A Mutual Feedback Search Scheme on Real-time Web

Yuqing Mao[1,2]
maoy0002@e.ntu.edu.sg

Haifeng Shen[2]
haifeng.shen@flinders.edu.au

Chengzheng Sun[1]
czsun@ntu.edu.sg

[1] School of Computer Engineering, Nanyang Technological University, Singapore

[2] School of Computer Science, Engineering & Mathematics, Flinders University, Australia

ABSTRACT

Real-time information has become ubiquitous on the Web with the growing use of social media today. The more user-generated content is published, the higher are the demands of searching for required information from such content in real-time. However, both conventional search engines and real-time search engines still follow the original search scheme and ignore the characters of real-time Web, where seekers and providers are online simultaneously in real-time search processes such that they can acquire feedback from each other. In this paper, we present a novel mutual feedback search scheme on real-time Web, which can make seekers be satisfied for receiving information required and providers be acknowledged for publishing information required at the same time. This scheme has been implemented in a prototype system, which uses pre-structured epistemology to represent a seeker's information needs and manage the mutual feedback between seekers and providers. Preliminary usability testing has shown that the proposed scheme is particularly effective for real-time Web search where both seekers and providers take their choices through mutual feedback based on a structured information package.

Author Keywords

Micro-blogging, information seeking, real-time Web search, mutual feedback, epistemology.

ACM Classification Keywords

H.3.3 Information Storage and Retrieval: Information Search and Retrieval. H5.3. Information interfaces and presentation (e.g., HCI): Group and Organization Interfaces.

INTRODUCTION

Real-time Web [9] (e.g. *Facebook*'s newsfeed or *Twitter*'s tweet) has achieved a resounding success in recent years. With the increasing popularity of micro-blogging systems (e.g. *Twitter*), more and more breaking news was actually first reported through *Twitter* or other micro-blogs. The reason is that publication and circulation of user-generated information are extremely easy and rapid in social media comparing to conventional information publication. For example, on the 7th January 2010 morning, a modest but noticeable earthquake of magnitude 4.1 rattled San Francisco. In no time, *Twitter* was inundated with tweets about the event, but the link to the official news of the event released from *United States Geological Survey* was only discovered by *Google* web page search 24 minutes after [2].

The rapid growth of real-time Web also attracts more information seekers demanding effective ways of searching for required information from such huge volume of the user-generated real-time content. Conventional search engines (e.g. *Google* or *Bing*) are inadequate in searching for real-time content because they crawl and index web pages periodically so that they cannot follow after real-time content well.

Therefore real-time search, including those specialized real-time search engines such as *Collecta*[1], *OneRiot*[2], *Scoopler*[3], and *Twitter Search*, together with the real-time search services provided by most conventional search engines, e.g. *Google* [6], has gained momentum recently. Real-time search can find the user-generated information, e.g. activities, blog posts, newsfeeds, and tweets, as it is produced on a range of social media systems.

However, real-time search still follows the original search scheme as conventional search engines, where the seekers submit queries to a search engine and receive the results from the search engine based on their relevance to the queries judged by the search engine. Such scheme has been adopted for decades and is still adequate for conventional Web page search. Nevertheless, it is not effective on real-time search because it ignores an important character of real-time Web which implies that the content is generated online by information providers. That is definitely different from conventional Web search, where the content of Web pages is supposed to be self-contained and generated offline.

On real-time Web, when a seeker is searching for real-time information, most providers of the relevant information are constantly updating the content and it is very likely that the

The 11th International Workshop on Collaborative Editing Systems (IWCES'11), Jointly with ACM CSCW, March 20, 2011, Hangzhou, China.

[1] http://collecta.com

[2] http://www.oneriot.com

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

providers are online and ready to communicate with the seeker in real time. Therefore during the search process, as the information seeker and provider are online at the same time, it is possible for them to receive feedback from each other in real time.

On the one hand, the provider's feedback can ensure the seeker's needs are satisfied because real-time interactivity is effective for the information seeker acquiring information directly from the provider. On the other hand, seeker's feedback can help the provider publish high quality content because a general purpose of publishing content on social media is for attention, e.g. followers on tweet or comments of blogs, and the provider will also be satisfied in nature if the published content is acknowledged.

