Refine Search Results Based on Desktop Context

Xiaoyun Li(✉), Ying Yu, and Chunping Ouyang

Department of Computer Science and Technology,
University of South China, Hengyang, Hunan, China
lxy.yy@yeah.net, yyingu@sina.com, ouyangcp@gmail.com

Abstract. During a search task, a user's search intention is possible inaccurate. Even with clear information need, it is probable that the search query cannot precisely describe the user's need. And besides, the user is utterly impossible browse all the returned results. Thus, a selected and valuable returned search list is quite important for a search system. Actually, there are lots of reliable and highly relevant personal documents existing in a user's personal computer. Based on the desktop documents, it is relevantly easy to understand the user's current knowledge level about the present search subject, which is useful to predict a user's need. An approach was proposed to exploit the potential of desktop context to refine the search returned list. Firstly, to attain a comprehensive long-term user model, the operational history and a series of time-related information were analyzed to achieve the attention degree that a user paid to a document. And the keywords and user tags were focused on to understand the content. Secondly, working scenario was regarded as the most valuable information to construct a short-term user model, which directly suggested what exactly a user was working on. Experiment results showed that desktop context could effectively help refine the search returned results, and only the effectively combination of the long-term user model and the short-term user model could offer more relevant items to satisfy the user.

Keywords: Desktop context · Information retrieval · User model · Personal information space

1 Motivations

In big data era, the increasing growth of various kinds of digital resources leads to the difficulty and inefficiency to find accurate information that user need. Considering the whole process of information retrieval, the reasons can be concluded into three defects. Firstly, people do not know what they need, that is the search intention is unclear. Secondly, the search intention probably cannot be described accurately and adequately by the short search query. And thirdly, it is totally impossible to check numerous returned result pages.

In order to satisfy the user need in a better way, an approach was proposed to exploit the desktop context information, and two user models were constructed. One is the long-term user model that is responsible for collecting all the data related to a user's interest preference. This model is relatively reliable and comprehensive.

J. Li et al. (Eds.): NLPCC 2015, LNAI 9362, pp. 209–218, 2015.
DOI: 10.1007/978-3-319-25207-0_18

The other one is a short-term user model that is focusing on find out what exactly the user is working on. But this model is temporary and the data is relatively sparse. To achieve a better understanding of the user' search need for a specific task, the short-term model is not sufficient. The experiment results showed that combining the two models helps predict the user need and refine the search returned results.

2 Related Works

To overcome those inherent defects mentioned above, many research works contributed to improve the performance of retrieval systems. Most works were based on collecting implicit feedback information to gather user personal information. For example, Mouse dynamics were regarded relatively unique from person to person by Zheng [1], so they used the point-by-point angle-based metrics of mouse movements for user authentication. They proved that an ordinal regression model for user feedback could greatly improve the accuracy of a recommender system. And a user clicking on an item has been proven as implicit evidence that the user was interested on it [2,3,4]. Lee [5] even found that a user's first search result click could provide valuble insight into this user's subsequent interaction with the returned result list. Radinsky [6] used time-series models to represent the dynamics of search behavior over time and the results showed that it could effectively improve ranking and query suggestions. These works, to some extent, improved search performance but did not solve the intrinsic problems like data sparsity and cold start.

Lots of research works focused on other factors to improve personalized information service. Social annotation [7] was explored by Lin as the expansion term resource, and used the term co-occurrence method to demonstrate that the expansion terms extracted from social annotation were better than those from feedback documents. Otsuka [8] used the seasonal and topical facets on the interfaces to provide appropriate terms in the systems of Community Question Answering. White [9] focused on all user's search histories, not only one user's. They analyzed a user's current search task, and mined other users' historical behavior who had performed similar tasks to leverage the current user's on-task search behavior. Lu [10] built a probabilistic model to identify implicit local intent queries, and leverage user's physical location to improve Web search results for these queries. Considered different people usually have different perceptions about the same document, for each document, Xu [11] got a personalized document profile for each individual user to better summarize his/her perception about this document, then constructed user profile as the sum of all of the user's personalized document profiles to better characterize a user's preferences. This method was proved effective to achieve better personalized ranking on the Social Web. Liu [12] proposed methods for analyzing and modeling user search behavior in search sessions, and generated prediction models of document usefulness from behavior data collected in a controlled lab experiment. And the documents predicted useful and not useful by the models were used to modify the queries in each search session. Their results showed that these models could lead to consistently improved performance.

