#TwitterSearch: A Comparison of Microblog Search and Web Search

Jaime Teevan
Microsoft Research
One Microsoft Way
Redmond, WA USA
teevan@microsoft.com

Daniel Ramage
Stanford University
353 Serra Mall
Stanford, CA USA
dramage@cs.stanford.edu

Meredith Ringel Morris
Microsoft Research
One Microsoft Way
Redmond, WA USA
merrie@microsoft.com

ABSTRACT
Social networking Web sites are not just places to maintain relationships; they can also be valuable information sources. However, little is known about how and why people search socially-generated content. In this paper we explore search behavior on the popular microblogging/social networking site Twitter. Using analysis of large-scale query logs and supplemental qualitative data, we observe that people search Twitter to find temporally relevant information (e.g., breaking news, real-time content, and popular trends) and information related to people (e.g., content directed at the searcher, information about people of interest, and general sentiment and opinion). Twitter queries are shorter, more popular, and less likely to evolve as part of a session than Web queries. It appears people repeat Twitter queries to monitor the associated search results, while changing and developing Web queries to learn about a topic. The results returned from the different corpora support these different uses, with Twitter results including more social chatter and social events, and Web results containing more basic facts and navigational content. We discuss the implications of these findings for the design of next-generation Web search tools that incorporate social media.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *search process.*

General Terms
Measurement, Design, Human Factors.

Keywords
Social media, Web search, microblogging, social search, Q&A.

1. INTRODUCTION
Many popular social networking services enable users to write brief status messages that they can share with their network of friends and, often, with the general public. Among these services, one of the most popular is Twitter; in 2010, over 15% of U.S. adult Web users are expected to use it [7]. Status updates on Twitter (also called *tweets*) are short snippets of text that provide news about the person posting, commentary on links, directed discussion, location information, the poster's current mood, or any other content that can fit into 140 characters.

WSDM'11, February 9–12, 2011, Hong Kong, China.

In addition to using microblogging services like Twitter to share information, there is evidence that people use them to find information. For example, people sometimes post status updates that are questions directed to their social connections ([10], [19], [20]). Because many status updates are public, people also gather information by searching collections of status updates to find recent posts on a particular topic. For example, Twitter provides a search interface to access public tweets, and Bing and Google have both recently begun to provide online search of Twitter posts. However, very little is understood about what motivates people to search a corpus of status updates, and about how such search behavior differs from that observed on traditional Web search engines. Microblogging content has very different properties than content on the Web; tweets are short, frequent, and do not change after being posted, while Web pages are rich, generated more slowly, and evolve after creation. We expect these differences to affect why people search, the types of content they search for, and how they go about finding it.

This paper presents the first systematic overview of search behavior on Twitter and what differentiates it from Web search. To better understand what motivates users to query Twitter, we begin by looking at questionnaire data. These qualitative responses help frame the subsequent analysis of large-scale Twitter and Web query logs to understand how the observed motivations translate into practice. We compare aspects of the queries issued to Twitter with those issued to traditional Web search engines, and study how the same users search both mediums for the same content. Our findings reveal that:

- People search Twitter to find temporally relevant information and information related to people. Memes, Twitter user names, and celebrity names are all popular Twitter queries.
- Twitter search is used to monitor content, while Web search is used to develop and learn about a topic. Twitter queries are more common, repeated more, and change less than Web queries. Some individuals issue the same query to both Web and Twitter search engines to capture these different uses.
- Twitter search results include more social content and events information, while Web results contain more basic facts and navigational content. The language used by Twitter results and Web result snippets is very different.

These findings suggest rich ways search engines can use social information finding behavior to improve the search experience. We conclude with a discussion of how this might be done.

2. RELATED WORK

Researchers have sought to understand how and why people search the Web using a variety of techniques. Query log analysis, the technique we use in this paper, is a valuable approach because it gives insight into people's self-motivated search engine use at a large-scale. Researchers have studied query logs to understand what people search for, including the types [5] and topics [26] of queries issued. Researchers have also used query logs to explore how search engines are used, revealing typical session behavior [15] and the prevalence of re-finding [28]. Others have used log analysis to understand how Web search behavior differs from behavior observed with specific corpora (e.g., news or images [2]) or from different entry points (e.g., mobile devices [16], [1]).

In this paper, we use query log analysis to contrast people's Web search behavior with their *social search* behavior on microblogs. Social search refers broadly to the process of finding information online with the assistance of social resources. Although social search can include behaviors like asking others online for assistance (e.g., [1]), when the term is used in the context of a search engine it refers to searches conducted over existing databases of socially-generated content such as blogs [22], tagged URLs [31], or archives of questions and answers [1].

To compare social and Web search, Evans et al. [8] conducted a between-subjects study where eight participants completed two search tasks. For one task, participants used non-social, online resources (e.g., search engines) while for the other condition they used social resources (e.g., emails to friends, or searches over Q&A sites). They found searching over social databases rarely produced task-relevant results and was less likely to prompt deep thinking than Web search. Morris et al. [20] conducted a within-subjects study where participants posted a question to their social network as their Facebook status message, and simultaneously searched the Web. They found that over half of their participants received responses from their social network before they had completed their Web search. Participants viewed Web responses as more authoritative and objective, but appreciated the personalized and trustworthy nature of answers received from their networks.

Some researchers have used query log analysis to gain insight into social search. Mishne and de Rijke [18] studied the queries issued to a blog search engine, and found that people were particularly likely to search for named entities (e.g., people, or products) and blogs on a topic of interest. People's overall search patterns, however, remained similar to Web search patterns. Sun et al. [27] compared blog queries with news queries, observing that queries often refer to people and temporally relevant content. Social tags have also been studied as proxies for queries in the retrieval of Web results [31]. We focus here on understanding the queries people issue when searching over a corpus of short status updates made to a social networking site.

Researchers have shown that finding information is an important use for status updates. For example, Lampe et al. [17] found university students used Facebook "to get useful information." Java et al. [14] identified "information seekers" as a primary category of Twitter users. Zhao and Rosson 0 found people use Twitter for "gathering useful information for one's profession or other personal interests," and "seeking for help and opinions," and Naaman et al. [21] found that questions to followers made up about 5% of posts that they manually coded. Honeycutt and Herring [10] similarly found that tweets directed at specific Twitter users were sometimes meant to "solicit information." Morris et al. [19] explored the use of status messages to find information by asking questions. They identified several reasons why people asked questions instead of searching, including a trust in the responses provided by friends and a belief that friends could provide better answers to subjective questions.

