

Analysis of character network heat based on web retrieval technology

Minglei Wu^{1,a} and Jingchang Pan^{2,b,*}

¹School of Mechanical, Electrical & Information Engineering, Shan Dong University, Weihai; School of Rong Cheng, Harbin University of Science and Technology, Wei Hai, China

²School of Mechanical, Electrical & Information Engineering, Shan Dong University, Wei Hai, China
^awuming8511@126.com; ^bpjc@sdu.edu.cn

Abstract. Information technology is now developing rapidly, the Internet has also obtained widespread popularization. The amount of information on the network is increasing exponentially, whose information sources are widely distributed and varied. If the information can't be managed in an orderly manner, it will be difficult for the user to extract the information they need from such a massive amount of information. Although the current search engine give people a lot of convenience in searching for information, but the search engine can't reflect the user's personalized information demand with facing a wide variety of users with different information needs, knowledge background and interest. In this paper, a recommendation method based on search results with java as a technology is implemented. This method takes Baidu hot people ranking as an example for verifying.

1 Introduction

In the 1970s, research and application of personalized recommendation systems began to emerge. Grundy is universally recognized as the first personalized recommendation system that is put into use, and this system recommends relevant books for each user by creating a user interest model [1]. By the 1990s, the theoretical framework of the personalized recommendation system had been mature [2-4]. Tapestry mail processing system calculates the similarity between users for recommending [5]. GroupLens to establish user information group, the group of users can publish information, through the social information filtering system to calculate the similarity between users, and then to other users within the group to recommend [6]. Ringo uses the same social information filtering method to recommend music to users [7]. Other recommended systems include Amazon's book recommendation system, Jester's jokes recommendation system, Phoaks' page recommendation system [8-10]. Since the beginning of the 21st century, under the great data age, the individual demand for information has become more and more intense, and more and more personalized recommendation systems have been developed. The system of this paper mainly meet the user's personalized information needs through the second search analysis of the current search engine result.

*Jingchang Pan: pjc@sdu.edu.cn

2 Basic principles of search engine

Search engine is not a search for the Internet, it searches for pre-organized index database. Similarly, the search engine also can't understand the content on the page, it can only match the text on the page. Search engine workflow is shown in Fig. 1.

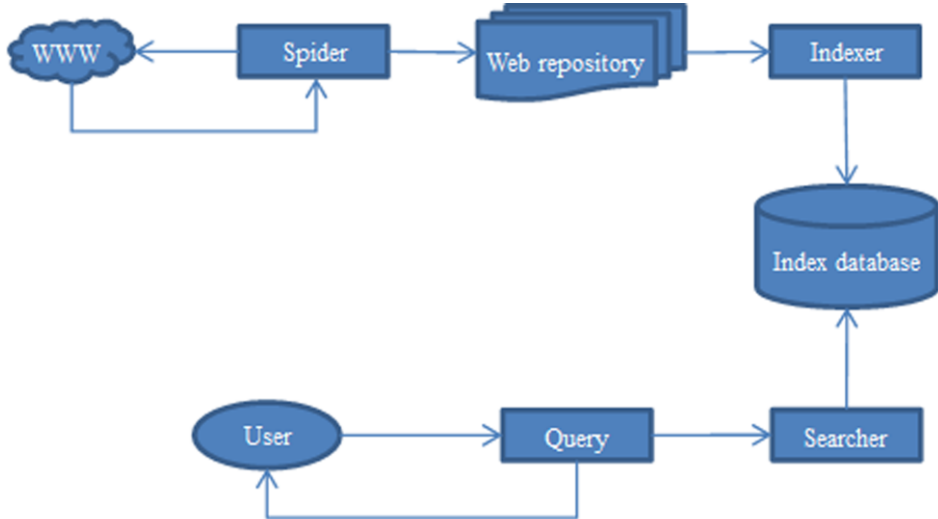


Fig. 1. Search Engine Workflow

Search engine workflow can be divided into four steps: Firstly, the web crawler crawls the page from the Internet according to the relevant algorithm (depth first, breadth first priority); secondly, the information on the page is processed after crawling the page; thirdly, after the post-processing information will be saved in the index database; fourthly, the search result is sorted by the search engine and displayed to the user when the user searches the index information about the query in the index database.

In 1990, Alan Emtage of the University of Montreal used Archie engine for related information retrieval in a specific network, thus creating a new field of modern search engines. At present, the Internet search engine can be divided into three types according to the information collection methods and services provided by different ways.

(1) Director Search Engineer. The search engine is build based on manual or semi-automatic way, and the Internet page is browsed by a lot of manpower which will be compiled into HTML files, classified and arranged by a certain order in order to enable users to refer look up the index.

(2) Crawler-Based Search Engineer. The data source is automatically searched and indexed in a certain strategy by a spider robot program.

(3) Meta Search Engineer. There is actually no search engine in itself, but it relies on other original engine index or search interface to complete the task, that is, it doesn't have its own data, but submits the query request to more than one search engine at the same time. Then these results with repeated exclusion, rearrangement and other processing as a result of their own return to the user.

3 Principles of the proposed method

Although the current search engine brings a lot of convenience for searching information resources, but search engines fail to meet the user's personalized information demand due to

a wide variety of users with different information needs, knowledge background and interest. The search engine can't analyze the user's search intent by interaction, the user's preferences and interests can't be tracked by capturing the user's records, and the user's feedback can't be used, etc. Furthermore, the engine returns too much useless information and too little personalized results, so it is difficult to find interesting documents in a simple way. Personalization recommendation is a valid method for lightening the user's burden on information retrieval.