In a personalized system, one of the most difficult works is to understand a user's need. Kotov [13] focused on cross-session search tasks, predicted if a user would return to the present task, and next time, identified queries from earlier sessions on the same task. By this method, a search engine could recommend queries to re-find helpful past results. Actually, there are lots of reliable and highly relevant personal documents precisely existing in a user's personal computer, which are very useful to predict a user's need. Indeed, lots of researchers had paid much attention on desktop information to raise the quality of information retrieval [6, 14] and personalization [6, 15]. Work task and interaction context also have been proven effective to predict information need. Work tasks and search tasks play different roles when a user interacted with an information system, and different work tasks would lead to different types of search tasks [16].

However, these works still ignore lots of precious information on a user's personal computer.

3 Desktop Context Model

Although there is plentiful precious personal information in a user's computer, in a particular time, the user generally works on a specific task, such as writing an academic paper, searching some specific information, shopping online or exploring some news. Under this special consideration, the system model was separated into two main parts, as shown in following figure.

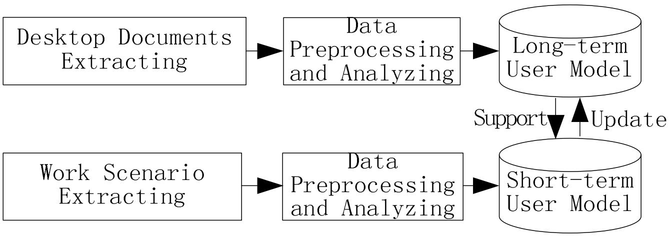

Fig. 1. Desktop Context Model

These two parts were dependent but connected. Nowadays, there are dramatically increasing volume of digital files in a user's computer, such as various documents, images, and multimedia files. These files were regarded as the desktop documents in this paper. For one part, desktop context extracting helped analyze the user's interest in the digital world, and helped establish a relatively complete profile of the user's personal information space. For another part, work scenario extracting focused on understanding what exactly a user was working on. This was very useful to predict the user's current need. Here the user's long-term model and the short-term model were mutually supportive to assure the best information retrieval service.

3.1 Analyzing Document Information

A user saved those desktop documents in his or her personal computer, so it is reasonable to assume that these documents were quite valuable for this user. However, in the long run, these documents did not own the same importance. To understand a user, a traditional way was to analyze the documents' content the user explored. Actually, there was some pretty important information hidden in the operational history, which, to some extent, reflected the real requirement degree for the document. Therefore, two kinds of information were employed to find those documents that owned highly values, as shown in figure 2.

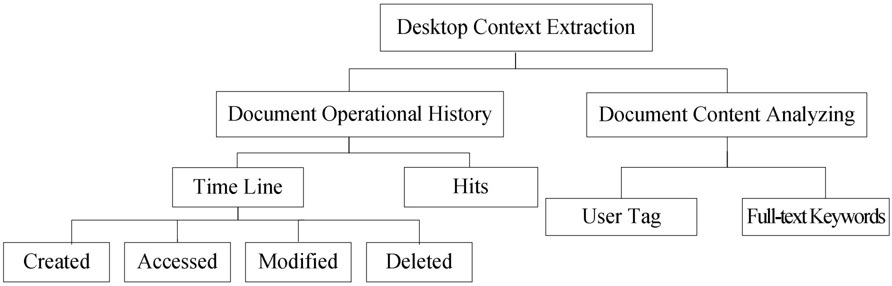

Fig. 2. Desktop Context Extraction

Extracting Document Operational History
According to Fig 2, this subsection discussed two kinds of operational information: time line and hit history.