Nonetheless, little is known about what information seeking behavior on Twitter (and, in particular, keyword search over Twitter status updates) actually looks like. This paper gives the first account that we are aware of into what search over microblogging content looks like, and what differentiates from traditional Web search. We begin by looking at the results of a small-scale questionnaire to build a picture of why people search Twitter. With these motivations in mind, we use large-scale query log analysis to build a richer understanding of how these motivations translate into practice. We compare the queries issued to Twitter and Web search engines, the behavior of individuals moving between the two corpora, and the results returned via searches over both corpora.

3. WHY PEOPLE SEARCH TWITTER

To get an initial picture of why people search Twitter content, we asked 54 Twitter users at Microsoft, "When you search Twitter, what kind of information are you looking for?" Respondents provided a freeform, typed answer to the question. Answers were coded using a grounded theory approach [9], with a two-phase process that involved a first pass through all of the responses to develop a coding scheme of answer types, followed by a second pass to label each response, possibly in multiple categories if multiple themes were mentioned. We further supplemented these questionnaire responses with structured interviews of four active Twitter users at Microsoft, each with several years on the site and hundreds to thousands of followers.

Our respondents' basic age demographics are in line with Twitter's core user base [6]. Thirty four respondents were male, and the median age range was 36-45 years old. Nonetheless, it is likely that this group is not representative of Twitter users in general, as all respondents were employees of the same company and most reported being very familiar with Twitter. Although their responses can provide insight into the reasons why some of Twitter's core users search Twitter content, in this paper they are primarily used to motivate the analysis of large-scale query logs collected from a more representative sample of users.

The median time respondents had spent using Twitter ranged from 1-2 years. Most respondents (83%) reported reading tweets one or more times per day, and over half (59%) reported writing tweets one or more times per day. The mean number of people followed was 370.4 (median 159.5), which is more than typical [11]. The most popular applications used to access Twitter were TweetDeck (tweetdeck.com), Twitter (twitter.com), and Seesmic (seesmic.com).

Forty seven (87%) of the respondents reported having searched a corpus of Twitter posts and provided reasons why. While some of these motivations matched the common motivations behind Web search, others, including a desire for timely or social information, are relatively unique. We briefly examine the types of motivation in the sections below.

3.1 Timely Information

Many participants reported an interest in searching Twitter to find timely information. Current events were of particular interest to our respondents. Twenty three (49%) of those who had searched Twitter described having looked for information related to news (e.g., "technology news, trends"), topics gaining popularity (e.g., "information on a currently trending topic"), or summaries of events colleagues were attending (e.g., "event hashtags"). Participants' motivations for performing current event searches included keeping up with what was happening (e.g., "to keep up with events") and understanding trends (e.g., "basically a 'why is this trending?' kind of thing").

Another type of timely information respondents sought, mentioned by four participants, was real-time information. For example, one participant reported looking for, "Regional/local information (police incident, weather, etc)." Others wanted reports of traffic (e.g., "traffic jam") or the status of online services (e.g., "down services").

3.2 Social Information

Participants also reported using Twitter to find social information. Twelve participants (26%) described searching for information related to other Twitter users. Motivations for finding other users varied. Sometimes the intent was to find individuals with specific interests (e.g., "locate people with similar interests"). Other times it was to discover what particular individuals were saying (e.g., "accounts by certain people"). Some participants were interested in placing the tweets they observed from an individual in context. For example, one mentioned, "Occasionally I want to know how others responded to someone's tweet so I'll search for @replies," referring to the Twitter convention of preceding a user name with the '@' symbol. Another respondent used Twitter search to "figure out a cryptic comment from a user based on past" tweets. Several searched specifically for information about themselves, often looking for replies.

In addition to looking for information about people, another type of social information respondents sought was a general picture of people's overall opinions on particular topics. Four participants described searching to learn the community buzz. For example, one reported wanting to find, "what people are talking about with regards to an upcoming Microsoft event or product." Others looked for "movie reviews" or "marketing campaigns."

3.3 Topical Information

We also observed motivations for searching Twitter that more closely matched traditional Web search motivations. In particular, seventeen participants (36%) described searching for specific topics. For example, one mentioned, "astronomy or science stuff." Another, "Topic of interest (example: digital forensics)."

However, even the topically motivated searches on Twitter appeared to contain themes related to timely and social information. During the structured interviews, several of those interviewed described using Twitter searches to find public sentiment about topics of interest, regardless of whether those topics were originally discovered via feeds from followed users or through non-Twitter channels. One participant found the results of these supplemental searches to be more compelling than the information provided by the users he followed. This user considered many posts from his followed network to be outside of his topical interests, and thus read Twitter content by issuing six standing search queries. As search became his primary source of Twitter content, he began to use following merely as a way to "bookmark" people. While unusual, this user's experience illustrates the importance of search on Twitter: as users' networks grow and as interfaces for finding and filtering information on Twitter improve, standing queries and search in general may become more important to consuming social media.

People also reported using Twitter to try to re-find previously encountered information. Two participants described trying to return to previously viewed tweets. For example, one reported, "What I don't find are old tweets to/from people about a certain thing. Say I know someone sent me a link a year ago that is now somehow relevant - I can't usually find those things." Re-finding using Web search engines is very prevalent [28].

4. HOW PEOPLE SEARCH TWITTER

Building on these qualitative insights into a subset of Twitter user's motivations for searching Twitter, we used log data from a much larger and more representative sample of Twitter users to develop a quantitative understanding of microblogging search behavior along temporal, social, and topical lines. In this section, after introducing the logs analyzed, we contrast the queries people used on Twitter and the Web, examine temporal querying behavior through multi-query sessions, and provide a focused analysis of the overlapping queries that people issued to both Twitter and the Web.

