and followers.

## INFORMATION NEEDS IN MICROBLOGS

A casual perusal of Twitter shows that people use the service for many reasons, including social search (cf. Sakaki et al., 2010). The following tweets exemplify this use of Twitter:

3. Does anyone know of a good tutorial on setting up a hadoop cluster?

4. I need to find a good restaurant in SOMA.

5. What is the maximum number of files in a directory?

These questions broadcast their authors' information need to members of their social network. In these cases, tweets act as queries in an informal social service, akin (though not identical) to platforms such as Aardvark (vark.com), a service that matches users' questions to people in their social network who are most likely able to answer them.

While the tweets in the preceding paragraph clearly play the role of queries in the information retrieval sense, other types of information need expression are common in Twitter data. Consider the following tweets:

1. Who needs a good webhosting and on top of that a stately income? http://bit.ly/dv9fP7

2. how's everyone doing today?

3. I'm also bothered by the lack of transparency of the release of detainees to Europe. Why won't the Administration say who they transferred?

4. @jpdimond please tweet what you hear for those of us who can't make it #chi2010

The first tweet actually points to an external resource whose topic treats the question posed in the tweet itself. Thus this tweet is in a sense suggesting a worthwhile question and offering fodder in its context. The second tweet solicits responses from the author's followers. The author does want information. But the desired information comes only from the author's social group. In the third tweet the author responds to an earlier message, using a question for elaboration, inviting the person to whom she has responded to continue their conversation. Finally, the fourth tweet demonstrates an author with an information need--to follow CHI 2010 conference events, posing this need as a request. Although the request is directed at a particular, the author also broadcasts this need to all of his followers, serving at least three purposes:

- To suggest that others might post updates about CHI

- To identify that although he is interested in CHI, he won't be there.

- To identify hash tag he will be using to identify tweets referencing the CHI 2010 conference.

Tweet number 4 also raises an important point: many statements of information need on Twitter are not phrased as questions. Phrases such as *I need to find, I'm looking for, I need, I'm trying* all indicate appeals for the input of others. Likewise, not all tweets containing questions express an information need, as the first tweet shows. In a similar vein, questions may be rhetorical, expressing an opinion without necessarily inviting a response.

In the remainder of this paper we analyze a sample of Twitter data to help readers come to terms with the different types of questions that authors post on microblogs.

## DATA
To inform our analysis we acquired two corpora of tweets. In this section we describe these data, as well as operational definitions that we brought to bear on them.

### Corpora
Microblog data is easy to come by. Using Twitter's streaming API we sampled data at random for a 24-hour period on April 19, 2010, generating a corpus with 2,022,544 tweets. We call this the *general corpus.*

However, for this study we were particularly interested in analyzing questions in a naturalistic setting and thus we built a more focused corpus, as well. Our *community corpus* consists of tweets written mainly by people who are interested in issues related to information retrieval. The corpus also contains followers and friends of these IR community members and is thus quite general in scope, with a kernel of IR interest. We identified the core set of authors in this corpus by tracking tweets that were written in a two-week period; one week before and after the date of the announcement of paper acceptances and rejections for the 2010 ACM SIGIR conference (March 24, 2010). Our hope was to use this event as an opportunity to identify people in the IR community who use Twitter. To accomplish this, we used the Twitter streaming API, tracking all tweets using any of the following words:

| sigir | sigir2010 | geneva | workshop |
|---|---|---|---|
| retrieval | iiix | cfp | hcir |

**Table 1. Words used to track tweets using Twitter's streaming API during the initial data collection phase.**

The words *geneva, workshop,* and *retrieval* intentionally admitted tweets by authors not involved in information retrieval research. The last three search terms (*iiix, cfp,* and *hcir*) were included because a good deal of discussion about SIGIR treated related conferences and calls for papers.

After tracking this activity we had collected tweets by 49 individual users. To augment this set, we downloaded each of the 49 users' friends and followers, for a total of 8,736 authors in our "community."

For each author we retrieved his or her 50 (or fewer if they had not tweeted 50 times) most recent tweets. However, due to a desire to acquire additional information, the 50 tweets per user were collected after the initial author-gathering period: tweets were downloaded on May 10, 2010. Taking at most 50 tweets per author, our corpus contained 375,509 tweets.

