A Location-Based Personalized Focused Search System for Physically Challenged People

Xin Shen[1, a], Wei Wang[1, b *], Junyi Huang[1, c], Yue Wu[1, d]

[1]Zhejiang Provincial Key Laboratory of Service Robot, College of Computer Science, Zhejiang University, Hangzhou 310027, China

[a]sxstar@zju.edu.cn, [b]wangwei_eagle@zju.edu.cn, [c]jy03163938@zju.edu.cn, [d]wy1988@zju.edu.cn

Keywords: Focused Search System; Location-Based Service; PersonalizedRetrieval

Abstract. Physically Challenged people are a special part of society, which need more care. Because of the huge differences in physical conditions, qualities of life, locations and personal backgrounds, there exists a lot of challenges for disabled people searching related information and receive professional and personal guide. This paper proposes a novel focused search system for physically challenged people, which uses the location information and provides the more effective search results. The search system consists of two function models, one is a focused search engine, and the other is a map of the related institutions. The search engine is designed for the multimedia files of disabled recovery, including words, pictures, audios and videos. The map of related institutions offers the location-based service showing the near occasion addresses intuitively, where the physically challenged people may get professional guide. With this location-based focused search system, the special community has an easy access to the technical database and relevant information. The system is highly significant in the rehabilitation information acquisition field.

Introduction

The basicallyprinciple of focused search or topic search was proposed by Menczer [1] [2] and Chakrabarti [3]. After that, Diligenti introduced the context graphs to the focused crawling [4]. Recently, De Groc [5] has also proposed the improved algorithm about the focused searchtechnology since 2011, which can be applied to the big data and become more intelligent.

As we know, there are no contributions to applying focused search to the physically challenged people in the world now, who have the deficiency ability to fetch information. In this paper, we will introduce and implement a novel location-based personalized focused search systemfor physically challenged people. Using this system, the disabled people can get the useful, meaningful and professional information with easy access.

Personalized focused search technology

Because of the variable types of the disabled people, it is quite meaningful to do the personalized searching based on the users' hobbies. Naive Bayes is a useful mothed to train the classification models. Focused on the document classification model, in our opinions, users' hobbies and representations of documents are consistent. Users' hobbies are vectors of the conditional probabilities, which represent the probability distributions of the classification model. Search engine will reorder the search results based on the probabilities of recommendation.

Considering the different kinds of physically challenged people, the personalized focused search engine uses the technologies of the contend-based filtering and collaborative filtering to select and reorder the searching result. First of all, we make the basic filtering rules for different types of the disabled people, and the details are following.

(1) If a user belongs to retarded people, we will erase the plain text information.
(2) If a user belongs to deaf people, we will erase the audio information.
(3) If a user belongs to blind people, we will erase the image information.
(4) If a user doesn't belong to the listed types above, no information filters.

In general, the framework of personalized focused search engine is shown as Fig.1.
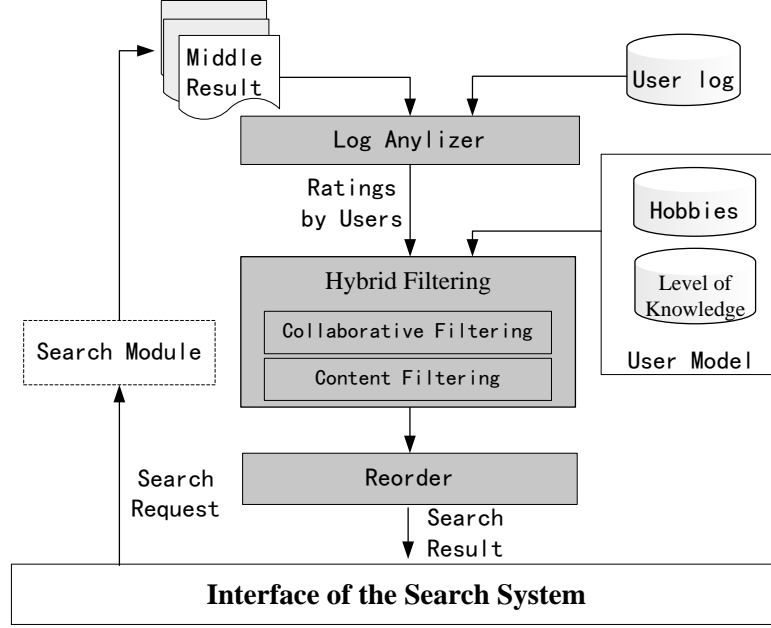