This paper takes Baidu hot people sorting as an example, and this proposed method with JAVA technology based on the Baidu search result for specific keywords of user's interests. The concrete steps are as follows:

- (1) To enter the interested content in the txt file;
 - (2) To automatically transmit the interested content one by one to the Baidu search engine to search by this proposed method;
 - (3) To return the Baidu search result to the txt file one by one by this proposed method;
 - (4) To copy the result to excel for analysis. (Processing results shown in Fig. 3, 4).
- The method flow is shown in Fig. 2.

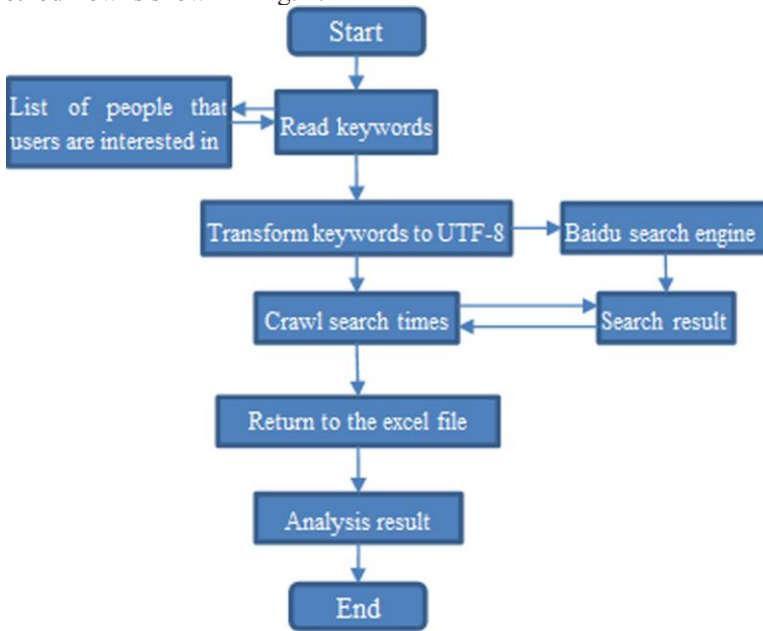


Fig. 2. The Proposed Method Flowchart

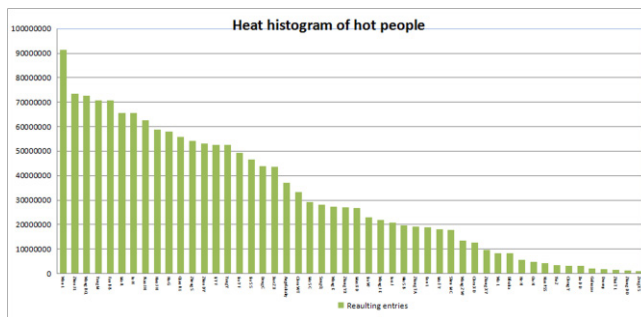


Fig. 3. Heat Histogram of Hot People

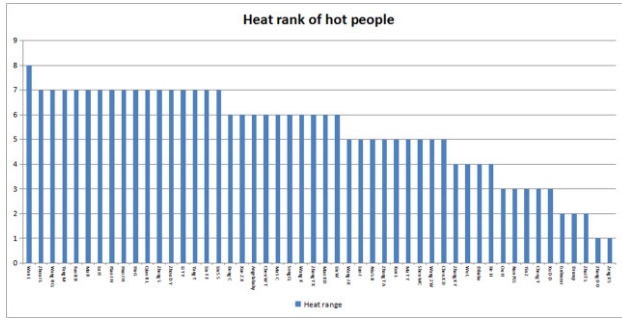


Fig. 4. Heat Rank of Hot People

4 Summary

In summary, this paper proposes a personalized retrieval method based on JAVA according to the personalized problem in Web retrieval, and this method can accurately grasp the corresponding Baidu resulting entry according to the user's interest. This method could not only be applied to the analysis of hot people, but also could be applied to such as Taobao, Jingdong and other large shopping sites to analyze and predict the behavior of consumers.

This research was financially supported by the National Natural Science Foundation of China (U1431102).

References

- [1] E. Rich, User Modeling via stereotypes, *Cognitive Science*. 4(3) 329-354.
- [2] P. Resnick, N. Iacovou, et al, GroupLens: An open architecture for collaborative filtering of netnews, in *Proc 1994 Computer Supported Cooperative Work Conf*, Chapel Hill, 1994, pp. 175-186.
- [3] W. Hill, L. Stead, et al, Recommending and evaluating choices in a virtual community of use, in *Proc. Conf Human Factors in Computing System*, Denver, 1995, pp. 194-201.
- [4] U. Mochio, *Webu shinkaron : hontō no daihenka wa korekara hajimaru*, Tokyo, 2006.
- [5] D. Goldberg, D. Nichols, et al, Using collaborative filtering to weave an information tapestry, *Comm ACM*, 35(12) 61-70.
- [6] J. A. Konstan, B. N. Miller, et al, GoupLens: Applying collaborative filtering to usenet news, *Comm ACM*, 40(3) 77-87.
- [7] U. Shardanand, P. Maes, Social information filtering: Algorithms for automating ‘Word of Mouth’, in *Proc Conf Human Factors in Computing Systems*, Denver, 1995, pp.201-217.
- [8] G. Linden, B. Smith, York J. Amazon.com recommendations: Item-to-item collaborative filtering, *IEEE Internet Computing*, 7(1) 76-80.
- [9] K. Goldberg, T. Roeder, et al, Eigentaste: A constant time collaborative filtering algorithm, *Information Retrieval J*, 4(2) 133-151.
- [10] L. Terveen, W. Hill, et al, PHOAKS: A system for sharing recommendations, *Comm ACM*, 40(3) 59-62.