### Operationalizing and Sampling Questions
As noted above, not all "questions" on Twitter end with a question mark. Indeed the linguistic literature on the semantics of questions is large. Here we enlist a portion of that literature to help us operationalize the idea of a question in order to draw a meaningful sample of user questions from our corpus of Twitter data.

To guide our analysis, we refer to Karttunen's description of *question embedding verbs* in (Karttunen, 1977). Question embedding verbs are verb phrases that lend a declarative sentence interrogative semantics. The sentence, *I would like to know where you will be after the plenary* is, in Karttunen's analysis, the same as asking, *Where will you be after the plenary?*

To the best of our knowledge, no canonical list of question embedding verbs exists. Thus we combined an analysis of the verbs listed by Karttunen (p. 6) and our own reading of a large number of tweets to arrive at the following working definition of what constitutes a question in our analysis. A tweet contains a question if:

- It contains a question mark that is not part of a URL.

- It contains the phrase *I* [try*,like, need] to find.

- It contains the phrase *I* [try*, like, need] to know.

- It contains the phrase *I*m looking for* .

- It contains the phrase *I* wonder*.

In these cases the * sign is a wildcard, signaling 0 or more instances of any character. The list above is admittedly ad hoc, but in our initial analysis focusing on tweets that match these patterns yielded plausible samples.

In all sampling described below, we removed tweets written in languages other than English.

## CREATING A TAXONOMY OF QUESTIONS ON TWITTER
Our goal was to create a taxonomy that serves two purposes. First, we hoped to articulate generalities that pertain to the diversity of Twitter questions. Secondly, we aimed to build a taxonomy that would identify types of questions that would benefit from further analysis, such as information retrieval, visualization, or routing. Thus, we wanted our taxonomy to be both descriptive and actionable.

To create the taxonomy, five individuals analyzed a sample of 100 putatively question-bearing tweets. The tweets were taken uniformly from the community corpus. A sample size of 100 was chosen to balance the need for capturing the population's diversity with participants' ability to judge each question carefully.

The five judges consisted of this paper's authors and three information science graduate students. All participants reported familiarity and comfort with the Twitter service.

Each participant was given the sample of 100 tweets and instructions to classify these tweets with respect to their authors' *purpose* in writing them. We distinguished purpose-based classification from *topical* classification. A similar distinction (and a description of its motivation) is given in (Morris et al., 2010a). Aside from requesting a taxonomy based on *purpose,* the only strictures that guided the process were these: the taxonomy should only be one level deep (i.e. no categories will have sub-categories) and all categories must entail more than one tweet from the sample. Participants were also asked to provide a label for each category.

### Results of Initial Efforts

Our five readers developed taxonomies that shared many features, but that focused on distinct aspects of the samples. Summary statistics for these initial taxonomies appear in Table 2.

While Table 2 shows obvious differences in the granularity of the judges' taxonomies, we found many similarities as well. Space considerations prevent a thorough analysis of these, though we return to them below.

### The Final Taxonomy

After collecting all five initial taxonomies, we created a "final" taxonomy by analyzing the judges' initial work and by revisiting the community-based sample of 100 tweets in light of these efforts. As stated above, the goal in creating this taxonomy was to build a classification of Twitter questions that would have both descriptive power and utility with respect to designing systems that provide access to information in microblogs. Several mandates guided this process, including *accommodation* (do the classes provide suitable slots for the real range of Twitter questions?) and *expressiveness* (do the classes capture and manifest useful distinctions between types of Twitter questions?).

| Judge | mean | median | std. dev. |
|-------|------|--------|-----------|
| 1 | 12.5 | 11.5 | 7.21 |
| 2 | 12.5 | 7.5 | 11.02 |
| 3 | 9.9 | 7 | 8.43 |
| 4 | 12.38 | 7.5 | 11.41 |
| 5 | 7.08 | 5.5 | 3.23 |

**Table 2. Summary statistics of the number tweets per category created by each of five judges during the initial round of taxonomy creation.**

The list below describes the taxonomy that ultimately emerged from our analysis. In addition to the eight classes that comprise the final taxonomy, for the purposes of this paper, we defined a class called *not a question*, for tweets in the samples that met our rubric for inclusion but which were not actually questions.