4.1 Collecting Twitter and Web Queries

Information about the queries people issued to the Twitter search engine and to Web search engines was sampled from the Web browser logs of opt-in users of the Bing Toolbar. The toolbar is a commercial Web browser plug-in that provides augmented search features and reports anonymous Web usage behavior to a central server. Our analysis makes use of data from millions of users, and includes hundreds of millions of page visits. The data were collected during the two week period of November 11 - 24, 2009.

In addition to containing other URLs, the browser logs contain query URLs associated with multiple search engines, including those for search queries issued to general purpose search engines like Bing, Google, and Yahoo, and to vertical search engines like Twitter. It is possible to extract the queries issued to each engine from the URLs, and associate the queries with user IDs and timestamps. After filtering the data for spam and robots, we sampled 126,316 queries issued to Twitter by a subset of 33,405 users from the United States. Although people can search Twitter content in other ways, including via several major search engines and Twitter application tools (e.g., TweetDeck), the search interfaces can vary greatly. We focus only on queries issued to the Twitter search engine for consistency. All users included in the study issued at least one query to Twitter during the study time period. We also extracted the 2.5 million queries the same subset of users issued to Bing, Google, and Yahoo during the study period. Queries containing non-ASCII characters were ignored.

Table 1. Characteristics of the text of Twitter queries as compared with Web queries and queries common across both corpora. All differences are significant ($p < .01$).

	Twitter	Web	Common
Query length (chars)	12.00	18.80	11.69
Query length (words)	1.64	3.08	1.93
Is a celebrity name	15.22%	3.11%	38.20%
Mentions a celebrity	6.51%	14.86%	7.75%
Contains @	3.40%	0.14%	0.60%
Is username w/out @	2.37%	0.01%	3.25%
Contains #	21.28%	0.08%	0.2%
Is hashtag w/out #	4.35%	2.99%	5.88%

Requests for additional results beyond the top 10 or 15 initially returned were treated as part of the same query instances. When a user visits a Web search result and then returns to the search result page, the toolbar logs the result page as a new URL visit, even though the results are shown as part of the same search activity. For this reason, we treated the appearance of all duplicate queries that occurred within a fifteen minute window to the same Web search engine with no other queries intervening as the same query instance. In contrast, because Twitter results are not hyperlinked, the query URL does not get revisited between result visits. The results displayed for a repeat visit to the same URL are often different, and these were treated as new queries.

4.2 Queries Issued

We begin our analysis of how Twitter and Web search differ by characterizing differences in the text of the queries our population issued to both search engines. When comparing queries, differences in case, stop words, white space, and punctuation were ignored, with the exception of '#' and '@', as these characters have special meaning when used at the beginning of a word in a tweet. Table 1 summarizes several key differences between Twitter and Web query strings. All differences are significant ($p < .01$ with a paired two-tailed t-test). We see, for example, that Twitter queries are significantly shorter than Web queries. Also shown are statistics for the queries that were most commonly issued to both Twitter and a Web search engine by the same user. These queries are discussed in greater detail in Section 4.4.

Table 2 shows the queries issued by the largest number of unique individuals when searching the Web and when searching Twitter posts (as well as issued to both by the same individual; again, these are discussed in greater detail later). The most popular Twitter queries appear to relate to the topics identified via our qualitative analysis. Some are clearly temporally based, and related to relevant holidays (e.g., *thanksgiving*), recently released movies (e.g., *new moon*), or popular Internet memes (e.g., *#youknowyouruglyif*). One would not expect to find queries related to individuals' social networks among the most commonly issued search terms, as these are highly idiosyncratic (e.g., the usernames of specific friends). However, many of the queries are related to people (e.g., *lady gaga*).

In contrast to the most popular Twitter queries, the most popular Web queries are navigational in nature (i.e., intended to get to a particular Web resource) [5]. Note that the popular Web queries shown in Table 2 are somewhat different than the top Web queries overall, because they are based on our subsample of the population that has also run Twitter search queries. For example, *twitter* is not actually the most popular Web query, although it is among our Twitter-using population. Nonetheless, the prevalence of navigational intent is consistent with previous research [28]. It is not until the twenty-fourth most popular Web query in our data set that a non-navigational query (*porn*) appears. Of the 100 most popular queries in our data set, only 24 were not navigational in nature. A sample of these popular non-navigational Web queries can be found in the Common query column of Table 2, as 10 of the 24 are represented there.

Table 2. Ten queries issued by the most unique users to search the Web and Twitter, as well as the queries most likely to be used to search both by the same individual.

Web	Twitter	Common
twitter	*new moon*	*new moon*
youtube	*#youknowyouruglyif*	*justin bieber*
facebook	*justin bieber*	*adam lambert*
google	*adam lambert*	*taylor swift*
myspace	*#theresway2many*	*miley cyrus*
youtube com	*taylor swift*	*taylor lautner*
yahoo	*lady gaga*	*lady gaga*
ebay	*modern warfare 2*	*robert pattinson*
craigslist	*thanksgiving*	*chris brown*
myspace com	*#wecoolandallbut*	*modern warfare 2*

The most popular queries shown in Table 2 illustrate how Web search and Twitter search are used in very different ways, with one presenting many navigational queries and the other presenting temporal Internet memes. Many topics of interest, however, are similar: non-navigational Web queries and non-meme Twitter queries are highly represented by the queries individuals use to search both corpora, listed in the third column of Table 2. But the absence of navigational queries and the presence of memes only begin to scratch the surface of what makes Twitter queries different. We now look more closely at the prevalence of celebrity queries on Twitter, the use of specialized Twitter search syntax, and differences in query popularity.

4.2.1 Celebrity Queries

Although celebrities were a popular topic among both Twitter and Web searches, celebrity names emerged as an overwhelmingly popular category of query issued to Twitter. To identify celebrity queries, we began with an existing list of 234,008 popular celebrity names, compiled by the Bing search engine. We then manually coded the top 100 queries issued by the most people on the Web, in Twitter, and in common as to whether they referred to a well-known person, and added those that did to the list. Twitter queries were significantly more likely to be a celebrity name; 15.22% of the Twitter queries were, compared with only 3.11% of the Web search queries. On the other hand, Web queries were much more likely to include a celebrity name and additional content (e.g., *lady gaga is a man*); 14.86% of Web queries did, compared with 6.51% for

Twitter. Twitter celebrity queries appear to be motivated by a desire for timely information (such as breaking news about a particular person), rather than by a desire learn more about a particular aspect of that person. The presence of celebrity user accounts on the Twitter service, Twitter's youthful demographics, and the real-time nature of gossip likely contribute to the overall higher prevalence of celebrity queries on Twitter than on Web search engines.

