http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (20): 5840-5844, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.5840.5844

Research on SNS Real-time Search Sorting Algorithm Based on User Context

Qingfeng Wu
Software School, Xiamen University,

Abstract: With the development of mobile social network, traditional search engines have not met the users' needs of real-time search because of inadequacy of real-time information such as users' location and events. A real-time search model based on mobile social network is introduced in this study. According to the characteristics of social network under multi-element context, the model sorts the search results by user context such as time, location, relation and attention and user weight. Taking PageRank algorithm based on hyperlinks page rank as reference, a new algorithm, SocialRank is put forward in user weight counting. The experiment results show that the model improves real-time and relevance of search results greatly in time and location.

Key words: Multi-element context, real-time search, mobile social, sorting algorithm, social rank

INTRODUCTION

Mobile social network is transforming from the service based on context (photos, videos, states, notices and comments) to the based on location and event. With the transformation, the elements including events, people, location and time will be integrated into a context model (Qing, 2010). Traditional search engines have not met the users' needs of real-time search for social network in a content degree because that they can't mining users' real-time information adequately such as location, event and time. For example, after the China-Africa Huangyan Island Event users input Huangyan Island in traditional search engines, what are showed may be the geography overviews of Huangyan Island or the latest news of confrontation of Huanyan Island but users possibly aren't interested in that information. Taking another example, user searches food when they go to a tourist area, what he expects is the local information of the most famous food but not of other areas. The former example shows the users' need of latest news and the later one shows the users' need of real-time location.

Recently, most social networks have realized the real-time of information but they just take the timeliness of information into account without considering other user context information. A real-time search model of social network based on multi-element user context which can ensure that persistently and dynamically updated latest news will be provided when user search on a specific social network is put forward in this study. What's more, according to the shortage of social network search engines and the characteristics of the social networks under multi-element context, a search sorting algorithm based on elements such as time, location, relevance, attention and user weight is proposed which is expected to get user search experience better.

Next the above-mentioned search model will be introduced. First key steps of this search model will be described and then the six factors of search sorting algorithm will be introduced. At last the experiment results of this algorithm will be showed.

REAL-TIME SEARCH ENGINE MODEL

For mining context information such as users' location and event in mobile social network adequately and improving the efficient and quality of search engine, a real-time search engine of based on social network under multi-element context is designed in this study. In this search engine, search results will be sorted by relevance to get more effective results.

Model structure: This model is a system prototype based on social network under multi-element context. The real-time newest data will be obtained by the Web Service interface of the system. When user input key words and user's information (multi-element context), the background service program will search in database according to the key words and then sort the search results by the factors such as time, position, relevance, attention and user weight and return the result to the client. This model sends search request periodically to get latest search results. Results will return to users in time once new search results are produced so that the real-time of this model is achieved. The model structure is as Fig. 1 shows.

Flow of data analysis and result sort: A series of operations such as data analysis, data calculating and sorting of results searched preliminarily at server-side is not only the key of the algorithm but also is the core of real-time search. The flow is as Fig. 2 shows.

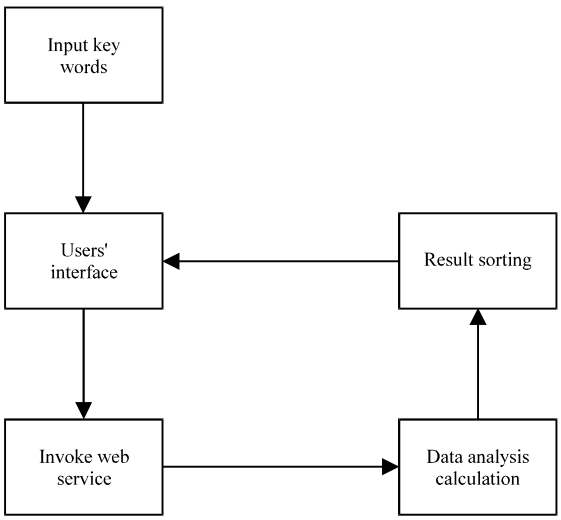

Fig. 1: Model structure

ALGORITHM FACTORS OF REAL-TIME SEARCH SORTING

The mobile social network under multi-element takes users as center and integrates events, people, position, time and other elements into a context model. Relationship between users is built by attention, concern, comment and new forwarding. As a result, sorting of search results should take the following points considering.