1. **Suggest a question/answer pair – Promotional:** Tweets in this category encourage the reader to consider a question, and offer a way to answer or understand that question (typically by visiting an external website). Tweets in this category are probably not sent from one of the user's bona fide friends; i.e. it is likely to be 'spam' or some other solicitation.

2. **Suggest a question/answer pair – Social:** Tweets in this category encourage the reader to consider a question, and offer a way to answer or understand that question (typically by visiting an external website). Tweets in this category probably are sent from one of the user's bona fide friends; i.e. they are not spam.

3. **Request information: factual / clarification:** Tweets in this category seek information that is likely to have an answer, or at least likely to have tangible sources for addressing the question. These tweets may also seek information by way of asking for clarification of a previous tweet.

4. **Request information: opinion (individual or group):** Tweets in this category seek information that consists of opinions of people. Those opinions may be held by individuals, or the question may seek the consensus of a group with respect to a particular topic.

5. **Invite an action:** Tweets in this category encourage their audience to consider taking a particular action. These tweets may contain invitations, announcements of actionable events, etc.

6. **Express an opinion or current status:** rhetorical questions: Tweets in this status express a question, but they do so in a rhetorical manner in order to express the author's opinion or current status. Thus these tweets use a question to articulate a statement that is not directly seeking information (though it might incidentally cause a follower to respond to it).

**7. Coordinate action:** logistical planning / arrangement: Tweets in this category express questions that authors pose in order to coordinate interactions with (either face-to-face or virtual) with other individuals or groups of people.

**8. Not clear / miscellaneous:** If a tweet contains a question, but its intent is indiscernible or does not fit into any of the other categories, it belongs to this class.

**9. Not a question:** If a tweet has been included in the sample but does not express a question in any way it belongs in this category.

## ANALYZING THE TAXONOMY

The process of creating the taxonomy, while based on data, was admittedly heuristic. To assess the success of our design with respect to accommodation and expressiveness, we undertook a second round of classification. During this round, three judges (one author and two students[2]) classified question-bearing tweets according to the final taxonomy.

During this round, two samples were classified. The first, *sample 1,* was simply the original 100 tweets taken from the community corpus. *Sample 2* was 100 tweets sampled uniformly from the general corpus.

The goals during round 2 were twofold. First, we hoped to assess the ability of the taxonomy to accommodate a representative range of questions. Since the taxonomy was built using the community-specific corpus, we included tweets from the population of Twitter posts at large during this stage.

Additionally, we wanted to gauge inter-rater agreement with respect to our induced classes. If the classes are indeed based on visibly expressed qualities in tweets, different judges should, within unavoidable limits of subjectivity, agree on the best class for a given question.

To measure inter-rater agreement numerically we employed the Fleiss kappa statistic (Fleiss, 1970). Fleiss kappa is a generalization of the more well-known Cohen's Kappa; while Cohen's measure assesses only pair-wise agreement, Fleiss kappa measures agreement among $m$ raters:

$$\kappa = \frac{p_{obs} - p_e}{1 - p_e} \qquad \textbf{(1}$$

where $p_{obs}$ is the observed proportion of agreement over all $N$ judged cases across $m$ raters, and $p_e$ is the expected level of agreement. The two variables in Eq. 1 lie on [0, 1]; thus kappa can be less than 0, with larger values indicating greater agreement. For a full derivation of kappa, readers may refer to the work by Fleiss cited above.

[2] For logistical reasons, the second author and third student were unable to complete round 2.

*In our analysis kappa equaled 0.47, and 0.497 for samples 1 and 2, respectively.*

Interpreting kappa values is notoriously difficult. Among other problems, the magnitude of kappa depends on $k$, the number of classes. In our case $k=9$, a fairly high number, indicating that we may expect to see low levels of agreement.

