

# Using Twitter to Assess Information Needs: Early Results

Max L. Wilson

Future Interaction Technologies Lab  
Swansea University, UK  
m.l.wilson@swansea.ac.uk

## ABSTRACT

Information needs tell us why search terms are used, helping to disambiguate, for example, what exactly people are looking for with queries such as ‘Orange’ or ‘Java’. It is hard to understand goals and motivations, however, from the keywords entered into search engines alone. This paper discusses the pilot analysis of 180,000 tweets, containing search-related terms, to try and understand how people describe their own needs and goals. The early analysis shows that some terms academically associated with searching behaviours were infrequently used by twitter users, and that the use of terminology varied depending on the subject of search. The results also show that specific topics of searching tasks can be identified directly within tweets. Future analysis of the still on-going 5-month study will constitute more formal text analytical methods and try to build a corpus of real search tasks.

## INTRODUCTION

Search is a very loaded term. We seek, search, look, find, and explore information. Traditionally information retrieval has focused on matching keywords to documents, which we now see in most web search engines. Information needs, however, tell us whether searchers have entered ‘orange’ in order to find information about citruses, colours, or corporations. Further, information needs are typically part of larger work tasks [2], where the goal of searching for ‘orange’ may be to write a report, plan a food shop, manage a diet, or buy a phone, etc. Understanding information needs and work tasks, therefore, tells us whether interfaces need to be supporting activities such as: exploration, synthesis of information, comparison, or evaluation [11]. Further, understanding information needs tell us how we should design interfaces that support effective human computer interaction during information retrieval.

In this paper, the early stages of an analysis into how people describe and converse about their own information needs are presented. After discussing related work on information needs and analysing twitter, the method and results of this pilot stage analysis are presented. The paper concludes with some potential findings, before discussing the future plans for the full analysis of a 5-month archive of tweets.

## RELATED WORK

### Information Needs

Gaining insight into real information needs is not trivial. Advances however, have been made by, for example,

studying search engine logs [4] and comparing keywords with relevance judgements [13]. Broder [1] noted that web searches typically fall under three categories of: transactional, navigational, and informational. Transactional queries are for web-based activities, such as buying, downloading, printing, etc. Navigational queries are simply to find a known website. Finally, Informational queries are those performed while trying to learn. Rose and Levinson [13] extended these into a hierarchy of goal types, such as types of learning, and types of transactions. Other research (e.g. [10]) has been trying to automatically infer goals based on click behaviour of a searcher over time.

The value of understanding information needs and goals is further emphasized by the inclusion of context when setting search related tasks in studies. TREC tasks [3], which are used to benchmark the performance of search systems, are created in association with *topics* so that it is clear what constitutes accurate results. Capra and Kules [9] further identified the types of contextual information that are important to provide to study participants when creating exploratory search tasks for user studies.

Jarvelin and Ingwersen, in discussing many aspects of information seeking, also noted that separate research areas have focused on both information needs and *perceived* information needs, where search is more closely related to how users currently understand their information needs [5]. Part of exploratory search and learning often involves first understanding a problem space, and then resolving it.



Figure 1: Tweets that included the exact text: 'searching the net...', shown in a word tree.

### Using Twitter as a Resource

Twitter is becoming a popular medium for communication, and recent work has begun analysing: networks, how

people communicate, and what they talk about [6]. Pear Analytics, for example, classified tweets as being either: News, Spam, Self-Promotion, Pointless Babble, Conversational, or Pass-Along. Their results showed that around 40% was babble, 37% was conversational, and, in third place, Pass-Along constituted 9% of the tweets [7]. Similarly, the Web Ecology Project released a sentiment analysis of tweets regarding Michael Jackson's death [8]. In comparison to a typical archive of tweets, the Michael Jackson archive included a significantly larger portion of negative tweets.