4.2.2 Specialized Syntax

In addition to celebrities, people clearly use Twitter to search for specific users. As a Twitter convention, the '@' symbol is used to refer to a user's alias (e.g., *@oprah* or *@perezhilton*). The terms in a tweet that are preceded by the '@' symbol are hyperlinked and point directly to the referenced user's profile.

Of the Twitter queries, 3.40% contained an '@' symbol, while only 0.14% of the Web queries did. Most Web queries (87.77%) that contained the '@' symbol had it in the middle of the term, usually as part of an email address, whereas most Twitter queries (87.16%) had it at the beginning. Some Twitter queries appeared to reference usernames without using the @username convention. We compiled a list of usernames by removing the '@' symbol from all one-word Twitter queries that started with the symbol, and found that usernames without the '@' symbol constituted 3.25% of Twitter queries and 0.01% of Web queries. Although searching for other user accounts was popular, the use of the '@' symbol appears much less common in Twitter query logs than it is in the body of the tweets; boyd et al. [4] report that 36% of posts mention another user using the @username convention.

The hash symbol ('#') in Twitter is generally used in *hashtags*, a convention adopted by Twitter users to self-tag posts. The terms in a tweet that are preceded by the '#' symbol are hyperlinked; clicking on the link issues a search for tweets containing the associated tag. Many hashtags are compound words, such as *#cheatingexcuses* or *#dontmeantobrag*. Although Twitter queries are shorter than Web queries, the words in Twitter queries are on average longer (7.31 characters) than Web query words (6.10 characters). This may reflect the popularity of the relatively long hashtags in Twitter queries, as hashtags average 13.88 characters long.

Many Twitter queries (21.28%) contained a hash symbol, compared with only 0.08% of the Web queries. Twitter queries with a hash symbol were much more likely (99.91%) to have it at the start of a word than Web queries were (42.65%). When used in Web queries, the hash symbol was used primarily to represent the term "number" (e.g., *#46 on la lakers*). Additionally, many Twitter queries appeared to reference hashtags without the preceding '#'. We compiled a list of hash-less hashtags by collecting all one-word Twitter queries preceded by a hash symbol and removing the '#'. These comprised 4.35% of Twitter queries, and only 2.99% of Web queries.

The use of operators like '@' and '#" in Twitter queries is in some ways analogous to the use of other advanced query operators such as '+', '-', or quotations. Like these other operators, the presence of '@' and '#' can improve search success by reducing ambiguity. For example, if a search for *oprah*, began with the '@' symbol, it would clearly indicate that content from or directed to Oprah's official Twitter stream was being sought, whereas if the '#' symbol were used it would indicate an interest in news about Oprah. However, unlike Web search operators, '@' and '#' are explicitly part of Twitter's user-generated content and are regularly employed by Twitter content creators and consumers. We observe that they are also commonly used in the context of search. Prior large-scale log analysis of Web search [30] found that only 1.12% of Web queries contained advanced operators ('+', '-', quotations, or 'site:'), whereas in our sample 24.23% of the Twitter queries studied contained either '@' or '#'. Even still, large numbers of Twitter username queries are not preceded by an '@' and tag-word queries are missing a '#', suggesting that users may sometimes fail to employ these advanced operators or choose not to.

4.2.3 Query Popularity

In general, people are more consistent in the queries they issue to Twitter than to Web search engines. Each Twitter query was issued, on average, 3.08 times, while Web search queries were issued 1.76 times; similarly, only 23.19% of Twitter queries were unique, while 49.73% of Web queries were.

The consistency in Twitter queries may arise in part because searchers often issue queries on Twitter by clicking rather than typing. Many queries are issued via a click on a *trending topic* (i.e., a hyperlinked popular term), listed by the Twitter search box. It is impossible to distinguish via the logs whether a Twitter query was issued by the user typing the query into the search box or by the user clicking on a trending topic. However, based on a weeklong daily sample of trending topics, it appears that 30% of the trending topics have a unique URL format that we can identify in our logs, with the topic expressed in two different ways, separated by the advanced operator "OR." As 4,041 of the queries in the logs conform to this format, it is reasonable to assume that approximately 10% of the observed Twitter queries come from a click on a trending topic.

The use of hashtags likely further encourages this convergence in query terminology, since users tend to converge on their use of tags when others' tags are visible to them [25]. Because hashtags are hyperlinked in the Twitter interface and issue a Twitter search for the hashtag when clicked, many queries in the logs probably come directly from clicks on hashtags. Popular Twitter queries are much more likely to contain a hashtag than unpopular queries. The 50 most popular Twitter queries (representing 21.19% percent of query volume) contain a hashtag 50.73% of the time. In contrast, queries that occur once (representing 22.57% percent of query volume) contain a hashtag only 7.06% of the time.

In addition to clicks on trending topics, the use of Twitter search to learn about timely topics such as current events or celebrity news likely accounts for some of the query homogeneity, as only a limited set of topics are current at any one time. Popular Twitter queries are more likely to contain a celebrity name than unpopular queries. The 50 most popular Twitter queries are a celebrity name 24.92% of the time. In contrast, queries that occur once are a celebrity name only 4.03% of the time.

Finally, the more narrow demographic of Twitter users as compared to general Web users likely results in the more limited search vocabulary appearing within Twitter queries.

4.3 Temporal Aspects of Search Behavior

We now look at temporal aspects of Twitter and Web search behavior, including differences in session behavior and in repeat

Table 3. Characteristics of Twitter sessions as compared with Web search sessions and sessions with queries common across both corpora. All differences are significant ($p < .01$).

	Twitter	Web	Common
Number of queries in session	2.20	2.88	6.13
Number of unique in session	1.52	2.67	4.88
Secs between queries in session	9.38	13.63	20.56
Percent of repeat queries	55.76%	34.71%	46.30%

queries. Our observations are summarized in Table 3. All differences are significant ($p < .01$ with a paired two-tailed t-test).