Additionally, there is little agreement in the literature about what constitutes "significant" kappa scores. Following heuristic, and admittedly debated, guidelines, kappa scores in the high 40%'s are often referred to as evidence of "moderate" agreement, though others consider this bar too high (i.e. our results may be considered low). To contextualize this, chi-squared tests computed using the *irr* (inter-rater reliability) library of the R programming language[3] gave $p \ll 0.01$.

What we find promising about this result, however, is not the raw magnitude of the observed kappas, but the similarity of agreement between samples 1 and 2. Since the taxonomy was built only with reference to the community data set, we find the relatively high score on the less focused data to provide evidence of the taxonomy's generalizeability.

Figure 1 shows the category assignments as designated by each of our three raters during round two. As a point of reference, we base this presentation on a "gold standard" classification. The gold standard was obtained by taking a vote among the three judges and assigning the majority winner as the "correct" class (with ties when no judges agreed chosen at random).

Each row in Figure 1 refers to a sample: sample one, the community corpus, is the top row; and sample two, the general corpus, is on the bottom. There are four graphs in each row, with the leftmost showing the proportion of the sample's 100 tweets assigned to each of the nine classes by the gold standard. Each of the remaining graphs shows how the various judges allocated their classifications. The graphs are laid out in such a way that if there were perfect inter-rater consistency, all four graphs would look exactly the same.

All tweets in the graph are color-coded and arranged into columns that match the gold standard's organization. For example, the top row shows that Judge 2 labeled a large number of tweets as "expressions of opinion," and that all of these categorizations match the gold standard (in that there are no yellow bars anywhere else in his graph); however, he also categorized some tweets as "expression of opinion" which the gold standard categorized as "Suggest a question/answer pair – Social," (blue) "Request information: opinion" (red) and "Request information: factual / clarification" (green). This is in contrast to Judge 1

[3] http://rproject.org

who allocated the gold standard's "expression of opinion" tweets over several classes, or columns.

A glance at Figure 1 shows two main results. First, the large blocks of uninterrupted color in most panels reaffirm our quantitative analysis: judges often agreed on the appropriate class for a given tweet.

However, some classes see greater or lesser agreement than others. Promotional tweets (shown in the leftmost column in pink), for instance, see strong agreement. These tweets are often "spam," and are quite easy to recognize as such. On the other hand, posts sent from a friend, posing a question and pointing to a link are more nebulous. In Figure 1 these are shown in purple. We can see that Judges 2 and 3 distribute these purple tweets across many categories, suggesting either that the category is ill-posed or that it is difficult to identify this class of question.

A final point about Figure 1 relates to the overall distribution of tweets across classes. The infrequency (both in the gold standard and in each judge's assessment) of miscellaneous classifications suggests that our taxonomy is sufficiently accommodating. On the other hand, we can see that a very large number of the general sample tweets were finally labeled as either fact-based inquiries (green) or opinions (yellow). This is especially stark in the general corpus data. These results suggest that perhaps these classes could be refined further.

## ALTERNATIVE TAXONOMIC CONSIDERATIONS

The taxonomy presented in the previous section is accommodating and expressive. However, it is not the last word in how users conceive of question-asking on Twitter. In this section we consider other factors that differentiate these questions. Figure 2 suggests an alternate way of dividing the space of microblog questions. The first phase of taxonomy creation is where this dynamic was most strongly felt.

The vertical and horizontal dimensions of Figure 2 each correspond to a crucial aspect of every sampled Twitter question. The vertical dimension, Audience, refers to intended readers of questions. Broadly speaking, questions fell into two classes along this dimension. Either authors directed their questions to particular people (typically using the @ sign, though not always) or they posed questions to their readership at large (i.e. those people within their network).

Some questions were directed to several individuals. But for expository purposes, we count these as special cases of the "individual" mode. A second dimension that spans this information space is harder to articulate—we have labeled it *information need.* Some questions seek a tangible response. We have termed these *immediate information needs.* At the risk of oversimplification, authors write questions in this class to find out actionable information. In this sense, these questions are pragmatically similar to queries in an ad hoc information retrieval system.