### GATHERING INFORMATION NEEDS FROM TWITTER

With the aim of better understanding real information needs, Twitter was analysed as a worldwide resource of people's public discussions, to find conversation about searching behaviour. Although Pear Analytics said that twitter is mostly used for babble and conversation, these are the elements of their taxonomy, as opposed to news, spam, and self-promotion, that will provide value for this study. Figure 1 shows a basic example, where people used the exact words: 'searching the net...'. The analysis described here is of the first 2 weeks of a larger 5 month investigation into the ways people describe their own searches on twitter.

### Method

To gather tweets that describe searching behaviour, a Twitter search was automatically queried every hour for the most recent 100 tweets for each of the 10 search related terms<sup>1</sup> listed in Table 1. The terms, mainly selected from academic publications from search communities, were also passed through a thesaurus to identify and consider additional English language terms. Alternatives, as in those not used, were checked with a single search on twitter to assess current frequency of use on twitter. The chosen terms were those above a significant drop-off point. This process was performed for two weeks during this pilot analysis. To catch as wide a net as possible, all tweets including these terms were archived without any analysis of whether they were describing searches. That is, although Figure 1 shows a basic example of where people explicitly talk about searching the '[inter]net', this research has aimed to discover real-world information needs and work tasks, which may involve search behaviour in real or physical environments, as in Figure 3. Further, each of these terms were queried in their past, present, and future variations, such as the query 'find OR finding OR found'.

To analyse the tweets, several methods are being considered. The initial analysis here is designed to be more qualitative to a) reveal early interesting qualitative insights, such as in Figure 2, and b) help inform the way that the final dataset should be more formally analysed. Initially, for visualization, tag clouds were considered, however these

revealed very little about what people searched for. After a more structured semantic analysis of tweets, however, tag clouds of identified *search topics* may provide interesting insights. At this stage, aside from some high-level statistics, Word Trees, using IBM's ManyEyes project [14], were used to manually and qualitatively explore the content.

searched twitter for keyword : poop <

Figure 2: This exact phrase appeared in 2 separate tweets.

### Results

In total, 189,452 unique tweets were captured from 163,564 authors. Additionally, 14,959 re-tweets were archived, where users echo the tweets of others to their own network.

Table 1: Showing a breakdown of the tweets collected during the first 2 week archiving process.

Term	Unique Tweets	ReTweets	Authors
Exploring	21,287	1,414	19,119
Finding	26,333	1,107	25,656
Foraging	910	1,627	790
Hunting	26,534	1,123	22,666
Investigating	19,255	2,016	14,488
Looking	22,783	1,267	21,142
Retrieving	3,506	1,500	3,269
Searching	25,493	1,788	20,095
Seeking	15,767	1,380	12,987
Studying	27,584	1,737	23,352
Totals	189,452	14,959	163,564

### Frequency of term use

One contribution of this analysis is to see the popularity of different terms as people describe their searching actions. 'Studying' was the most popular term used, but, despite being a popular metaphor for how people may search [12], the 'Foraging' term, and its temporal variations, were hardly used. Similarly, and perhaps surprisingly, the term 'Retrieving', and its variations, were used significantly fewer times than many of the other terms. The terms 'Searching', 'Hunting', and 'Finding' were also popular terms, but 'Hunting' in particular was often used in relation to sport, as discussed below. While 'Looking', as might be expected, was quite popular, two terms relating to exploration ('Exploring' and 'Investigating') were also quite popular. The term 'Seeking', while perhaps quite an academic term for search, was used almost half as often as most other terms, but significantly more than the term 'Retrieving'.

<sup>1</sup> Unfortunately, the term 'browse' failed during this pilot study, but has been fixed for the 5 month study.



**Figure 3: Tweets described searching behaviour in both physical and digital environments.**

*Language associated with terms*

Another contribution of performing this qualitative analysis, is in being able to see how different terms are used to describe different kinds of searching. Figure 4, for example, shows that the ‘Finding’ terms were often associated with finding an ideal or optimal results. When followed by the word ‘my’, however, the task was often re-finding, and usually for locating where technology was in the home. A third regular use of the ‘Finding’ terms was followed by the word ‘out’, which typically represented more exploratory tasks.