In this paper, we present a novel mutual feedback search scheme on real-time Web, which can make seekers be satisfied for receiving information required and providers be acknowledged for publishing information required at the same time, by supporting the mutual feedback between seekers and providers in the real-time search processes.

In this mutual feedback search scheme, a seeker initiates a real-time search process by making a pre-structured epistemology [5] to represent her/his information needs on a specialized real-time search engine, which automatically collects relevant results according to the structure from a range of social media systems in real time. The seeker and the information provider(s) then receive feedback from each other on the filled fields in real time so that effective responses can be achieved in the search process.

The scheme has been implemented in the *Baijia* prototype system. For the purpose of conducting internal usability testing, we have also implemented a micro-blogging system for information providers to publish real-time content. Preliminary usability testing has shown that the proposed scheme is particularly effective for real-time Web search where both seekers and providers take their choices through mutual feedback based on a structured information package.

The rest of this paper is organized as follows. First we describe some work related to real-time search. Then we sketch the mutual feedback search scheme and present a prototype real-time search engine that implements the proposed scheme, followed by discussions of a preliminary usability testing. Finally we conclude the paper with a summary of major contributions and future work.

RELATED WORK

Real-time Web refers to the practices that enable users to receive information as soon as it is published, mainly involving the increasingly popular micro-blogging systems and social networking sites, which is different from real-time communication or collaboration tools such as instant messaging or Google Wave [3], where the information is shared between users of a predefined group rather than published to general public.

Real-time search is therefore a new compelling area that allows a seeker to acquire information immediately from the vast majority of such content on the Web. For example, users can search for latest tweets posted on *Twitter*, or friends' newsfeed on *Facebook*, with their built-in real-time search services. Some third-party real-time search engines also emerged such as *Scoople*, *Collecta*, *OneRiot*, or *Status Search* [8], where public users can retrieve real-time updates on a specific topic from various social media systems such as *Twitter* and *Facebook.*

However, these real-time search engines still follow the original search scheme as conventional search engines - they simply retrieve real-time news through micro-blogging API based on a seeker's query and then display the results to the seeker. What they essentially overlook is a character of real-time Web which implies the seeker and the provider are online at the same time so that it is possible to display their feedback to each other on both sides in the search process. Such mutual feedback will benefit the effective real-time search and satisfy both seeker and provider.

It is worth discussing some practices that follow different search schemes, although they are not specially designed on real-time Web. For instance, Q&A (Question and Answer) systems such as *Yahoo! Answers*[4] allow a seeker directly posts her/his question on the system and provider(s) will prepare their answers accordingly. The human-powered search engine *ChaCha*[5] allows a seeker to directly interact with a real person (a search agent instead of an information provider) to communicate her/his information needs.

In summary, while existing work hasn't seized the essential of real-time Web, the proposed mutual feedback search scheme makes full utilization of the character of real-time Web where the seeker's search activity and the provider's publish activity occur simultaneously. That is, a seeker could get feedback from provider(s) to know the details about her/his information needs, while provider(s) could get feedback from seeker(s) to know what information is required and polish her/his real-time content.

THE MUTUAL FEEDBACK SEARCH SCHEME

Overview

Figure 1 Left shows the search scheme of a conventional search engine, which needs to use a page index database to index every Web page crawled from various Web servers so that a seeker's queries can generate matches from the database. Similarly, a real-time search engine following this scheme needs to use a real-time content base to temporarily hold user-generated real-time information retrieved from social media sites so that a seeker's queries can generate matches from the content base, if it does not retrieve such information directly through micro-blogging systems' API.