However, a second type of information need is common in Twitter questions. We have called this mode of information need *persistent.* Tweets of this type may indeed seek a response but their authors' expectations with respect to replies is not as intense as it is in the *immediate* class. Questions of this type might be strictly rhetorical, encouraging a reader to consider a question leisurely. Or the question may pose a problem that is important, but which the author does not anticipate will generate answers in a narrow window of time or discourse.
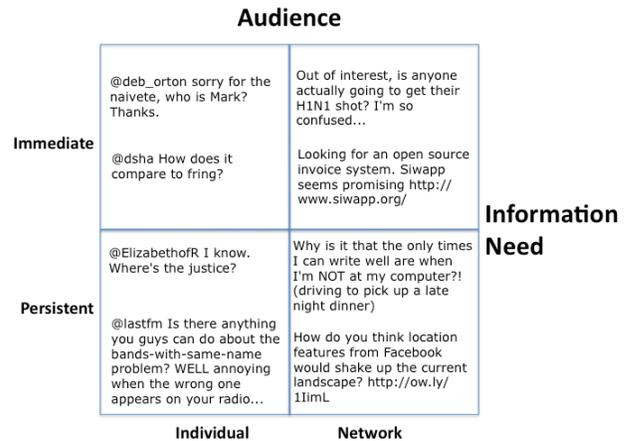


**Figure 2. A Two-dimensional schematization of questions used by users of Twitter.**

Each of the four quadrants in Figure 2 contains two tweets that exemplify the respective classes of Twitter questions. While some of these classifications are self-explanatory, a few are murkier and of special interest. Consider the bottom, left-most tweet (directed to `last.fm`, the Web-based music service). On first blush this tweet seems more *immediate* than *persistent* in its information need. However, we have classified it as *persistent* because of an (admittedly interpretive) aspect of the tweet. On its surface, the tweet requests information (*why…*) but in fact, insofar as it is addressed to a corporate Twitter account, we anticipate that its author is making a suggestion. He or she is pointing out a question that possibly many users have, with the hopes that `last.fm` will address it. Thus the tweet is pointing out (for a particular individual account) a persistent question.

### The Social Nature of Questions

In the initial round of classification, all five judges included a class such as *conversational,* wherein the tweet author expressed a question in efforts to engage another user in conversation, as opposed to asking a question to solve a keen and immediate information need.
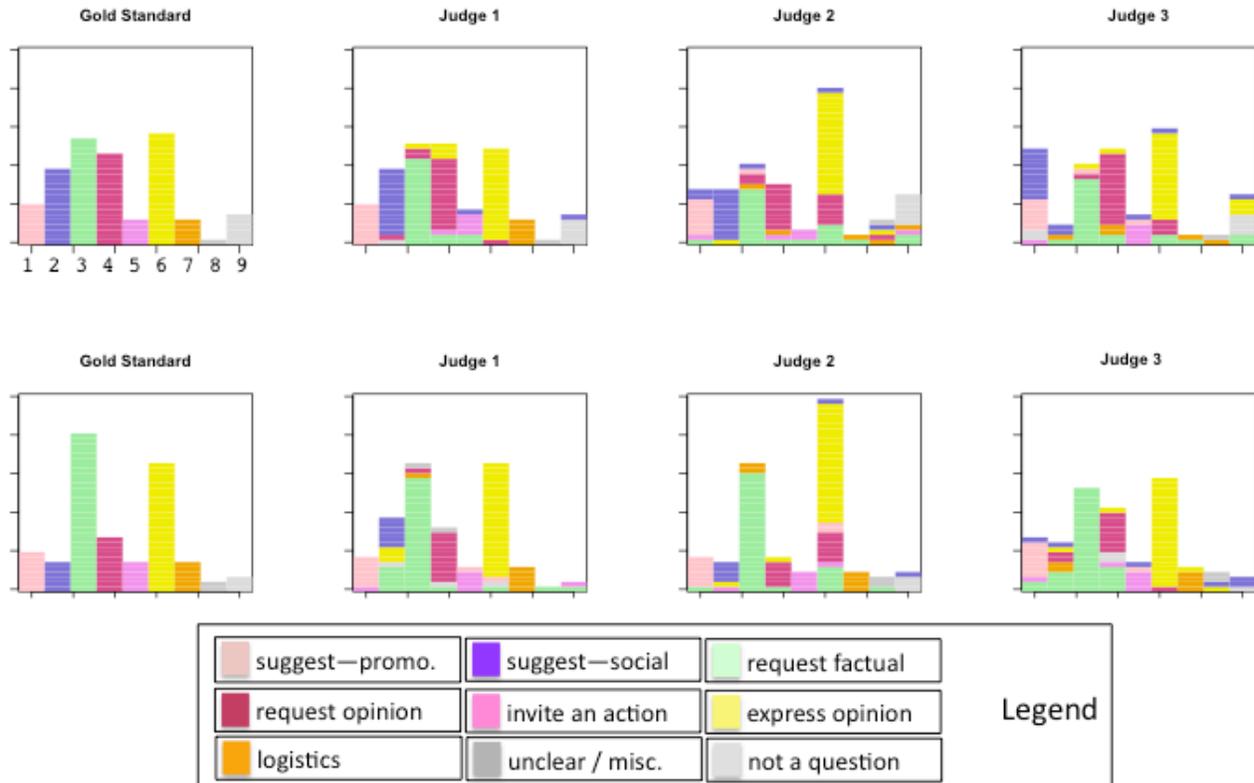
**Figure 1. Taxonomic assignments by each judge. Each column of each panel corresponds to a class in the taxonomy. Class numbers from Table 3 appear under top-right panel. These are color-coded: one color per class. Bars indicate which tweets each author assigned to each class. Top row corresponds to the community corpus; the lower row shows data for the general corpus. The leftmost panels (labeled Gold Standard) show the majority vote (with random tie breaks) for each tweet.**