4.3.1 Search Sessions

We begin by looking at how Twitter and Web search sessions differ. A session is a series of queries issued by an individual in close succession, often (but not always) with all queries being related to the same topic. Using a common approach for identifying sessions [15], we treat queries that occur in a sequence without 15 minutes of inactivity to be part of the same session.

The Twitter search sessions studied were shorter than the Web search sessions, both in terms of the number of queries they contain and the amount of time they span. On average, Twitter search sessions involved only 2.20 queries, while Web search sessions involved 2.88 queries. When multiple queries were issued within the same session, Twitter users waited 9.38 seconds on average before issuing the next query, compared with 13.63 seconds on a Web search engine.

Twitter session behavior often appears to involve monitoring of tweets of a particular query, with people refreshing the results after a short interval to see what is new. In contrast, overlapping but non-duplicate queries being more common with Web search, with people were more likely to change and modify their query within a session. There were many fewer unique queries in the average Twitter search session (1.52) than there were in a Web search session (2.67). Queries in Twitter sessions were issued 1.45 times each, compared with 1.08 times each on the Web.

4.3.2 Re-Finding

We also observe many more repeat queries overall on Twitter than is typical for Web search. In our data 34.71% of the Web queries were issued by the same individual more than once; this is very similar to the repeat query rate (32.59%) observed by Teevan et al. [28]. In contrast, 55.76% of the Twitter queries were issued more than once. In Web search, repeat queries often lead to re-finding. For Twitter, where, as we saw in our qualitative data, re-finding is difficult, it appears people use repeat queries instead to monitor topics over time, both within and across sessions.

4.4 Common Cross-Corpus Queries

To better understand how people integrate Twitter and Web search, we further analyzed the search behavior of the 4,277 people who issued the same query to both a Web search engine and to Twitter. We call the 3,534 unique queries that were issued to both services by the same individual *common queries*. An example of common query behavior can be seen in Table 4. In this example, the person begins by searching for information about the movie *New Moon* on Twitter, using both the query term *new moon* and the associated hashtag *#newmoon*. In a subsequent session, the person searches for *new moon* on the Web, and in a final session, the person searches on Twitter and then conducts a Web search to try to find places online to view the movie.

Table 4. An example of the Web and Twitter search sessions surrounding the common query *new moon*.

Query	Time	Corpus
new moon	Nov 20 at 10:46PM	Twitter
#newmoon	Nov 20 at 10:46PM	Twitter
new moon	Nov 21 at 5:36PM	Web
new moon	Nov 24 at 1:33AM	Twitter
watch new moon full movie	Nov 24 at 1:40AM	Web
new moon whole movie online	Nov 24 at 1:59AM	Web
watch new moon full movie	Nov 24 at 2:05AM	Web

This example is representative of many of the things discussed in greater detail below. People appear most likely to carry informational needs (like learning about a new movie) across Twitter and the Web. The linking query (e.g., *new moon*) is usually short in overall length like Twitter queries and short in word length like Web queries. Analysis of the surrounding session context suggests people use Twitter to monitor the query topic and the Web to learn more about it. Although these two activities occurred at the same time in many instances (as in the third session), they were also often treated as separate activities (as in the first and second session).

4.4.1 Queries

In Section 4.2, we observed that many Web queries are navigational, and that many Twitter queries relate to memes or social interaction. In contrast, the common queries used to link searches across the two corpora appear to be informational. For example, there is a much higher density of informational celebrity queries; 45.95% of the common queries were a celebrity name or contained one, compared with 21.73% on Twitter and 17.97% on the Web.

The linking query was usually a succinct representation of the common need, short in overall length like Twitter queries and short in word length like Web queries. As can be seen in Table 1, the common query length is more similar in terms of characters and words to the Twitter queries than to the Web queries. However, the average common query word length of 6.07 characters is close to the average Web query word length of 6.10 characters, and much shorter than the Twitter query word length of 7.31 characters. This may be because Twitter words are long in part due to the presence of long hashtags. Common queries are only somewhat more likely to contain a hash symbol than Web queries (0.22% of common and 0.08% of Web queries do), and are much less likely to contain one than Twitter queries (where 21.28% do). However, many common queries (5.88%) contain hashtag text without the preceding '#', more than Twitter (4.35%) or Web (2.99%) queries.

4.4.2 Temporal Aspects

Sessions with common queries display temporal characteristics distinct from other Twitter or Web search sessions. In the sessions containing the common queries there were 46,307 queries (13,486 issued to Twitter and 32,821 to a Web search engine). As shown in Table 3, these common sessions contained more queries (6.13 per session) and people spent more time between queries in a session (20.56 seconds) than they did with either Twitter or Web search. The time between queries may be influenced in part by the fact that these sessions sometimes include movement between a Twitter and Web search engine. The number of times each unique query was issued in a session falls between that of Web and Twitter searches, as does the amount of re-finding.

People who used the same query to search both corpora were more likely to issue the common query on the Web first (61.92% of users searched the Web before moving to Twitter with the same query). Common queries were often issued to Twitter and to a Web search engine within the same session; 43.74% of the common queries were run over both corpora in at least one session. But for the other 56.26% of the common queries, the query was issued only to either a Web search engine or Twitter during each session. For 70.53% of the queries, the query was issued to just one corpus in at least one session.

Web search sessions were more likely than Twitter search sessions to include related queries on the same topic. Although people who issued common queries did so with roughly the same frequency on Twitter and the Web (searching for a common query 13,199 times on the Web and 12,070 times on Twitter), the common query was much more likely to also be part of another query (e.g., *watch new moon full movie*) on the Web than on Twitter. On Twitter, the common query appeared as part of another query by the same user 995 times, while on the Web it appeared 8,416 times, or 8.46 times more often.

5. WHAT PEOPLE FIND ON TWITTER

Another important aspect of the differences in searching on Twitter versus the Web can be understood by analyzing the text of the returned search results. In this section we discuss the data we collected about Twitter and Web search results, and present the language differences that emerged from these data.