- Time line

Considering that a user's interests might change over time, enlightened by Jaime Teevan's works [15], the time sensitivity was introduced. Time line included a document's created time, last accessed time, latest modified time, and deleted time. With these raw data, time sensitivity $timesst$ could be achieved by following formula.

$$timesst = del \cdot (\alpha \cdot crttime + \beta \cdot accsstime + \gamma \cdot mdftime) \tag{1}$$

where α, β and γ were parameters. $crttime$, $accsstime$ and $mdftime$ separately represented the time sensitivity of a document being created, viewed and modified. Time information was processed by fuzzy linguistic [14].

Besides, the deletion information was quite valuable. Once a document was deleted, it was sensible to believe that the user regarded this document as irrelevant or no longer useful. So the deletion time was also recorded here and it was denoted by del. But this record was only reserved for a period of time. Once the time exceeded a predefined threshold, then the document record would be deleted. Experiments showed that users rarely clicked those links that they had deleted. And excluding the corresponding item links from the returned results' list could effectively improve the user experience.

- Hit history

People own different behavioral habits like periodically cleanup disks or not. In a personal computer, there are various documents, fresh but useless or relatively old but frequently used. Thus, *hit* was introduced to record a document's hit history. And then combined with the time sensitivity, the importance of a document could be achieved by formula (2).

$$doc\ (timesst, hit) = A \cdot timesst + B \cdot hit + C \tag{2}$$

where *A*, *B*, *C* were constants that were achieved and adjusted in the experiments.

Extracting Document Content

The notion of relevance is the key of a personalized system. It is determined by the document content. We focused on user tags and keywords. User tags were generally marked by the authors, editors or peers. In our experiments, not all documents owned this kind of tags. And the keywords were extracted through full-text segmentation and preprocessing. In this paper, the document segmentation relied on the Institute of Computing Technology Chinese Lexical Analysis System1. The document weight was gained by TF-IDF, as shown in formula (3).

$$weight_i = \frac{n_{i,j}}{\sum n_{k,j}} \cdot \log\left(\frac{N}{n}\right) \tag{3}$$

where $weight_i$ represented the weight of word *i*; $N_{i,j}$ was the occurrence number of the word *i* in document *j*; the denominator was the occurrence number of all words in document; *N* was the total number of documents in the corpus, and *n* was the number of documents where the word *i* appeared.

Thus, the weight of word i could be achieved by formula (4).

$$word_i = doc_j \cdot \left(\sigma \cdot weight_i + \lambda \cdot tag_i\right) \tag{4}$$

where σ and λ were constants that were achieved and adjusted in experiments.

3.2 Analyzing Work Scenario

A user generally focuses on some particular task in a certain time. Consider an example scenario: while a user concentrated on topic A, like writing a paper, there might be in a particular time, the user switched to another topic B that might be generating a list to listen to the music. We called this task switch as the task context switch. No matter what the task was, there was a corresponding short-term context in the personal computer. We regarded current opened documents in the same work scenario as a

[1] http://www.nlp.org.cn/project/project.php?proj_id=6

work corpus, which would take more responsibilities than other documents in the personal computer to personalize information retrieval. Task identifying was not the research issue in this paper, so in our experiments, many work scenario were identified artificially.

Working corpus included local documents and online Web pages. It was more than a kind of real-time desktop resources analysis. Three kinds of information were employed to build a short-term user model:

1. Extracting the document content in the work corpus could display what kind of task the user was working on.
2. Analyzing the corresponding exploring time for each document might show how valuable the document was.
3. Several special actions, including print, collect, and save, implied that the user regarded these documents or pages as valuable.

No matter explicit or implicit, all these information, to some extent, reflected a user's interest preference. They could be used to build the short-term model.

4 Experimental Evaluation

There were 23 participants joined our experiments, who were asked to search pdf format documents separately through Bing, Baidu and the experimental system. The Discounted Cumulative Gain (DCG) [15] was introduced to measure the quality of the search engine result sets. DCG was a measure to evaluate the effectiveness of a search engine algorithm or related applications, which owned two assumptions:

1. Highly relevant documents were more useful when appearing earlier in a search engine result list (have higher ranks).
2. Highly relevant documents were more useful than marginally relevant documents, which are in turn more useful than irrelevant documents.