Readers will note that such as class is absent from our final taxonomy. The reason for this omission is that on further analysis of our data it became clear that almost all questions in our sample had conversational elements as evidenced by the presence of @ and # symbols, and the presence of replies to the initial tweets. The nature of the microblog medium, we decided, makes a truly non-conversational question exceedingly rare.

One fact that emerged from the initial round of taxonomy creation was that interaction between authors is a key factor in the creation of questions on Twitter. With this in mind we scrutinized a large number of tweets from both corpora and arrived at the following hypothesis:

> People are more likely to ask a question in a reply to a previous tweet than they are likely to ask a question outside the purview of a previous tweet.

That is, we hypothesized that it is more likely for someone to ask a question during a conversation with another person than it is for a person to ask a question "point blank." Table 2 shows the data from our corpora that bear on testing this hypothesis.

| General Corpus | | | |
|---|---|---|---|
| | Question | ¬Question | |
| Reply | 87,701 | 446,509 | 534,210 |
| ¬Reply | 178,874 | 1,309,460 | 1,488,334 |
| | 266,575 | 1,755,969 | 2,022,544 |
| Community Corpus | | | |
| | Question | ¬Question | |
| Reply | 12,806 | 47,903 | 60,709 |
| ¬Reply | 50,322 | 264,478 | 314,800 |
| | 63,128 | 312,381 | 375,509 |

**Table 2. Question and reply statistics for two Twitter corpora.**

Table 2 bears out our hypothesis. The proportion of replies that are also questions is 87,701/534,210 and 12,806/60,709 for the general and community corpora respectively. Numerically these fractions are 0.164 and 0.211, respectively. The proportions of non-replies that are also questions are 178,874/1,488,334 and 50,322/314,800 for

each corpus—or, 0.120 and 0.160. With these results in mind we tested two hypotheses:

$$H_0^a : \frac{87701}{534210} \approx \frac{178874}{1488334}$$

$$H_0^b : \frac{12806}{60709} \approx \frac{50322}{314800}$$

where we may read $\approx$ as "is generated by the same distribution as."

Due to the large sample size it is not surprising that a two-tailed test of proportional equality yielded *p<<0.001* for both hypotheses. The interesting point is that in both corpora, there is strong evidence that people ask questions when they are interacting with other people.

