Search Across Social Media Sites

Information Structure & Retrieval Spring 2010
Project 2

Eva Wilhelm

1 Introduction

While Social Media is far from being established in mainstream communication, it has clearly developed from a fringe phenomenon into a legitimate communication means with a solid user base that regards it as an integral part of everyday personal and professional communication. The topic is widely discussed in both popular and academic publications, however, research is largely focused on social aspects. Search behavior, searchability and other information retrieval aspects are rarely examined. In this paper, I will address some of these points.

First, I will review current research on how users search on Social Media sites, what they search for and how they navigate between sites. Subsequently, I will synthesize a list of new needs arising from Social Media search that go beyond general web search, and present an overview of how global search engines like Google and local search engines of Social Media sites try to accommodate to these needs and how these modifications are received by users. Finally, I will identify the needs that haven't been addressed yet or at least not satisfactorily, especially regarding the parallel use of several sites by a single user, and propose possible starting points for solutions.

2 Defining Social Media

Before any discussion about Social Media search can begin, the term "Social Media" needs to be defined, since several varying definitions can be found throughout academic literature, usually depending on the domain an article or paper originates from. For the purpose of this paper, I will use the term in the broadest possible sense. According to Smith et al. [2008], "from a sociological perspective, social media can be described as 'collective goods produced through computer-mediated collective action'." The term "collective goods" can be used to describe a diverse array of objects or even concepts, ranging from collaboratively authored articles on Wikipedia[1], to the more abstract notion of Social Capital[2] in the case of Facebook[3] [ibid.].

The diversity of the applications labeled "Social Media" poses one of the main problems when trying to infer rules or develop models that are applicable to the entire body of sites. Since the boundaries of Social Media are somewhat diffuse, it is difficult to conduct research covering its entirety. However, there are a number of characteristics classifying an application or site as Social Media, one being the emphasis on social relationships and collaboration. Hong and Davison [2010] make a case for a unified search model for Social Media, arguing that such a model could be used as a basis for developing more specific models for certain types of Social Media or to improve the quality of search results from Social Media retrieved by search engines.

Another aspect Social Media sites share are the problems when it comes to searching user-generated content [Cho and Tomkins, 2007]. Collaboratively created content available on Social Media sites is usually not controlled very closely or in a timely manner, thus, spam will inevitably appear in search results. Due to Social Media's immediate nature, relevance of this content is often time-sensitive, so relevant content can become useless when found at a much later point in time. Moreover, a large portion of Social Media content tends to be very personal and might not be of interest for the majority of users, but will appear in their search results none the less. Finally, recently

[1] http://wikipedia.org

[2] Social Capital can be defined as any type of social interaction that generates productive benefits.

[3] http://www.facebook.com

available options to assign each content object a specific level of privacy adds another new challenge for for search engines.

In general, only very little research on search behavior across Social Media types and sites exists. Current research usually covers one particular site or type of Social Media. Therefor, this paper synthesizes these findings and highlights results concerning web search or individual Social Media sites that might be applicable to other Social Media types.

3 How do users search?

When analyzing the way users search in Social Media, temporal, social and sequential aspects can be taken into account.

3.1 Browsing

Although Social Media sites cater to different search goals and facilitate many search strategies, one thing they have in common is that they facilitate browsing in some way. Whether you look at tags in delicious, hashtags on Twitter[4], links to friends' profile on Facebook or to related article on a blog post, browsing is one of the main ways to search in Social Media.

3.2 Social Search

Social Search (or Social Navigation) refers to any direct or indirect social interaction supporting a search (or navigation) process. These interactions can be categorized using two properties: dynamism and personalization [Dieberger et al., 2000]. In the former case, previous users have dynamically changed an information space, leaving navigational traces helping oncoming users to find their way around more easily. When searching for a product on Amazon, the listing of products viewed by users who had also searched for the currently viewed one, leads users onto a path taken by peers with a similar interest. In the latter case, users perceives search advice to be personalized, sometimes even a direct interaction is involved. When searching the IIT Galvin Library's website, users can get personalized contextual support via

[4]http://twitter.com

instant message from a library staff member.

Users can benefit from social input at any point during the search process: while developing a search strategy, selecting search terms or evaluating results [Evans and Chi, 2008].

While Social Search is not necessarily applied to Social Media, it would certainly benefit from the active user involvement and interaction inherent to Social Media. Social Bookmarking for example is a type of Social Search tool in itself. Users can improve their queries by looking up tags used on relevant links by previous users, thus expanding their own vocabulary [Klaisubun et al., 2007].

In cases when it is difficult to form a successful query, users resort to satisfying an informational need by directly asking their peers through Twitter or a status message on their Social Network. Although response times might be longer, it can be more efficient than trying numerous queries with a search engine [Morris et al., 2010] and it also yields fewer results to sort through. Generally, responses from a small group of people connected to users through a trust relationship or common interests will also yield more relevant results [Haynes and Perisic, 2009, Pirolli, 2009].