The DCG was defined as formula (5).

$$DCG(i) = \begin{cases} G(1) & ,if \quad i = 1 \\ DCG(i-1) + \dfrac{G(i)}{\log_2(i)} & ,otherwise \end{cases} \tag{5}$$

where $G(i)$ was the graded relevance of the result at position i. In order to determine whether a returned document was relevant to a user's search intention, each participant was asked to give a gain value to each returned document: 2 (if highly relevant), 1 (if relevant), or 0 (if not relevant).

For each query, DCG was cumulated for all ranks and offers us a simple method to measure the quality of a results set. However queries that had more relevant documents should have a higher DCG, so the DCG was normalized to a value between 0 (the worst possible DCG given the ratings) and 1 (the best possible DCG given the ratings).

The normalized DCG values could be averaged to measure the average performance of a search engine's ranking algorithm.

To evaluate the potential of desktop context, we compared the top-10 search results between the experimental system and two general search engines, *Bing* and *Baidu*, for the query "*Mercedes Benz*" for user A, who majored in automotive engineering and there were lots of documents about design, manufacture and operation of vehicle in his personal computer. According to the short-term model achieved from work scenario analysis, "*car design*" was taken into account. And the search results were filtered and re-ranked according to the long-term document model. The relevance of each result document was graded by the user, and the results are shown in following table.

The results in above Table 1 clearly showed that the experimental system used the keywords gained from work scenario analysis as the expansion query could provide more relevant items for the user than the other two general search engines and the sequence of gain value for each returned documents in refined results lists showed that those documents with more relevance were ranked better than its original lists.

Table 1. Comparison of top-10 search results between the experimental system and two general search engines

Bing Original Result	Gain	Baidu Original Result	Gain	Experimental Result	Gain
ec.europa.eu/competitior	0	worldauto.com.cn/html/sj	0	bric.com/archive/pdf/	2
mbusi.com/pages/MBUS	0	mcgraw-hill.cn/pdf/200708	0	bridgestone.eu/Englisl	2
mbusi.com/images/MB%	0	mercedes-benz.co.jp/e/nev	0	me.cycu.edu.tw/csme	2
daimler.com/Projects/c2	1	autechk.com/upfile/20092	1	abc-shanghai.com/en/	1
daimler.com/Projects/c2	1	total-lub.jp/lub/content/NT	1	cd1.edb.hkedcity.net/	1
ac-rerefined.com/Endors	0	businesscenter.jdpower.cc	0	abc-shanghai.com/en/	0
fire.state.mn.us/Respons	2	img1.toocle.com/uppic/a/	1	bharatbook.com/Sam	0
epa.gov/compliance/reso	0	westfalia-van.de/fileadmin/	0	web.mit.edu/aeroastro	2
iatfglobaloversight.org/d	0	english.people.com.cn/90	0	mercedes-benz.co.jp/	0
eaton.com/ecm/groups/p	0	mercedes-benz.com.tw/cc	0	mercedes-benz.co.jp/	0
DCG	1.64	DCG	1.29	DCG	6.86

To evaluate the performance of the system for the same query to different users with different background and information need, we randomly selected User B, who paid much attention to vehicle market information. According to the short-term model of User B, "*market report*" was taken into account when user B was searching. Compare the best ranking results of "*Mercedes Benz*" for user A and user B. Baidu served as the meta-search engine.

Table 2. Best ranking results and experimental results for two users for the same query