This result suggests that many questions on Twitter are not analogous to ad hoc information retrieval. That is, more often than not, questions on Twitter serve a social function, supplying momentum in conversations. This is not to minimize their informational role. Indeed their status as responses does not mean that reply questions don't entail information requests. Instead, our data suggests that people use Twitter to ask questions that they would like to have answered, but also questions that engage other people.

## DISCUSSION

Asking questions plays an important role in peoples' experience with microblog services. Using the rubric for identifying questions laid out in this paper, approximately 13% of our 2 million-tweet corpus were questions. As we have shown in this paper, the characteristics of these questions are complex.

The taxonomy we developed was both accommodating, in that the classes provide suitable slots for the real range of Twitter questions; and expressive, in that the classes capture and manifest useful distinctions between types of real tweets. Inter-rater consistency was relatively high and there was agreement between categorization within a general corpus of questions and a community-based corpus, which implies that the taxonomy is generalizeable as well. The categories that had the highest consistency levels were those that looked like spam. The more difficult categories to categorize were those that asked specifically for factual information or clarification; Social suggestions for question / answer pairs, and requests for opinions. Two categories: "request for factual / clarification," and "express an opinion" were very broad, and constituted the majority of the tweets, particularly in the general corpus. These categories might productively be sub-categorized.

It is not surprising that people express opinions on Twitter, nor is it surprising that people use Twitter as an informal social search service, posing questions in efforts to learn facts, opinions, and consensus.

But other aspects of questions on Twitter are less intuitive. We found that a large number of questions were in fact posed in service to directing readers to an external resource. For example, consider the tweet:

> How Much Does Apple Make From iPhone App Store? Bernstein Research Claims $440 Million (http://bit.ly/14cvMq)

This tweet was classified as "Suggest a question: social" (i.e. class 2) by all of our judges. As evidenced by Figure 1, a large number of tweets fell into this category. What is interesting about this class of tweets is that they do indeed pose a question but instead of marking an authors' information need, these tweets recommend a question that readers might consider. Additionally, they usually point to a document that treats the recommended question.

Tweets of this class speak to an important dynamic in the microblog experience. When people come to these sources, we find that they may reasonably expect to be presented with questions that they would otherwise not have considered. In this case the tweet, which might actually pose an immediate information need, also serves a persistent purpose, in that it gives users food for thought, and an opportunity to engage in discussion.

Also of interest is the fact that so many questions appear in replies to previous tweets. As Table 2 suggests, questions play a special role in microblog discourse, offering a way to maintain conversational momentum. We found that many of the questions are not analogous to traditional information retrieval models in that the questions are often performing two actions: they are questions that need answers, but they are also questions that engage others. Meanwhile, Twitter users are busy asking questions in response to tweets. If we hope to design effective tools for supporting information use in microblogs, keeping this dynamic—seeking while conversing—in mind will serve us well.

## CONCLUSION

It is common knowledge that people rely on social networks to address their information needs. In this paper, we have argued that this dynamic holds in the context of microblogs, but that social mediation of information needs also has qualities in microblogs that are novel.

Our data show that people use microblogs to find information by asking, much as they use search engines. But more commonly, questions in microblogs play a nuanced role in peoples' experience with information. People ask questions on Twitter as a rhetorical device, recommending to other readers points worth ruminating on and resources to guide this rumination. Likewise, people ask questions in the context of spontaneous interactions with other Twitter users. Just as they do in face-to-face conversations, questions arise during Twitter interactions. But unlike common interactions, these questions on Twitter are broadcasted to their authors' network of followers.

In future work we plan to pursue the implications of these findings with respect to system design. Creating systems

that provide meaningful ways of storing and retrieving information from microblogs will benefit from an understanding of the dynamics that people bring to bear when they ask questions via microblog services.

## REFERENCES

Borgatti, S. P., & Cross, R. (2003). A Relational View of Information Seeking and Learning in Social Networks. *Manage. Sci.*, *49*(4), 432-445.

boyd, d. m., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal Of Computer Mediated Communication-Electronic Edition*, *13*(1), 210.

Broder, A. (2002). A taxonomy of web search. *SIGIR Forum*, *36*(2), 3-10.

Cross, R., Parker, A., Prusak, L., & Borgatti, S. P. (2001). Knowing what we know: Supporting knowledge creation and sharing in social networks. *Organizational Dynamics*, *30*(2), 100–120.