5.1 Collecting Twitter and Web Results

To approximate the Twitter content returned for the queries in our sample at the time they were issued, we crawled the eight million posts provided by Twitter's *spritzer* stream for one week of the study period (November 17-24). The spritzer stream is a public stream containing messages sampled from all public Twitter posts. Its makeup is determined by Twitter. From this we sampled the tweets that contained the 50 most popular common queries for further analysis. The number of potential results per query ranged from several hundred to tens of thousands.

Twitter search results differ from Web search results in that the entire content of each result is presented to the user in the result list. In contrast, Web search results are typically presented as a list of hyperlinks, each with an algorithmically extracted snippet of text designed to help the searcher select which hyperlink to visit (although in some cases the snippet can fully satisfy the user's information need). To represent the Web search results, we extracted the title and summary text of all of the results presented by Bing from the search engine's query logs for the same queries from the same time period. While tweets are qualitatively distinct from Web snippets, both form the textual basis by which searchers are presented with results the search system deems relevant, and hence warrant comparison.

Very common and very rare terms were filtered from each query-specific result set, as is standard practice for the type of analysis we performed. Specifically, we removed the 20 most common terms and terms appearing in fewer than three results. After filtering, 42 of our initial 50 query result sets had at least 100 non-empty results from both Twitter and the Web; we explore the differences in Twitter and Web search results for these queries.

5.2 Language Differences in Results

The most immediate difference between the Twitter and Web result sets lies in the amount of information available following a query. The mean of the per-query average number of words in a Twitter result was 19.55, versus 33.95 for the Web snippets. The relatively short length of the tweets reflects Twitter posting behavior in the presence of the system's 140 character limit. In contrast, the relatively longer length of the Web snippets reflects the goals of the search engine in supporting its users' Web searching needs. Because Web snippets are associated with a Web page, more content can be found via link following. Twitter results, in contrast, provide the full text of the matching tweets and are usually read in entirety in the result list, although 34% of the Twitter results in our collection contain an external link.

Because the Web and Twitter result sets were collected for the same queries, we might expect that they would contain essentially similar content. And, indeed, many common terms are shared; for instance, both tweets and snippets for the singer *lady gaga* are likely to contain the term *music* (8% of tweets, 27% of Web snippets). But with a broader quantitative analysis, we can observe that Twitter's real-time and social dynamics do result in patterns of language quite distinct from those in the Web search snippets. In the remainder of this section, we characterize the difference in search results between Tweets and Web snippets.

Automatic analysis of Web search snippets and Twitter search result tweets is challenging. The language used in both datasets is generally unlike the text on which supervised models in natural language processing and machine learning are typically trained. Additionally, measures based on term overlap, such as *tf-idf*, tend to be noisy because of the results' short lengths. Therefore, methods that can readily adapt to the data at hand and operate in a lower dimensional space are particularly appropriate. We use Latent Dirichlet Allocation (LDA) [3], a popular unsupervised latent variable topic model from the machine learning community that has been applied to Web documents for information retrieval [29] as well as to posts from Twitter [24]. LDA assumes the existence of a small number of underlying topics, each represented as a multinomial distribution over words. The model assumes that each document (here, a Web snippet or tweet) is generated by first picking a document-specific distribution over topics, and then picking each word from some topic's word distribution in proportion both to how much the document uses the topic and how much the topic uses the word. We use LDA's per-document topic distributions as robust feature vectors for computing similarity between Web snippet results and tweets.

We trained LDA models on the 42 common queries issued to both Twitter and the Web for which we had sufficient data. For each common query, we created a balanced bootstrap dataset of *2X* documents by sampling with replacement *X* of the associated

Table 5. Six topics learned from Twitter and Web results for the query *lady gaga*. Two each are disproportionately used in *Web* search results, *Twitter* results, or *Common* to both.

	Description	Top words
Common	General music	album new songs releases *url* best list fame artist review won track release artistdirect nominated #musicmonday rating another grammy available
Common	Particular concert in New York	why more this tweets york see from new in view week yonkers usa contact als down date 1986 bekend open
Twitter	Social chatter about Lady Gaga	what you *url* but looks about rt weird do now she's will man omg wearing say listening hell bitch lmao
Twitter	2009 American Music Awards performance	*url* adam ama 2009 performance want lol so lambert amas awards rihanna american watching tonight ama's im happy ladygaga award
Web	Biographical info about Stefani Joanne Angelina Germanotta	an her wikipedia stage germanotta after better Stefani name by joanne interscope American encyclopedia artist performing angelina records free known
Web	Music-related multimedia content	listen mp3 free videos gaga's mp3s pop downloads watch myspace download streaming yahoo singles read profile pictures click per every

Twitter results (out of N total) and X Web search result snippets (out of M). To ensure neither Twitter nor Web results were over-represented, X was taken as the minimum of M and N. Models were trained with 30 latent topics each and symmetric Dirichlet priors set to 0.01. Although these parameters can affect the learned models, we found the qualitative results below to be relatively stable across different parameter settings.

Table 5 shows some of the topics learned on the *lady gaga* dataset. The topics are divided into three categories, with the high probability words from an example of each topic type shown in the figure. The *Twitter* topics (representing 5 of the 30 latent topics) are those topics whose total probability on tweets is at least twice as much as its total probability on snippets. *Web* topics (representing 10 of the 30 latent topics) are those whose total probability on snippets is at least twice as much as its total probability on tweets. *Common* topics (15 of the 30) are neither Twitter topics nor Web topics; their total usage on tweets and on snippets was within a factor of two. In this example and others in our test set, we observe that the common topics tend to be about information semantically related to the query (e.g., music for *lady gaga*); the Twitter topics tend to include more social chatter and current events (e.g., Lady Gaga's performance at the American Music Awards); and the Web topic tend to contain more basic facts and navigational results (e.g., biographical information).

To quantify the difference in language used in Twitter and Web search results, we computed the per-query average cosine similarity of each Twitter result with the centroid of the other tweets and with the centroid of the Web snippets. Similarly, we computed the per-

Table 6. Average cosine similarity to the centroid of Twitter posts versus the centroid of Web search result snippets for tweets and snippets. All differences are significant ($p < .01$).