Best Ranking of the Result				Experimental Re-ranked Result			
User A	Gain	User B	Gain	User A	Gain	User B	Gain
ieeexplore.ieee.org/Xpl	2	mercedes-benz-financi	2	wzb.eu/gwd/wpa/pdf/blo	1	caranddriver.com/asset	1
mercedes-benz.com.au/	2	mitsubishifuso.lv/Proje	2	ieeexplore.ieee.org/Xpl	2	mercedes-benz-financi	2
oliverwyman.com/ow/pd	2	aldrichquaihoi.com/ass	2	autonews.com/assets/PI	1	osmanauction.com/Ne	1
mercedes-benz.com/flee	2	caranddriver.com/asset	1	mercedes-benz.com.au/	2	bishkek.usembassy.gov	0
mercedes-benz.com/flee	2	osmanauction.com/Nex	1	oliverwyman.com/ow/pc	2	nrel.gov/docs/fy09osti/	1
wzb.eu/gwd/wpa/pdf/blo	1	smmt.co.uk/downloads/	1	mercedes-benz.com/fle	2	mitsubishifuso.lv/Proje	2
autonews.com/assets/PD	1	nrel.gov/docs/fy09osti/	1	mercedes-benz.com/fle	2	aldrichquaihoi.com/ass	2
mercedes-benz-classic.c	0	ec.europa.eu/competiti	1	wcoty.com/files/2010_1	0	smmt.co.uk/downloads	1
oldtimer-doctor.com/lin	0	bishkek.usembassy.gov/	0	mercedes-benz-classic.c	0	ec.europa.eu/competiti	1
mercedes-benz-classic.c	0	us-cdn.creamermedia.c	0	oldtimer-doctor.com/lir	0	us-cdn.creamermedia.c	0
Normalized DCG	1	Normalized DCG	1	Normalized DCG	0.89	Normalized DCG	0.85

The best ranking of the results was the best possible ranking for a query a search engine could do for a user, and its normalized DCG was 1.00. The results showed that experimental refined list owned a pretty high normalized DCG, which meant the refined ranking was closed to the best ranking results. Take all 23 participants into consideration, we compared the variation of the average normalized DCG in three situations, including searched only with long-term model, only with short-term model, and combined those two models. The results showed in figure 3.

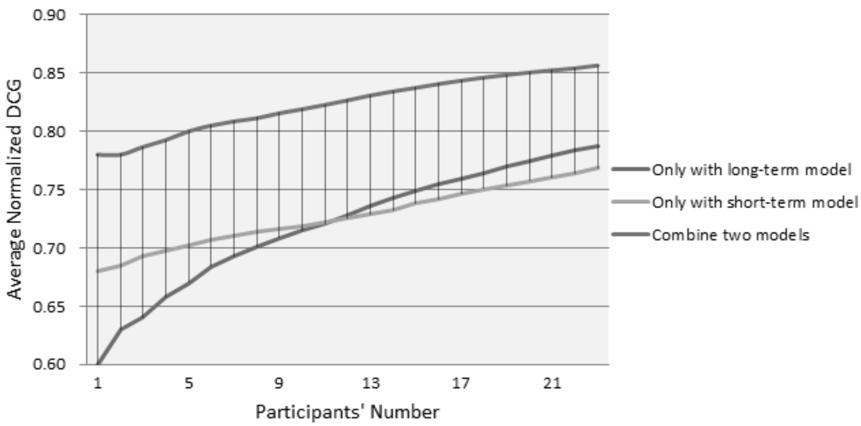

Fig. 3. Comparison of the average normalized DCG according to the participants' number

This figure showed that the short-term model served better for a specific search task than the long-term model. But in the long run, the long-term model showed more potential. These two models had their own merits and demerits. The combination of these two models could overcome the shortcomings of each models and support each other. The above figure displayed obvious advantage of this combination. The results also showed that with the increasing number of the participants, in all cases, the average value of the normalized DCG was gradually stabilized at a relatively high level, which indicated the potential of desktop context information that did help predict a user's current need and improve the personalized information retrieval.

5 Conclusion

This paper addressed the potential of desktop information, and proposed an approach to extract the desktop context to construct a relatively comprehensive personal information space. With the support of the long-term user model achieved by local resources extraction, we further concentrated on extracting the information of the work scenario that helped to understand what exactly a user need in a particular time. A series of experimental results reflected the potential of the desktop information that could improve traditional query expansion and effectively refine returned search results for the user. However, the present experiments were conducted in a limited field and we will continue to study to improve the information service in the future.