Time factor: Time is a very important factor in multi-element context. If one keeps attention on a certain event, what he expects is the latest information. As a result, the latest information takes the priority:

$$T_v = \frac{t\_factor}{now - createtime} \quad (1)$$

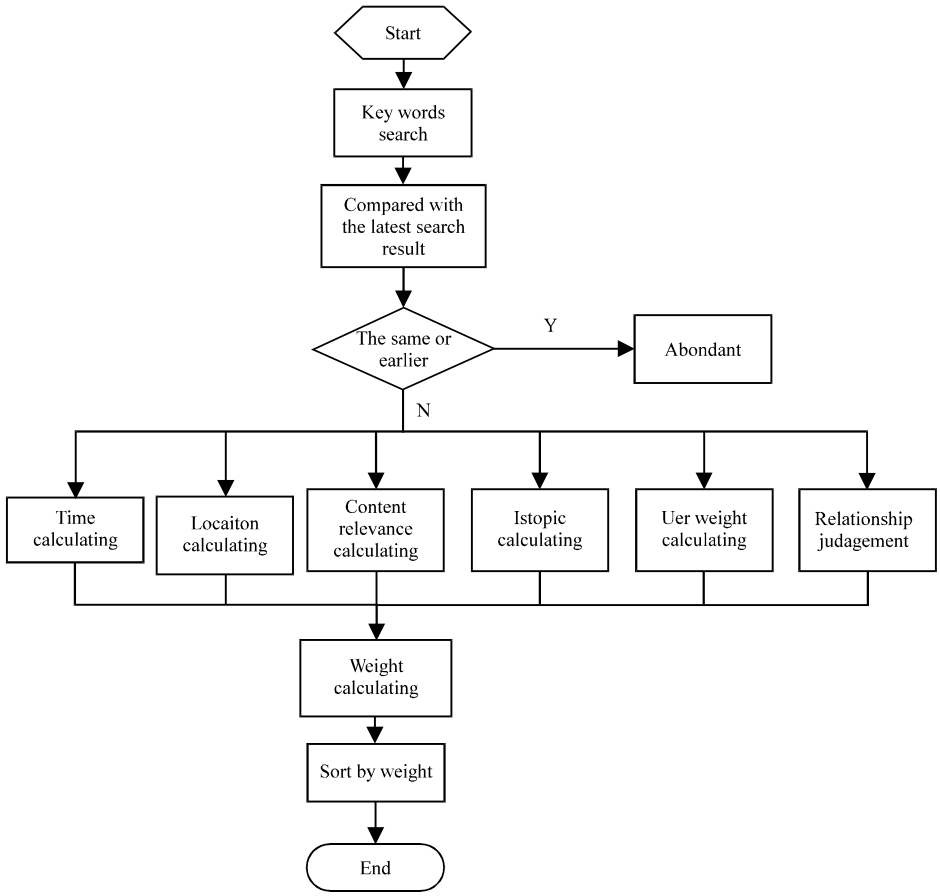

Fig. 2: Flow of real-time search sorting algorithm based on multi-element context

In Eq. 1, now means the time user starts to search, createtime means the time information is created and t_factor means correlation coefficient.

Location factor: If user comes in a new place, what he expects properly is the information matched with key words. For example, user searches food and hotels in a scenic spots, what he expects is the information there. For that, system provides a real-time search option of position fixing. Location factor is added into the search algorithm to affect the order of search results. If searched information is closer with users, the value of location factor will be bigger:

$$L = \frac{loc\_factor}{distance(A, B)} \quad (2)$$

In Eq. 2, distance(A, B) means the distance between user's location(place A) with location of searched information(place B) and loc_factor means the correlation coefficient of locations. The distance is calculated by longitude and latitude. It is 40075.04 km to walk around the earth on the equator. There is 360° a circle and 1° equals 60, so the length of every degree and every minute of the equator is calculated as follows:

$$40075.04 \text{ km}/360° = 111.31955 \text{ km}$$

$$111.31955 \text{ km}/60' = 1.8553258 \text{ km} = 1855.3 \text{ m}$$

As one minute equals 60 sec, the length of every second means 1855.3 m/60 = 30.92 m.

The computational formula of distance between two random locations is:

$$distance(A, B) = 111.12\cos\left\{\begin{matrix}1/[\sin\Phi A\sin F B + \\ \cos\Phi A\cos\Phi B\cos(\lambda B\text{-}\lambda A)]\end{matrix}\right\} \quad (3)$$

In Eq. 3, A's longitude is and latitude is. means B's longitude and means B's altitude and means distance.