[4] http://answers.yahoo.com

[5] http://www.chacha.com

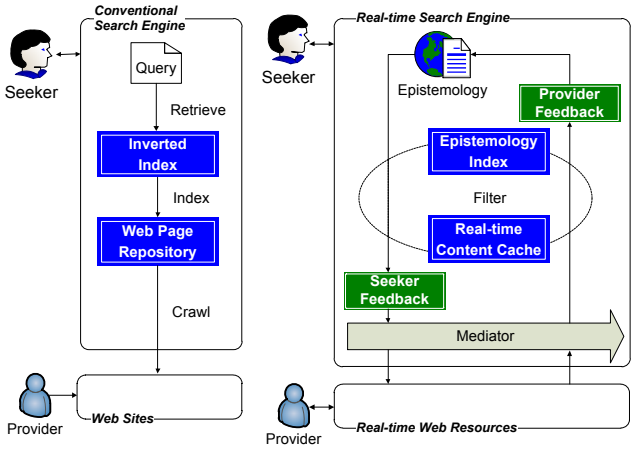

Figure 1. Comparison of two search schemes

Left: Scheme of conversional search engines;
Right: The mutual feedback search scheme

Figure 1 right shows the proposed new search scheme. In this scheme, seekers and providers can get feedback from each other so that the real-time search is more effective and both seekers and providers will benefit from the search process. The implementation of mutual feedback is based on the epistemology, which describes a seeker's information needs for a specific topic, filled with data records the interaction between the seeker and providers around that topic. The search process and key components in this scheme are described as followings:

A seeker conducting a real-time search first creates a pre-structured epistemology depicting a blueprint of the search process. The *Filter* component will then retrieve the *Real-time Content Cache* which cached user-generated real-time content from external social media systems, to discover real-time information that is relevant to the epistemology created by the seeker. At the same time, while a provider is publishing real-time content, the *Filter* component retrieves the *Epistemology Index* and finds epistemologies that can be filled out with the relevant content.

After the epistemology has been updated for both seeker and provider, the seeker can submit any enquiry about the real-time content as *Seeker Feedback* on the epistemology, which will be dispatched to the provider by the *Filter* component as it has collected the contact information of the providers whose content matches the epistemology. The provider can also submit more specified content as *Seeker Feedback* on the epistemology, which will then be filled into the relevant fields of the epistemology.

In addition, the real-time search engine is bridged with external real-time Web resources provided by micro-blogging users or social networking friends by the *Mediator* component.

A New Real-time Search Engine

We have designed and implemented a new real-time search engine following the proposed scheme in the *Baijia* prototype system. To conduct internal usability testing, we have also implemented a micro-blogging system for information providers to publish real-time content. The external micro-blogging system is connected to the search engine via the *Mediator* component within the search engine by using the micro-blogging system's API.

The search engine implements the user interfaces, processing algorithms, e.g. language-model-based similarity matching [1], operational-transformation-based concurrency control for web-based cooperation [7], and user-interest-based personalized information filtering [4], and data structures, e.g. epistemology, epistemology index, and real-time content cache for the components in the proposed scheme. In this section, we will discuss some user interface features, while processing algorithms and data structures are beyond the scope of the paper.

Interface of Feedback for the Seeker

Suppose a seeker is conducting a real-time search for information about the 2010 World Cup final, she/he has created a pre-structured epistemology titled "*World Cup Final*", as shown in Figure 2. The *Filter* component will update the epistemologies by filling each user-defined field, e.g. "*Spain in World Cup*" or "*Netherlands in World Cup*", with relevant real-time content. The content is organized as multiple threads based on different providers. Therefore those providers can work concurrently while the structure of the epistemology remains unchanged. That could make the epistemology more readable than search results ordered by time.

Thereafter, the seeker is able to submit feedback to the provider(s) in real-time. For example, the seeker can comment on the thread *"Pele: Spain Are Favorites to Win World Cup"* by *"How about the prediction of Octopus Paul"*. In real-time, if the provider has got such required information, she/he can response to the feedback by publishing another piece of real-time content, and the *Filter* component will update the epistemology accordingly thus the seeker is able to receive the provider's feedback immediately.

It is worth clarifying that although we implemented a micro-blogging system mainly for the purpose of conducting internal usability testing, with the *Mediator* component, the real-time search engine can talk to various external social media system via their APIs, e.g. *Twitter*, *Facebook*, and so on. For example, a seeker can broadcast a pre-structured epistemology to her/his friends on *Facebook* in order to ask them provide real-time information she/he is looking for.