The variations of the term ‘Exploring’ were typically used in regards to new places, such as cities and neighbourhoods. Many people self-reported as exploring twitter for the first time. Exploration, however, was often associated with abstract objects, such as ideas, options, and possibilities, but also with genre’s of music and film.



**Figure 4: Use of the term 'Finding' when followed by 'the'. This combination was often associated with ideal individual results, the ‘right’, ‘perfect’, or ‘best’.**

Perhaps interestingly, the variations of ‘Foraging’ were nearly always used in conjunction with food terms. Although people rarely used the term, people self-reported as foraging in cupboards, fridges, kitchens, and freezers, with the aim of locating food at mealtimes. When not

associated with food, the term foraging described behaviour in outdoor areas, such as yards or woods, but also within documents.

‘Hunting’, when not being used to discuss sport, was for: new jobs, people (including witches), and technology. Like the term ‘finding’, ‘Hunting’ was often used in association with adjectives representing optimal results, such as a best, cheapest, or perfect.

The ‘Investigating’ terms were typically used in relation to crimes. When used, however, they were often investigating the informational boundaries around such events, as investigating: claims, correctness, cause, and circumstance.

Like ‘Hunting’ and ‘Finding’, the term ‘Looking’ terms were often related to people, jobs, and technology, and their optimal variations, including ‘best’, ‘right’, and ‘perfect’. The term was also used, however, in association with looking for a new place to live, excuses, and the original copies of objects. People also often described looking for entertainment items, such as music, books, and movies.

When used, ‘Retrieving’ terms were related to gathering lost or distant items, often one’s daughter. The majority of subjects in these tweets, however, were digital, such as retrieving lost or archived passwords, records, files, pictures, and tweets.

‘Searching’ terms were used for a large range of subjects. While sometimes used in relation to optimal (best, next, perfect) technologies, ‘Searching’ was also used for food, missing people, soul mates, truth, music, friends, and pictures. The ‘Searching’ terms, however, produced the highest number of exact quoted search terms, discussed below. The ‘Searching’ terms also returned the highest number of tweets that described venues for search, such as Google, Facebook, eBay, Twitter, etc.

When not used for adult advertisements, the term ‘Seeking’ was primarily used for finding people for jobs, or a place to stay. It was also heavily used with exploratory and abstract terms such as ‘the truth’ and ‘to be understood’. ‘Seeking’ terms were also used in breaching peoples borders, such as ‘new lands’ and ‘faces’.

Finally, the studying term was primarily used when discussing forth-coming exams. Sometimes, however, studying was associated with self-driven learning on topics such as the bible, psychology, and photography. Consequently, the ‘Studying’ terms provide some interesting topics for learning tasks in studies, including the history of tobacco and the effects of erosion.

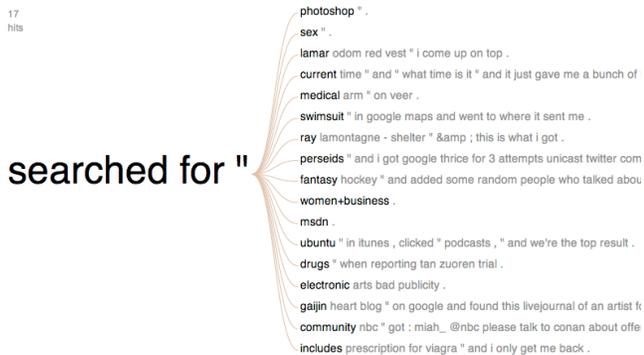
*Specific subjects of search*

Finally, a third contribution of the analysis is in identifying specific searching tasks. Figure 5, for example, shows three complicated self-reported information needs. The first represents a complex search need, where the user has two pieces of related information. The second and third represent more exploratory learning tasks. Figure 6,

however, shows that many twitter users directly provided search terms they had used, using speech marks. Figure 6 indicates those that explicitly used the past-tense variation of 'Searching' followed by the word 'for' and then speech marks.

searching the net about it , it was made by roland and was used by kanye west  
 searching the internet all day for a song i had heard .  
 searching the net re global food shortages .