Content relevance: How much the key words match with the text searched is assessed by TF-IDF algorithm (Xu *et al*., 2011).TF-IDF(term frequency–inverse document frequency) is a common weighting technology used in information retrieval and text mining. TF-IDF is also a statistical method to assess the importance of one word to a file set or corpora. The importance of one word will increase at a direct ratio with its frequency appears in the file but decrease at an inverse ratio with its frequency appears in corpora (Fei, 2010).The computational formula is as Eq. 4 shown:

$$C_v = \sum_{k=1}^{n} TF_{vk} IDF_k \quad (4)$$

In Eq. 4, n means number of key words for searching, TF_{vk} means the frequency of key word k appears in a certain JSON file and IDF_k means the inverse text frequency of key word k:

$$IDF_k = \lg\left[\frac{N}{DF(w_k)}\right] \quad (5)$$

In Eq. 5, N means the results number of searching and $DF(w_k)$ means the text numbers of results in which key word k appears at least once.

Topic correlation: In this system, user can insert a topic when he publishes fresh news. This function makes it convenient for users to raise a topic, scan hot topic and search topics. Suppose that there are some topics $U(t_{u1}, t_{u2}, \ldots, t_{un})$ in searched fresh news. It has better relevance if the topic collection of fresh news includes the key words. This relevance is represented as IsTopic:

If (keyword $\epsilon U(t_{u1}, t_{u2}, \ldots, t_{un})$
IsTopic = 1 = 1;
Else IsTopic = 0;

Weight of publisher: Weight of user who publishes fresh news is calculated with PageRank algorithm (Gao and Zhou, 2010).In PageRank algorithm, the value of website is assessed by the quantity and quality of external links and internal links. In PageRank, every link means a vote of this page, it means that the more page is linked, the more voted by other websites. At last, the importance of page will be judged by the value of PageRank. The more important page, the former it will be located in search results. In social network, one of user's fresh news is forwarded means that he is voted by this forwarder. If plenty of user's fresh news is forwarded, it means he has great prestige weight in this social network. Meanwhile, one has great prestige weight if a fan of great prestige weight forwards his fresh news. Therefore, SocialRank algorithm proposed in this study imitates PageRank to calculate the prestige weight of users. The weight of users publishing fresh news will be set according to the prestige weight:

$$PR(v) = (1-d) + d\sum_{u \subset f_v} PR(u)/N_u \quad (6)$$

In Eq. 6, PR(v) means the weight of user v who publish the fresh news, d means attenuation factor which

is between 0 and 1and usually is set 0.85, fv means fans forwarded user v's fresh news, PR(u) means the weight of fan u and N_u means the number of users forward fan u's fresh news.

Attention factor: Suppose that user A and user B follow each other in an application, they both take apart in a certain activity but they don't know the other take apart at the same time and they publish fresh news about this activity's topic in social network. Thus they will find that the other is nearby through real-time search of this topic. Then they can chat with each other even meet with each other to have more funny when taking part in the activity. Because of the characteristics of social network, users hope to get the fresh news of attention user related to the topic for search. As a result, in the search results, it should be taken into considering that whether the fresh news is published by attention users. Attention factor is expressed as IfFoucus, v means the user who published fresh news and $U(f_{u1}, f_{u2},\ldots, f_{un})$ means all the attention users of user v. So, it can be described as:

$$\text{If } (v \in U(f_{u1}, f_{u2},\ldots, f_{un}) \text{IfFocus} = 1;$$
$$\text{Else IfFocus} = 0;$$

To sum up, the weigh value of every piece of fresh news in preliminary search results is:

$$W_v = T_v + b.L + C_v + c.\text{IsTopic} + PR(V) + d.\text{IfFocus} + \text{others} \quad (7)$$

The algorithm above meets the demand of the real-time information for user, at the same time, it adequately mining the factors such as users' location, attention and weight to relevance sorting. In Eq. 7, Is Topic means if the key word is the topic of fresh news, If Focus means if the topic is user focuses on. And b, c and d are all means relevance factors and others means other related factors.

EXPERIMENTS AND ANALYSIS

This experiment is operated with correlation method. Input the key word "thesis defense" to search:

- In the situation not using real-time search algorithm, the search result is shown in Fig. 3

From the figures, the search result is just sorted by time but does not take other users' context factors into account when it doesn't use real-time search algorithm introduced in this study.