Interface of Feedback for the Provider

Figure 3 shows the interface of the micro-blogging system connected with the real-time search engine. Providers can easily publish real-time content through such a *Twitter*-like interface. As micro-blogging users always prefer a lightweight and concise interface, each epistemology

Figure 2. Interface of epistemology-based feedback for the seeker

dispatched from the search engine is succinctly presented as a short message.

While a provider is publishing real-time content, relevant epistemologies dispatched from the search engine are being filtered in and displayed in real time so that the provider knows which information to publish based on the feedback of seekers. For example, suppose a provider is publishing a piece of information about *"Spain sweat over Villa injury"*, the micro-blogging system immediately sends it to the search engine, which in turn filters all pre-structured epistemologies to find those asking for such information, e.g. the epistemologies titled "*Spain World Cup"* and tagged with "*injury"*.

The search engine then dispatches these matched epistemologies with seekers' feedback to the micro-blogging system and posts a micro-blog there, e.g. *"Currently interested information about #Spain #injury: #substitution(12) #squad(5) #influence(1)"*. The provider can judge which kind of real-time content is most interested (i.e. many seekers are interested in the substitution for Villa) according to seekers' feedback based on the number following each tag indicates, which is calculated as:

$$N = N_{epi} - N_{rec}$$

Figure 3. The micro-blogging interface

where N_{epi} is the number of epistemologies are expecting real-time content tagged by that keyword, and N_{rec} is the number of providers who has published any real-time content tagged by that keyword.

USER STUDIES

We conducted a preliminary internal usability testing of the real-time search engine in order to: 1) understand whether users (i.e. information seekers and providers) like this new concept and if yes what features they particularly like, 2) investigate what kind of search tasks where the proposed search scheme does better than the original search scheme, 3) study whether the system is easy-to-use and what special skills users need to use the system, and 4) get some feedback to improve the system.

Subjects, Tasks and Methodology

Five postgraduate students participated in this small-scale user study. Two of them have the experience of using real-time search services and three of them are current users of *Twitter* and have the experience of publishing real-time content on *Twitter*.

To investigate what kind of search tasks where the proposed search scheme might do better than the existing search scheme, we deliberately designed two real-time search tasks.

The first scenario was searching for real-time content about the *World Cup Final*, where we particularly focused on the real-time feedback. As the user study was conducted during the World Cup Final period, real-time was critical because any outdated content was meaningless.

The second scenario was searching for real-time content about the *new generation iPhone*, where we particularly focused on the mutual feedback. Because the seekers would like to get feedback from a current owner to know more details about that phone, and the providers would also like to publish relevant content according to the feedback from seekers.

The same set of participants did the two scenarios first with *Twitter Search* and with the real-time content published on *Twitter* and then with the *Baijia* real-time search engine and with the real-time content published on our micro-blogging system. They were not allowed to use any other

communication channels, e.g. phones or instant messengers, except the systems given to them.

At the end of the testing, participants were asked to fill out a questionnaire surveying novelty, quality, efficiency, and usability of the system in the scale from 0 (extremely disagree) to 5 (extremely agree) as well as the strengths and weaknesses of the system.

(1) How novel is the *Baijia* real-time search engine as compared to *Twitter Search*?

(2) How high is the quality of search results using *Baijia* as compared to using *Twitter Search*?

(3) How efficient is *Baijia* as compared to *Twitter Search*?

(4) How easy-to-use is *Baijia* as compared to *Twitter Search*?

(5) What are the strengths of *Bajia* as compared to *Twitter Search*?

(6) What are the weaknesses of *Bajia* as compared to *Twitter Search*?

Scenario 1

In this scenario, the participants were situated in context of the 2010 World Cup final. Two participants were designated as seekers to search what they were interested in the match, while the other three participants were designated as providers to publish real-time content.

Figure 4 shows that the participants were generally impressed by the system. The novelty of the system was highly appraised, while the usability of the system still has a big room to improve, which was quite expected as the system is still at its early stage. We also interviewed the participants in order to understand their views on the system.