Cosine similarity	To Twitter centroid	To Web centroid
From Twitter	0.52	0.35
From Web	0.28	0.41

query average cosine similarity of each Web snippet with the centroid of the other Web snippets and with the centroid of the tweets. Cosine similarity was chosen because we have found that other distance functions (including information theoretic measures like KL-divergence) are inferior on per-document topic distributions. All averaging and comparisons are done in the reduced topic space. The cosine similarity, as averaged within each query and then across five bootstrapped samples, are shown in Table 6.

We find that, despite the inherent similarities of the datasets due to their query-based selection, the language of tweets is significantly different from that of the Web results ($p < .01$ with a paired two-tailed t-test). More interestingly, we also see that, despite the great variety of linguistic expression found among Twitter users, the Web results are actually more topically diverse than are tweets. The average similarity of Twitter posts to the Twitter centroid is higher than the Web results' similarity to the Web centroid ($p < .01$), reflecting the Web results' tendency to cover a broader range of coherent topics related to a particular query than the collective chatter of Twitter's many authors.

6. DESIGN IMPLICATIONS

Thus far we have looked at why people search Twitter, examined how Twitter search differs from, and relates to, Web search, and compared the results found via Twitter and Web search for the same queries. In this section, we discuss what these findings suggest for the design of next-generation search tools, including how we might enhance temporal queries, enrich people search, leverage hashtags, employ user history, and provide query disambiguation.

6.1 Enhancing Temporal Queries

Recently several major search engines have begun to incorporate Twitter results into their general Web search results. Given our analysis, we expect such integration to be particularly valuable for queries where freshness or buzz matters, such as for queries related to celebrities. Search engines could also use trending Twitter queries to discover additional queries that have strong temporal components, and use knowledge of these queries to integrate into the search result page not only tweets, but also other timely information like news.

6.2 Enriching People Search

For search tools that provide microblogging results, the popularity of people search on Twitter suggests strong people search support is important. Current Twitter searches for a username return the most recent tweets that have that username in them. But our questionnaire revealed other things people were looking for with such queries. One respondent wanted to find the "top links or top stories by a user." Another wanted to "find common followers between users." Microblogging search tools could include a link to the target's account profile, recent, popular, or re-tweeted posts by the target, communications between the target and the seeker, and recent or popular hashtags used by the target.

People-related queries, particularly celebrity-oriented ones, were frequently searched on both the Web and Twitter search engines. This suggests a particular opportunity for incorporating more information into either result page. For example, Twitter search results could incorporate Web query suggestions for queries related to people or celebrity, as a way to provide microblogging searchers with links to learn more.

6.3 Leveraging Hashtags

The popularity of hashtags as Twitter queries suggests new ways to exploit tags in Web search results. Others have examined the role tags from social bookmarking sites like del.icio.us might play in improving Web search result quality [12] and clustering [23], and the ability to identify high-quality, non-spam results to tag queries is an important area of research to pursue. For greater coverage, social tags manually added to sites like del.icio.us could be automatically supplemented with hashtags that co-occur with a URL in tweets.

Our findings indicate that Web search queries that appear to be hashtags but that do not begin with a '#' are common. Search engines could use approaches similar to ours to identify queries with a "hashtag intent" and federate them to Twitter.

At an interface level, Web result pages could expose tags like Twitter does: as clickable links that run new queries. Web browsers could similarly expose a hashtag convention whereby Web page creators could include tags that become automatically hyperlinked to a Web search.

6.4 Employing User History

We observed many more repeat searches on Twitter than on the Web. This suggests query history could be useful on Twitter. One questionnaire respondent asked for "built-in search I can store and refresh." The Twitter searches an individual issues most often could be identified automatically and presented back to the user, much in the same way as trending topics are. Because some users use trending topics to get an idea of what others are talking about, suggesting personalized trending topics related to an individual's past queries may be particularly useful.

6.5 Providing Query Disambiguation

Content analyses of the tweets that match a query (such as what was presented in Section 5.2 and Table 5) might provide signal to improve ranking of Web search results by helping to disambiguate the most common query intents. If a query-specific Twitter topic were popular (e.g., 2009 American Music Awards performance for Lady Gaga), pages matching that topic could be ranked higher.

Twitter could also be mined to discover information needs associated with a query by looking at the associated questions people ask. For any given Twitter query, many Twitter users post questions; on average 17% of the tweets in our Twitter result set contain a question mark. Because these questions are intended to be answered by people, they typically exhibit richer expressions of underlying information needs than queries do in general. The content of these tweets might give insight into the information needs driving query traffic and could be used for ranking, in conjunction with associated query refinement suggestions, or to motivate a Twitter search for additional information.

Table 7. Categories of questions expressed in Twitter posts matching the query *new moon*, and the percent of questions (out of 50) that fell in each category.

Category	Percent of Qs
Celebrity premiere event	52%
Foreign language	16%
Opening night	12%
Recommendation	8%
Movie trivia	2%
Definitional	2%
New Moon books	2%
Viewing history	2%
Astronomy	2%
Spam	2%

To explore the potential value of query-associated Twitter questions, we looked at the questions included in the Twitter result set of two queries randomly chosen from the top 50 common queries: *new moon* (a popular movie released during the time period) and *adam lambert* (a celebrity). For each query, we randomly sampled 50 tweets containing question marks and manually categorized these posts by information need.

The categories for the query *new moon* are summarized in Table 7. Most questions were intended to find breaking news about the celebrity premier event for the movie's opening (e.g., "hey #newmoonpremiere has taylor lautner been interviewed yet?"), to check whether friends were interested in going to see the movie (e.g., "@*username* so excited for new moon?"), or to query the network to see if the movie was worth seeing (e.g., "@*username* is the movie new moon as good as the movie twilight?"). Several other question types were expressed only once, reflecting a long tail of question types that could be asked about *new moon*. We found a similar distribution of information needs for the *adam lambert* query.

The broad coverage of question types that we observed suggests a promising route to improved search interfaces for popular queries: mining Twitter streams to discover common classes of questions for which specific answers or links can be presented. Ideally, this process could be automated or partially automated with clustering or classification.