Fig.1. Framework of personalized focused search engine

We partition the users into different virtual communities based on their hobbies. Users in the same community share the similar hobby and level of knowledge.

Users submit the inquiry through the interface of the search system. The search model looks for the middle results from the disabled resource library. The log analyzer calculate the ratings to the resources based on the analysis of users history behaviors, such as the click times on the resources, visiting time, and bookmark. Hybrid filtering module finds the users' neighbors corresponding to the users' similarity. Using the collaborative filtering algorithm, hybrid filtering module may get the predict ratings. With these ratings, reorder module will adjust the searching results, and return the final result back to the users through the interface of the search system.

The algorithm routines are following:

Step 1:

Make a matrix of users' ratings, shown as Fig.2. The rows represent the users and the columns represent the resources. The goal of personalized recommendation is predict the rating to the unused resources.

	r_1	r_2	...	r_{n-1}	r_n
u_1		4	....	3	
u_2			...	5	
...	...	...	...	...	...
u_{m-1}	1		...		4
u_m		3	...	5	

Fig.2. Framework of personalized focused search engine

Step 2:

Get the users' similarities. The system uses the joint features and Pearson correlation to calculate the similarities [6]. Joint features are consisting of users' information, resources' information, and user-models information.

$$K\big((u,r,m),(u',r',m)\big) = K_U(u,u')K_R(r,r')K_M(m,m') \tag{1}$$

U, R, M mean the set of users, resources and user-models.

K_U, K_R, K_M come from the Pearson correlation:

$$K_U(u,u') = \frac{1}{|R(u,u')|} \sum_{i \in R(u,u')} r(u,i) \cdot r(u',i) \qquad (2)$$

$$K_R(r,r') = \frac{1}{|U(r,r')|} \sum_{i \in U(r,r')} r(i,r) \cdot r(i,r') \qquad (3)$$

$$K_M(m,m') = \frac{1}{|M(u,u')|} \sum_{i \in M(u,u')} r(u,i) \cdot r(u',i) \qquad (4)$$

Step 3:

Select the neighbors. Based on the users' similarities, the system chooses the k most similar users as the current user's neighbors.

Step 4:

Predict the ratings using the following formula.

$$p_{u,r} = \overline{r_u} + \frac{\sum_{i=1}^{k}(r_{i,r} - \overline{r_i}) * K_{u,i}}{\sum_{i=1}^{n} K_{u,i}} \qquad (5)$$

Step 5:

Reorder module will adjust the search results based on the personal hobbies.

Focused search engine

There are four modules of the focused search engine: Nutch focused crawler, Solr index service, back-end of search engine, and front-end webpage. The architecture of focused search engine is shown following Fig.3:
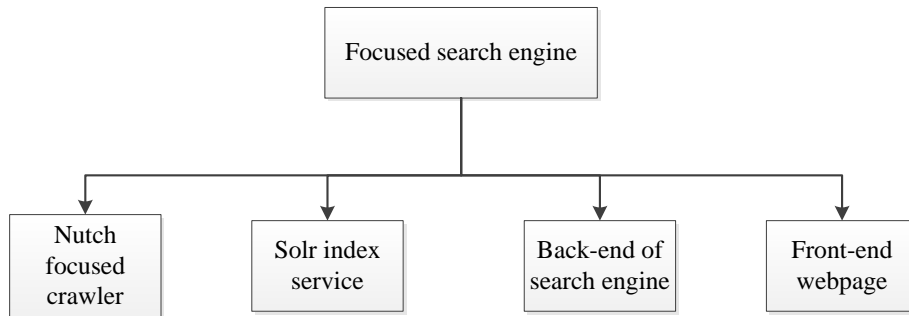

Fig.3. Architecture of focused search engine

Nutch is an open source search engine implemented in Java language. The module restricts the crawler behaviors. In order to remove the irrelevant information, the downloaded domain names are restricted in the white list.