3.3 Real-time Search

The availability of real-time information is one of the major advantages of Social Media. New posts or sites don't have to be found by a crawler and indexed afterwards, but can be published instantly. Social Media sites were the first to adopt Real-time Search, Twitter being the pioneer. However, Real-time Search is only able to provide a snapshot of recent information with little filtering and is therefore only appropriate when looking for timely information not necessarily provided by experts [Geer, 2010].

Jansen et al. [2010] show in their study that users indeed use Real-time Search in Social Media to keep up to date on a certain topic, repeating queries on a regular basis, sometimes multiple times a day, and usually search for current information in technology, entertainment and politics.

3.4 Multi-site search

Social Media users seem to commonly visit more than one site or application, be it in the same session or in subsequent sessions. There are several possible scenarios for this behavior, for example users might choose different sites for different informational needs, like searching Twitter for information about current events or looking up people on Facebook. Another scenario might be that users are lead from site to site while browsing. For example, a Twitter message might point to a blog post which in turn contains a tags cloud leading to the author's social media bookmarks or a link to a Facebook profile.

Current research by Gamon and König [2009] on how users navigate to, from and in between Social Media sites shows that users often directly choose Social Media sites when starting a session and quite frequently end their session there as well. This suggests a very controlled and focused use with satisfying results.

Unfortunately, the study doesn't distinguish between different types or sites, "Blogs" is the only category separate from the general "Social Site" category. The authors also point out that the data was solely collected from users of the Microsoft Live toolbar used to gather the data, and therefor might be biased. Nevertheless, a similar study with a more diverse user group and a more detailed categorization of sites might yield more insights into the way users frequent a number of sites. It could provide information about the sequence in which sites are visited, and evidence pointing to a habitual sequence or one informed by the informational needs at hand.

4 What do users search for?

Social Media covers almost every topic area that can be found on the Internet, so obviously queries are just as diverse as in general web search. In order to expose topics particular to Social Media search, it is interesting to compare the types of information from queries in each search area. Obviously, users searching Social Media have a certain interest in information about people. While web search queries are mostly directed at persons of public interest ("notable persons"), such as celebrities or experts on a certain topic, Social Media search queries are often about persons of personal interest that users

have an offline relationship with or that share views or interests [Kumar and Tomkins, 2009].

By far the most common target for web search queries are organizations (companies, non-profit, governmental), followed by notable persons, specific products and media titles (movies, CDs) [Kumar and Tomkins, 2009]. Social Media queries in contrast are targeted at current developments in technology, politics and entertainment [Jansen et al., 2010].

Furthermore, users are often interested in content from a particular user or group of users, regardless of the nature of the content [Smith et al., 2008, Cho and Tomkins, 2007]. Users might be looking for updates from favorite content creators because it is of importance for their offline relationship or they want to learn more about a person or simply for entertainment purposes.

The nature of the questions to be answered by the search process also influences if users choose to perform a search in Social Media. Questions requiring an opinion or recommendation are readily processed through Social Media search, but as soon as the question is of a more personal nature (e.g. health, finances) web search is perceived as more objective and therefor the preferable means to obtain the search goal [Morris et al., 2010].

5 Adaptions of global, on-site and specialized search engines

The most noticeable adaptations have been made in the area of Real-time Search. Google recently added several options to limit search results to a specific time frame or display unranked results strictly by recency. New results are added automatically when they are posted. Specialized Real-time Search engines like OneRiot[5], Scoopler[6] or Collecta[7] offer a similar service [Bradley, 2008].

General Social Media search engines don't seem to be very much in

[5]http://www.oneriot.com
[6]http://www.scoopler.com
[7]http://www.collecta.com

demand. In fact, the only one I could find that's trying to span a variety of Social Media and enabling controlled Multi-site Search is WhosTalkin[8]. Specialized search engines like Technorati[9] for blogs, boardreader[10] for forums and bulletin boards or Twitter's own search engine provide narrowed down results, but are not integrated at all.

Social Search elements are included by most search engines, global and internal. Google offers a list of related search terms at the end of each result page as well as trending topics on GoogleTrends. Twitter allows users to sort users they are following into list, enabling them to view updates filtered by personal criteria.

6 Further possibilities to improve Social Media Searches

A fairly obvious possible improvement would be a true Multi-site Search engine that enables users to choose their preferred site for each query. Current Real-Time Search applications do not offer this option, so most of the time Twitter results dominate the top of the result pages, since these tend to be the most recent. The Social Media Search engine WhosTalkin enables the users to view results by origin, but unfortunately doesn't include some of the most widely used Social Media sites (Digg, Facebook, delicious). The next step could be to offer an option to weigh the preferred sites based on trust or personal usefulness.

The Social Search aspect could be enhanced by integrating search with synchronous or asynchronous communication tools. To turn search from an isolated to a collaborative process, by using visualization tools, like the semantic neighborhood radar proposed by Papagelis et al. [2008], to raise awareness of other users conducting similar searches. Villa et al. [2008] show that collaboration, while not significantly improving search performance, did reduce search effort.