- When the real-time search algorithm introduced in this study is applied, search result is shown in Fig. 4

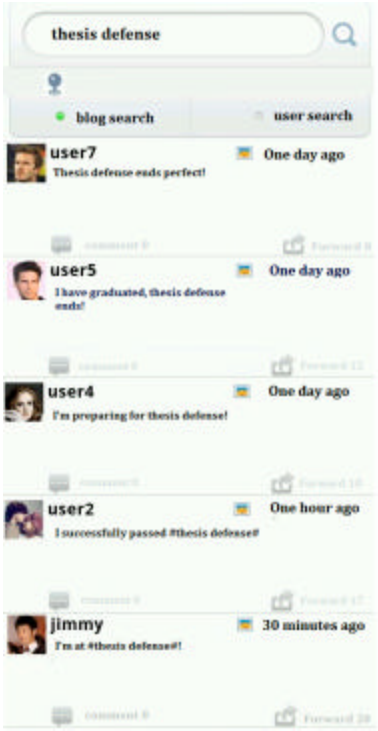

Fig. 3 :Search result 1

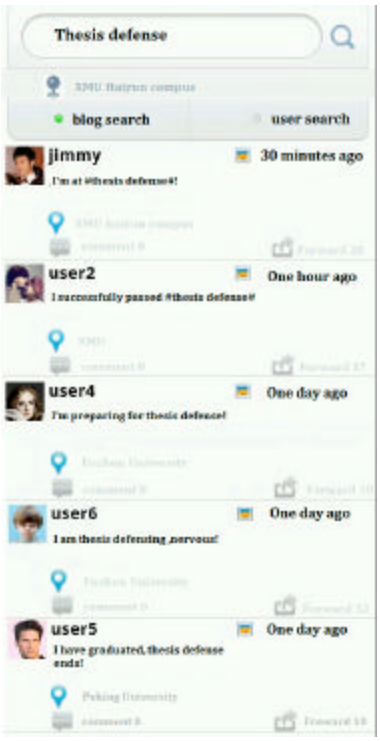

Fig. 4 :Search result 2

Table 1: Users' information

Username	Location	Attention No.	Microblog No.	Fans No.	Topics No.
Jimmy	XMU	384	10	789	34
User 2	XMU	274	6	500	34
User 4	Add4	256	2	349	78
User 5	Add5	178	2	187	28
User 6	Add6	128	1	200	78
User 7	Add7	158	1	86	0

And the information of users who published these search results is as Table 1 shows.

From the Table 1, search result is sorted relatively according users' context. Not only timeliness of news but also other factors such as users' location, relationship, topic, attention and users' weight are taken into account. The top information in search result has the following features: it is published short from the time for searching, its location is the same as user's, the key word for search is one of the topics of it, the user published this information is followed by the search user and from the number of fans we can see that the user' weight is bigger than others'. However, these factors of other information are worse. The bigger the value all the six factors of the information add, the topper the information will be located.

From these experiments, it can be concluded that the real-time search sorting algorithm of mobile social network based on multi-element context not only achieves real-time search result but also improves search efficiency. In this algorithm, the calculation of relevance between search content and topic factor optimizes the relevance and veracity of search, taking time factor and location factor into account adequately take advantage of users' context information to improve the real-time of search result, attention factor and user weight make it fit the characteristics of social network better to get better real-time search service. The time for search is acceptable for user and its performance is quite good. As a conclusion, both feasibility and rationality of this algorithm are verified.

CONCLUSION

Based on mobile social network, the study presents a real-time search model. Then according to the characteristics of social network under multi-element context, the model provides the search results by user context. A new algorithm named SocialRank is put forward in user weight counting. The experiment results show that the model improves real-time and relevance of search results greatly in time and location.

However, this algorithm has its own shortages in efficiency when there are huge users. This is because the algorithm takes PageRank as reference. In the next step, the calculation could be simplified by sparse matrix. What's more, the weight proportion of six factors mentioned above which needs to be further researched.

ACKNOWLEDGMENTS

Supported by the Key Technology R and D Program of Xiamen, Fujian under Grant No. 3502Z20103001, 3502Z20101002; the Key Technology R and D Program of Quanzhou, Fujian under Grant No. 2009G29; the Leading Academic Discipline Program, "Project 211 (the 3rd phase)" of Xiamen University; the Fundamental Research Funds for the Central Universities, Xiamen University under Grant No. 2011121023, CXB2012012, CXB2012013, 201212G007; Shenzhen Key Laboratory for High Performance Data Mining with Shenzhen New Industry Development Fund under grant No.CXB201005250021A.

REFERENCES

Fei, W., 2010. Study of performance evaluation of retrieval function of search engine. Wuhan University, Wuhan, China.

Gao, S. and Q. Zhou, 2010. RefRank: Sorting algorithm based on content quote of search engine. Tsinghua University, China.

Qing, Y.J., 2010. Evolution of mobile social network. November 2010. http://labs.chinamobile.com/mblog/588439_72024

Xu, J., M.N. Kang and G.Y. Dong, 2011. Study of Real-time search engine sorting algorithm based on social network. Northwestern Polytechnical University.