Acknowledgements. This work is supported by the National Natural Science Foundation of China (No.61402220), Hunan Provincial Natural Science Foundation of China (No.13JJ4076), the Scientific Research Fund of Hunan Provincial Education Department for excellent talents (No.13B101), Foundation of University of South China (No.2012XQD28), the Construct Program for the Key Discipline in University of South China (No.NHxk02), the Construct Program for Innovative Research Team in University of South China.

References

1. Zheng, N., Paloski, A., Wang, H.N.: An efficient user verification system via mouse movements. In: Proc. 18th ACM Conference on Computer and Communications Security, Chicago, Illinois, USA, October 17–21, 2011
2. Stern, D., Herbrich, R., Graepel, T.: Matchbox: large scale online bayesian recommendations. In: Proc. 18th WWW Conference, Madrid, Spain, pp. 111–120. April 20–24, 2009
3. Das, A.S., Datar, M., Garg, A., et al.: Google news personalization: scalable online collaborative filtering. In: Proc. 16th International Conference on World Wide Web, New York, NY, USA, pp. 271–280 (2007)
4. Lee, C.-J., Teevan, J., de la Chica, S.: Characterizing multi-click behavior and the risks and opportunities of changing results during use. In: Proc. 37th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2014), Gold Coast, Australia, July 2014
5. Radinsky, K., Svore, K., Dumais, S.T., Shokouhi, M., Teevan, J., Horvitz, E.: Behavioral Dynamics on the Web: Learning, Modeling and Prediction. Proc. ACM Transactions on Information Systems (TOIS) **31**(3) (2013)
6. Teevan, J., Dumais, S.T., Horvitz, E.: Potential for Personalization. Proc. ACM Transactions on Computer-Human Interaction (TOCHI) special issue on Data Mining for Understanding User Needs, March 2010
7. Lin, Y., Lin, H.F., Jin, S., et al.: Social annotation in query expansion a machine learning approach. In: Proc. the 34th Annual ACM SIGIR Conference, Beijing, China, pp. 405–414, July 2011

8. Otsuka, A., Seki, Y., Kando, N., et al.: QAque: faceted query expansion techniques for exploratory search using community QA resources. In: Proc. the 21st International Conference Companion on World Wide Web. Lyon, France, pp. 799–806, April 2012
9. White, R.W., Chu, W., Hassan, A., et al.: Enhancing Personalized Search by Mining and Modeling Task Behavior. In: WWW 2013, Rio de Janiero, Brazil, May 13–17, 2013
10. Lu, Y.M., Peng, F.C., Wei, X., et al.: Personalize web search results with user's location. In: Proc. 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2010, Geneva, Switzerland, pp. 763–764, July 19–23, 2010
11. Xu, Z., Lukasiewicz, T., Tifrea-Marciuska, O.: Improving personalized search on the social web based on similarities between users. In: Straccia, U., Calì, A. (eds.) SUM 2014. LNCS, vol. 8720, pp. 306–319. Springer, Heidelberg (2014)
12. Liu, C., Belkin, N.J., Cole, M.J.: Personalization of search results using interaction behaviors in search sessions. In: Proc. 35th International ACM SIGIR Conference on Research and Development In Information Retrieval. Portland Oregon, USA, pp. 205–214, August 12–16, 2012
13. Kotov, A., Bennett, P.N., White, R.W., et al.: Modeling and analysis of cross-session search tasks. In: Proc. 34th SIGIR. Beijing, China (2011)
14. Findlater, L., Moffatt, K., McGrenere, J.: Ephemeral adaptation: the use of gradual onset to improve menu selection performance. In: Proc. 27th International Conference on Human Factors in Computing Systems, Boston, MA, USA, pp. 1655–1664, April 4–9, 2009
15. Teevan, J., Dumais, S.T., Horvitz, E.: Personalizing search via automated analysis of interests and activities. In: Proc. 28th Annual ACM Conference on Research and Development in Information Retrieval (SIGIR 2005), Salvador, Brazil, August 2005
16. Li, Y.L., Belkin, N.J.: An exploration of the relationships between work task and interactive information search behavior. Journal of the American Society for Information Science and Technology (ASIS & T) **61**(9), 1771–1789 (2010)