Cross, R., Rice, R. E., & Parker, A. (2001). Information seeking in social context: Structural influences and receipt of information benefits. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, *31*(4), 438–448.

Cross, R., & Sproull, L. (2004). More than an answer: Information relationships for actionable knowledge. *Organization Science*, 446–462.

Efron, Miles (2010). Hashtag retrieval in a microblogging environment. *Proceedings of ACM SIGIR 2010*, Geneva, Switzerland, 787-788.

Evans, B. M., & Chi, E. H. (2008). Towards a model of understanding social search. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work* (pp. 485-494). San Diego, CA, USA: ACM.

Evans, B. M., Kairam, S., & Pirolli, P. (2009). Exploring the cognitive consequences of social search. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems* (pp. 3377-3382). Boston, MA, USA: ACM.

Fleiss, J.L. (1971). Measuring nominal scale agreement among many raters. Psychological Bulletin, 76, 378-382.

Hearst, M. A., Hurst, M., and Dumais, S. T. 2008. What should blog search look like?. In *Proceeding of the 2008 ACM Workshop on Search in Social Media* (Napa Valley, California, USA, October 30 - 30, 2008). SSM '08. ACM, New York, NY, 95-98.

Honeycutt, C., & Herring, S. C. (2009). Beyond Microblogging: Conversation and Collaboration via Twitter. In *Proceedings of the 42nd Hawaii International Conference on System Sciences* (pp. 1-10). IEEE Computer Society.

Jansen, B. J., Zhang, M., Sobel, K. and Chowdhury, A. Twitter power:Tweets as electronic word of mouth. *Journal of American Society for Information Science*, 60(11) (2009), 2169-2188.

Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why we Twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis* (pp. 56-65). San Jose, California: ACM.

Karttunen, L. (1977). Syntax and semantics of questions. *Linguistics and Philosophy*, *1*(1), 3–44.

Lenhart, A., & Fox, A. (2009). Twitter and status updating. *Pew Internet Project Data Memo.* http://www.pewinternet.org (last check 2010-05-25).

Morris, M. R., Teevan, J., & Panovich, K. (2010). A Comparison of Information Seeking Using Search Engines and Social Networks. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*. Presented at the Fourth International AAAI Conference on Weblogs and Social Media, AAAI Press.

Morris, M. R., Teevan, J., & Panovich, K. (2010). What do people ask their social networks, and why?: a survey study of status message q&a behavior. In *Proceedings of the 28th international conference on Human factors in computing systems* (pp. 1739-1748). Atlanta, Georgia, USA: ACM.

Naaman, M., Boase, J., & Lai, C. (2010). Is it really about me?: message content in social awareness streams. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work* (pp. 189-192). Savannah, Georgia, USA: ACM.

Takeshi Sakaki, Makoto Okazaki, Yutaka Matsuo (2010) Earthquake shakes Twitter users: real-time event detection by social sensors. *International World Wide Web Conference.* Raleigh, North Carolina, USA 2010, 851-860

Twidale, M. B., Nichols, D. M., & Paice, C. D. (1997). Browsing is a collaborative process. *Information Processing & Management*, *33*(6), 761–783.

Wagner, C., Strohmaier, M. (2010). The Wisdom in Tweetonomies: Acquiring Latent Conceptual Structures from Social Awareness Streams, *Semantic Search 2010 Workshop (SemSearch2010),* in conjunction with *The 19th International World Wide Web Conference (WWW2010)*, Raleigh, NC, USA, April 26-30, ACM, 2010.

Watters, A. (2010 4). Just the Facts: Statistics from Twitter Chirp. *Read Write Web*. Retrieved May 20, 2010, from http://www.readwriteweb.com/archives/just_the_facts_st atistics_from_Twitter_chirp.php

Zhao, D., & Rosson, M. B. (2009). How and why people Twitter: the role that micro-blogging plays in informal communication at work. In *Proceedings of the ACM 2009 international conference on Supporting group work* (pp. 243-252). Sanibel Island, Florida, USA: ACM.