Participants were very pleased with the experience that they could actually give feedback to providers to publish what they were looking for. For example, a seeker first submitted a query "goal" in *Twitter Search*, but she/he found there was no information about who assisted the goal in the search results. There is no result for another query "assist" at that time so that she/he had to wait until any provider published such information. In contrast, using the *Baijia* search engine, the same seeker added tags "goal", "assist" to a pre-structured epistemology, and when a provider who was publishing the real-time content - *"Iniesta grabbed the winning goal!"* noticed that someone was interested in knowing the assist, she/he then published another piece of real-time content - *"Fabregas provided the crucial assist for Iniesta's winning goal"*.

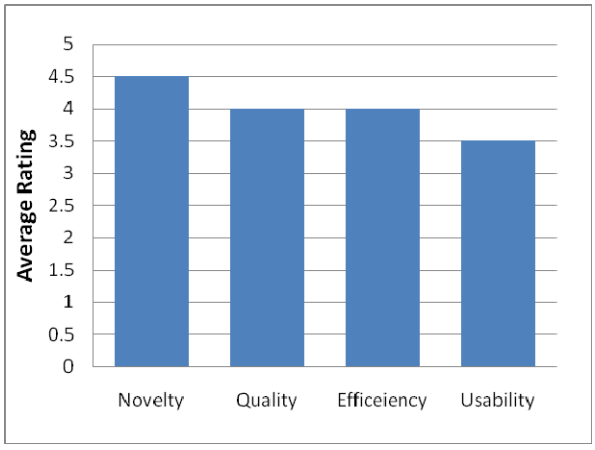

Figure 4. Average rating in scenario 1

The providers were also very impressed by the real-time feedback from the seekers. One participant commented on that feature during the interview:

"When I was posting the information about De Jong's kicking Alonso on the chest, I saw that someone also wanted to know about the red card. Why not publish something relevant because I also thought he should receive a red card? Then I posted: That should be a red card, but the referee is under pressure. Immediately that seeker and I started discussing about the referee."

As for the quality of search results, all participants agreed that *Baijia*'s search results were much better structured than *Twitter Search* and some participants did want that structure for their search tasks.

As for the efficiency of search, the seekers totally submitted 16 queries in *Twitter Search* and receive 53 real-time tweets published by the providers. In contrast, each seeker created an epistemology of 11 fields in *Baijia* search engine and received 60 pieces of real-time information published by the providers in our micro-blogging system.

As for the usability of the system, most participants thought *Baijia* search engine was generally easy to use, although a lot more could be improved. This was somewhat unexpected because the system is still at its early stage. We attribute this to the participants' skills in using various real-time search and micro-blogging systems.

Scenario 2

In this scenario, three seekers were interested in buying the new generation iPhone. However, they have been bored by the perpetual advertisements and stereotyped reviews on the Web. Therefore, they turned to search for real-time content about the latest and just critiques on the product. Two providers were either the current owners or technical experts of the product.

Figure 5 shows the evaluation results and by comparing Figure 5 with Figure 4, we found that the participants in this scenario were more satisfied with the quality of search results but less satisfied with the efficiency of search than those in the previous scenario.

We observed a lot of feedback between the seekers and providers in this scenario. This is because the seekers were not clear about what they were looking for in the beginning, but when they knew the provider had already bought an

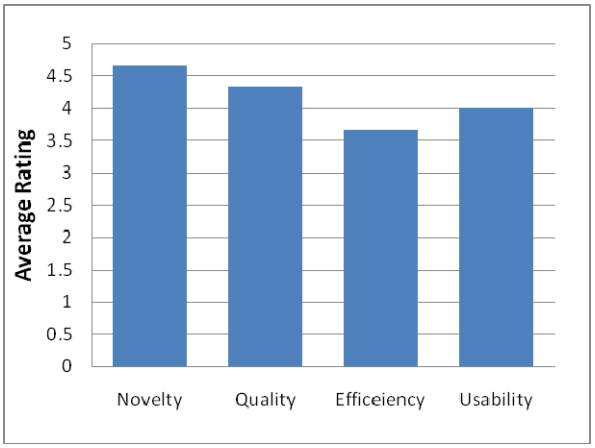

Figure 5. Average rating in scenario 2

iPhone, they would like to know more from the provider as the interaction went on.