[8] http://www.whostalkin.com
[9] http://technorati.com
[10] http://boardreader.com

Furthermore, visualization like social graphs could be used as a supplement on search result pages to facilitate choosing search results [Donaldson et al., 2008].

Another way of improving relevance of search results could be to use clues based on a user's social network [Haynes and Perisic, 2009]. So if, for example, the user searched for "jaguar" and a lot of his friends belong to car fan clubs, it is probable that the user was searching for the car model rather than an animal species. Results can be ranked accordingly. One problematic issue of this approach is privacy, the majority of users would have to willingly expose this type of information, which is not very likely at the moment.

7 Conclusion

Although a number of studies concerning Real-time Search and Social Search/Social Navigation, there is very little material about search in Social Media specifically. The same observation can be made about studies regarding Social Media Use across different sites and the relevance of Social Media portfolios.

Improvements benefitting global Social Media Search seem to be made by applying known problems from related disciplines like web search or by generalizing results from studies about individual sites or types of Social Media. The absence of a widely used Social Media search engine suggests that this approach is not producing satisfying results.

Before any legitimate suggestions for improvement on Social Media Search can be made, sufficient data about Social Media use needs to be collected. Since Social Media sites a often used for personal purposes and privacy is a concern to the majority of users, information about activities on these sites is not released for public access. Therefore, it is difficult to link profiles from different Social Media sites to an individual user, making any automatic data collection is impossible. More personal methods like surveys, interviews and observation would be a better starting point.

References

Phil Bradley. Search engines: A selection of social media search engines. *Ariadne UK*, (57), 2008.

Junghoo Cho and Andrew Tomkins. Guest editors' introduction: Social media and search. *IEEE Internet Computing*, 11(6):13–15, 2007.

A. Dieberger, P. Dourish, K. Höök, P. Resnick, and A. Wexelblat. Social navigation: techniques for building more usable systems. *interactions*, 7: 36–45, November 2000.

Justin J. Donaldson, Michael Conover, Benjamin Markines, Heather Roinestad, and Filippo Menczer. Visualizing social links in exploratory search. In *HT '08: Proceedings of the nineteenth ACM conference on Hypertext and hypermedia*, pages 213–218, New York, NY, USA, 2008. ACM. ISBN 978-1-59593-985-2. doi: http://doi.acm.org/10.1145/1379092.1379132.

Brynn M. Evans and Ed H. Chi. Towards a model of understanding social search. In *SSM '08: Proceeding of the 2008 ACM workshop on Search in social media*, pages 83–86, New York, NY, USA, 2008. ACM.

Michael Gamon and Arnd König. Navigation patterns from and to social media, 2009.

D. Geer. Is it really time for real-time search? *Computer*, 43(3):16 –19, March 2010.

Jonathan Haynes and Igor Perisic. Mapping search relevance to social networks. In *SNA-KDD '09: Proceedings of the 3rd Workshop on Social Network Mining and Analysis*, pages 1–7, New York, NY, USA, 2009. ACM.

Liangjie Hong and Brian D. Davison. Wanted: A unified model for search in social media. In *Informal Proceedings of The Third Annual Workshop on Search in Social Media (SSM 2010)*, 2010.

Bernard J. Jansen, Gerry Campbell, and Matthew Gregg. Real time search user behavior. In *CHI EA '10: Proceedings of the 28th of the international*

conference extended abstracts on Human factors in computing systems, pages 3961–3966, New York, NY, USA, 2010. ACM.

Piyanuch Klaisubun, Phichit Kajondecha, and Takashi Ishikawa. Behavior patterns of information discovery in social bookmarking service. In *Web Intelligence, IEEE/WIC/ACM International Conference on*, pages 784 – 787, 2007.

Ravi Kumar and Andrew Tomkins. A characterization of online search behavior. *IEEE Data Eng. Bull.*, 32(2):3–11, 2009.

Meredith Ringel Morris, Jaime Teevan, and Katrina Panovich. What do people ask their social networks, and why?: a survey study of status message behavior. In *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems*, pages 1739–1748, New York, NY, USA, 2010. ACM.

Athanasios Papagelis, Manos Papagelis, and Christos Zaroliagis. Iclone: towards online social navigation. In *HT '08: Proceedings of the nineteenth ACM conference on Hypertext and hypermedia*, pages 237–238, New York, NY, USA, 2008. ACM.

Peter Pirolli. An elementary social information foraging model. In *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems*, pages 605–614, New York, NY, USA, 2009. ACM.

Marc Smith, Vladimir Barash, Lise Getoor, and Hady W. Lauw. Leveraging social context for searching social media. In *SSM 08*, 2008.

Robert Villa, Nicholas Gildea, and Joemon M. Jose. A study of awareness in multimedia search. In *JCDL '08: Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries*, pages 221–230, New York, NY, USA, 2008. ACM.