7. CONCLUSION

In this paper, we have presented the first systematic investigation of how people search Twitter content, and how their usage of Twitter search differs from general Web search. Via large-scale analysis of query logs, we have discovered and quantified distinctions in the search behavior of users that issue queries to both Twitter and Web search engines. Some of these users' self-reported motivations for searching Twitter included an interest in timely information (e.g., related to news or events) and social information (e.g., related to other users or popular trends). By analyzing the queries themselves, we demonstrated differences in the types of queries issued to the Twitter search engine compared with a Web search engine. Twitter queries were shorter, but contained longer words, more specialized syntax, and more references to people. Large distinctions in query frequency were also apparent, with Twitter search often used to monitor for new content while Web search was used to develop and learn about a topic. Twitter queries were more common, repeated more, and changed less than Web queries. Twitter results included more social content and events, while Web results contained more facts and navigation.

It is our hope that these distinctions will provide those working to improve microblog search with a richer understanding of the information needs that lead people to search on Twitter, the Web and across both. Ultimately, we hope this understanding enables a new generation of search tools that merge the topical depth and breadth of Web search engines with the real time and highly social content offered by microblogging services.

8. REFERENCES

[1] Ackerman, M. and Malone, T. (1990). Answer Garden: A tool for growing organizational memory. *SIGOIS 1990*, 31-39.

[2] Arguello, J., Diaz, F., Callan, J. and Crespo, J.-F. (2009). Sources of evidence for vertical selection. *SIGIR'09*, 315-322.

[3] Blei, D., Ng, A. and Jordan, M. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3: 993-1022.

[4] boyd, d., Golder, S. and Lotan, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. *HICSS 2010*.

[5] Broder, A. (2002). A taxonomy of Web search. *SIGIR Forum*, 36(2): 3-10.

[6] ComScore (2009). Retrieved on July 2010. http://www.marketingcharts.com/interactive/younger-us-demos-12-less-likely-to-tweet-8679/comscore-twitter-age-distribution-users-april-2009png/

[7] eMarketer (2009). U.S. Twitter usage surpasses earlier estimates. Retrieved on July 2010. http://www.emarketer.com/Article.aspx?R=1007271

[8] Evans, B., Kairam, S. and Pirolli, P. (2010). Do your friends make you smarter? An analysis of social strategies in online information seeking. *IP&M*, 2010.

[9] Glaser, B.G. and Strauss, A.L. (1977). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine.

[10] Honeycutt, C. and Herring, S. (2009). Beyond microblogging: Conversation and collaboration via Twitter. *HICSS 2009*.

[11] Huberman, B., Romero, D.M. and Wu, F. (2009). Social networks that matter: Twitter under the microscope. *First Monday*, 14: 1-5.

[12] Heymann , P., Koutrika, G. and Garcia-Molina, H. (2008). Can social bookmarking improve Web search? *WSDM 2008*, 195-206.

[13] Horowitz, D. and Kamvar, S.D. (2010). The anatomy of a large-scale social search engine. *WWW 2010*, 431-440.

[14] Java, A., Song, X., Finin, T., and Tseng, B. (2007). Why we Twitter: Understanding microblogging usage and communities. *WebKDD/SNA-KDD 2007*, 56-65.

[15] Jones, R. and Klinkner, K. (2008). Beyond the session timeout: Automatic hierarchical segmentation of search topics in query logs. *CIKM 2008*, 699-708.

[16] Kamvar, M., Kellar, M, Patel, R, and Xu, Y. (2009). Computers and iPhones and mobile phones, oh my!: A log-based comparison of search users on different devices. *WWW 2009*, 801-810.

[17] Lampe, C., Ellison, N. and Steinfield, C. (2008). Changes in use and perception of Facebook. *CSCW 2008*, 721-730.

[18] Mishne, G. and de Rijke, M. (2006). A study of blog search. *ECIR 2006*.

[19] Morris. M.R., Teevan, J., and Panovich, K. (2010). What do people ask their social networks, and why? A survey study of status message Q&A behavior. *CHI 2010*, 1739-1748.

[20] Morris, M.R., Teevan, J., and Panovich, K. (2010). A comparison of information seeking using search engines and social networks. *ICWSM 2010*, 291-294.

[21] Naaman, M., Boase, J., & Lai, C. H. (2010). Is it really about me? Message content in social awareness streams. *CSCW 2010*, 189-192.

[22] Ounis, I., Macdonald, C., and Soboroff, I. (2008). On the TREC blog track. *AAAI 2008*.

[23] Ramage, D., Heymann, P., Manning, C. D., and Garcia-Molina, H. (2009). Clustering the tagged Web. *WSDM 2009*, 54-63.

[24] Ramage, D., Dumais, S., and Liebling, D. (2010). Characterizing microblogs with topic models. *ICWSM 2010,* 130-137.

[25] Sen, S., Lam, S., Rashid, A.M., Cosley, D., Frankowski, D., Osterhouse, J., Harper, F.M., and Riedl, J. (2006). Tagging, communities, vocabulary, evolution. *CSCW 2006*, 181-190.

[26] Spink, A., Wolfram, D., Jansen, B. J. and Saracevic, T. (2001). Searching the Web: The public and their queries. *JASIST*, 52(3): 226-234.

[27] Sun, A., Hu, M., and Lim, E-P. (2008). Searching blogs and news: A study on popular queries. *SIGIR 2008*, 729-730.

[28] Teevan, J., Adar, E., R. Jones, and M. A. S. Potts. (2007). Information re-retrieval: Repeat queries in Yahoo's logs. *SIGIR 2007*, 151-158.

[29] Wei, X. and Croft, B.W. (2006). LDA-based document models for ad-hoc retrieval. *SIGIR 2006*, 178-185.

[30] White, R. and Morris, D. (2007). Investigating the querying and browsing behavior of advanced search engine users. *SIGIR 2007,* 255-261.

[31] Xu, S., Bao, S, Fei, B., Zhong, S., and Yu, Y. (2008). Exploring folksonomy for personalized search. *SIGIR 2008*, 155-162.

[32] Yi, J., Maghoul, F., and Pedersen, J. (2008). Deciphering mobile search patterns: A study of Yahoo! mobile search queries. *WWW 2008*, 257-266.

[33] Zhao, D. and Rosson, M.B. (2009). How and why people Twitter: The role that microblogging plays in informal communication at work. *Group 2009*, 243-252.