For example, a seeker got a piece of information about "*iPhone antenna problem*" from a provider in the pre-structured epistemology for "*the latest iPhone*". As the seeker really cared about such flaw, he immediately requested more about the detail of that problem. After reviewing the real-time update from the provider, the seeker wanted to know how to solve the problem, and on and on.

The seekers were more satisfied with the quality of search results in this scenario as they got a lot of valuable information from our system. As a consequence, the search efficiency could possibly be compromised by too many extra queries or unsuccessful queries. In contrast, they could not get comparable information from *Twitter Search* because such information, e.g. the provider's personal views, could only be acquired through continuous feedback between the seeker and provider, which is hardly achievable using *Twitter Search.* One participant commented during the interview:

"In Twitter, I only focus on what I published. I can't see other's feedback unless I follow them. However, I have no reason to follow a seeker I don't know, although I am glad to get feedback from them and reply to them in time. I think the interaction in Baijia is also convenient for me, not only for the seekers."

The results of the user studies have indicated that: 1) the concept is novel and the mutual feedback search scheme is the strength of our system, 2) our system is particularly more effective for a seeker to acquire a structured information package consisting of diverse but coherent information through real-time feedback from and to providers, and 3) the system is generally easy to use, particularly for users who have the experiences with real-time search services and micro-blogging systems, but it certainly has a big room to improve. We are conscious that we need to conduct more rigorous evaluation tasks ahead.

CONCLUSION AND FUTURE WORK

Real-time Web has grown in popularity with the rising of various micro-blogging systems and social networking sites. Conventional search engines generally do not accommodate real-time content, while specialized real-time search engines still follow the original search scheme where the seeker can only receive information returned by the search engine passively and the provider has nothing to do with the seeker in the search process.

The proposed search scheme fully takes advantage of the character of real-time Web by supporting mutual feedback between seekers and providers in real-time search process. Therefore the seeker can get required information more effectively and the provider can publish more valuable real-time content. Furthermore, we have implemented the *Baijia* prototype system following the scheme, which uses pre-structured epistemology to represent a seeker's information needs and also manager feedback between seekers and providers in search processes. Initial internal usability testing of the system is positive, but we are conscious that we need to conduct more rigorous evaluation tasks ahead.

We are improving the prototype system based on the initial user feedback. We are working towards connecting our search engine with public social media systems, e.g. getting real-time content through *"Twitter Firehose"* so that we can conduct more realistic and rigorous evaluation tasks.

REFERENCES

1. Fang, H. and Zhai, C. Semantic term matching in axiomatic approaches to information retrieval. In *Proc. SIGIR '06*. ACM Press (2006), 115-122.
2. Google real-time search: 6 min. to spot quake. http://news.cnet.com/8301-30685_3-10428590-264.html
3. How people are using Google Wave. 2010. http://wave.google.com/using-wave.html.
4. Lam, W., Mukhopadhyay, S., Mostafa, J., and Palakal, M. Detection of shifts in user interests for personalized information filtering. In *Proc. SIGIR '96*. ACM Press (1996), 317-325.
5. Mao, Y., Shen, H., and Sun, C. Supporting exploratory information seeking by epistemology-based social search. In *Proc. IUI '10*. ACM Press (2010), 353-356.
6. Shankland, S. Relevance meets the real-time web, 2009. http://googleblog.blogspot.com/2009/12/relevance-meets-real-time-web.html.
7. Shen, H., Xia, S., and Sun, C. integrating advanced collaborative capabilities into web-based word processors. In *Proc. CDVE'07*. Y. Luo, Ed. LNCS. Springer-Verlag (2007), 1-8.
8. Sullivan, D. What Is Real Time Search? Definitions & Players. 2009. http://searchengineland.com/what-is-real-time-search-definitions-players-22172.
9. Wikipedia. Real-time web, 2010. http://en.wikipedia.org/wiki/Real-time_web.