After fetching the websites, the index is needed to be created. Our system adopts the Solr to create the index. Solr is short for Apache Solr, am open source search server, which developed in Java language and implemented by HTTP and Apache Lucene. The principle of Solr is using XML to add the documents to the searching set through the HTTP, and response the XML/JSON query to search the set through the HTTP too.

Map of the related institutions

To make the convenience of the physically challenged people, we introduce the map of the related institutions into the search system. There is more than 7000 geographic locations information of the rehabilitation for the disabled people.

Firstly, we will get the geographic locations information of the rehabilitation for the disabled people automatically. In the original database, there is nearly 9000 rehabilitations' information. Via

the API of Baidu map, we remove the error data and get the batch of addresses. Finally, about 7000 useful rehabilitation information left. With the subsystem, users can acquire location-based service based on the map of the related institutions.

The whole work flow chart of map of the related institutions is as follows:
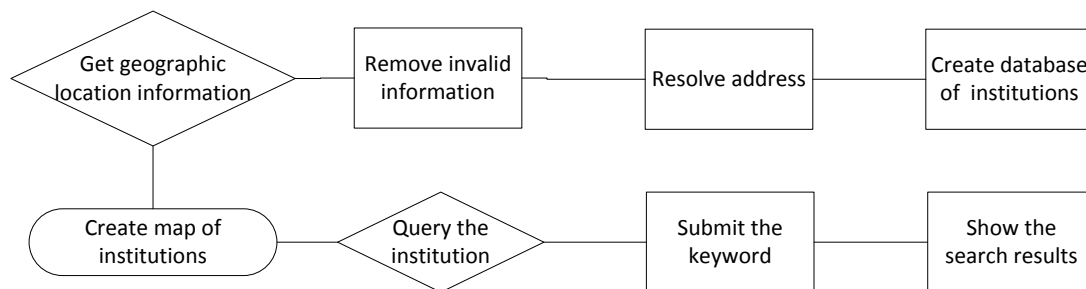

Fig.4. Work flow chart of map of the related institutions

Conclusion

Based on the personalized focused search technology, physically challenged people are able to get their professional information with easy access. Invalid information has been removed. That characteristic is the huge advantage comparing to other general search engines for disabled people. Owing tothe location-based service of the rehabilitationinstitutions, the system is friendlier to our target audience. Every useris able to get personalized search results in our system, which provides a good experience to these special crowds.

Acknowledgement

References

[1] Menczer F. ARACHNID: Adaptive retrieval agents choosing heuristic neighborhoods for information discovery[C]//MACHINE LEARNING-INTERNATIONAL WORKSHOP THEN CONFERENCE-. MORGAN KAUFMANN PUBLISHERS, INC., 1997: 227-235.

[2]Menczer F, Belew R K. Adaptive information agents in distributed textual environments[C]//Proceedings of the second international conference on Autonomous agents. ACM, 1998: 157-164.

[3] Chakrabarti S, Van den Berg M, Dom B. Focused crawling: a new approach to topic-specific Web resource discovery[J]. Computer Networks, 1999, 31(11): 1623-1640.

[4] Diligenti M, Coetzee F, Lawrence S, et al. Focused Crawling Using Context Graphs[C]//VLDB. 2000: 527-534.

[5] De Groc C. Babouk: Focused web crawling for corpus compilation and automatic terminology extraction[C]//Web Intelligence and Intelligent Agent Technology (WI-IAT), 2011 IEEE/WIC/ACM International Conference on. IEEE, 2011, 1: 497-498.

[6]Sarwar B, Karypis G, Konstan J, et al. Analysis of recommendation algorithms for e-commerce[C]//Proceedings of the 2nd ACM conference on Electronic commerce. ACM, 2000: 158-167.