**Figure 5: One complex search task and two exploratory tasks described by twitter users.**



**Figure 6: Twitters often labelled, using speech marks, exact specific terms they had queried different services for.**

## CONCLUSIONS

This work has reported the early pilot analysis of a work-in-progress investigation into tweets that included searching-related terminology, archived in the first 2 weeks of a larger 5 month study. The analysis revealed early insights into how often, and in regard to which forms of search, different search terms were used by twitter users when discussing their own searching behaviours. Where previous research has typically tried to deduce information needs from search engine logs, this research is trying to identify information needs from publically available conversations on the web.

In completion of the full 5 month long study, more formal text-analysis techniques will be applied, perhaps including a sentiment analysis [8], to find out if, for example, the search behaviours that people feel are worth tweeting mainly surround difficult or novel searches. Further, such an analysis may be able to identify the frequency, subject, and success of different types of searching goals [13]. Part of the aim, therefore, will be in building a resource of realistic search tasks for different types of searching contexts, which can be used in future user studies, and informed by people's own self-driven descriptions of searching behaviour. The research described here, however, provides early insights into how people describe and communicate their own searching activities to others.

Understanding how people perceive their searching activities and needs can help inform the design of interfaces for human computer interaction during information retrieval.

## REFERENCES

- [1] Broder, A. A taxonomy of web search. ACM SIGIR Forum, 36, 2 (2002). 3-10.
- [2] Byström, K. and Hansen, P. Work tasks as units for analysis in information seeking and retrieval studies. in Bruce, H., Fidel, R., Ingwersen, P. and Vakkari, P. eds. *Emerging Frameworks and Methods*, Libraries Unlimited, Greenwood Village, CO, 2002, 239-251.
- [3] Harman, D.K. The TREC conferences. (1997). 247-256.
- [4] Jansen, B.J. and Spink, A. How are we searching the World Wide Web?: A comparison of nine search engine transaction logs. 42, 1 (2006). 248-263.
- [5] Järvelin, K. and Ingwersen, P. Information seeking research needs extension towards tasks and technology. 10, 1 (2004). 10-11.
- [6] Java, A., Song, X., Finin, T. and Tseng, B., Why we twitter: understanding microblogging usage and communities. in KDD'07, (2007), ACM New York, NY, USA, 56-65.
- [7] Kelly, R. Twitter Study - August 2009, Pear Analytics, 2009, 1-13.
- [8] Kim, E., Gilbert, S., Edwards, M.J. and Graeff, E. Detecting Sadness in 140 Characters: Sentiment Analysis of Mourning Michael Jackson on Twitter, Web Ecology Project, Boston, MA, USA, 2009, 1-15.
- [9] Kules, B. and Capra, R., Creating exploratory tasks for a faceted search interface. in HCIR'08, (2008).
- [10] Lee, U., Liu, Z. and Cho, J., Automatic identification of user goals in web search. in WWW'05, (2005), ACM New York, NY, USA, 391-400.
- [11] Marchionini, G. Exploratory search: from finding to understanding. Commun. ACM, 49, 4 (2006). 41-46.
- [12] Pirolli, P. and Card, S., Information foraging in information access environments. in CHI'95, (1995), ACM Press, 51-58.
- [13] Rose, D. and Levinson, D., Understanding user goals in web search. in WWW'04, (2004), ACM New York, NY, USA, 13-19.
- [14] Viegas, F., et al. Manyeyes: a site for visualization at internet scale. IEEE Trans. Visualization and Computer Graphics, 13, 6 (2007). 1121-1128.