IM Search: An Agent-based Personalized Metasearch Engine

Meijia Wang, Qingshan Li, Yishuai Lin
Software Engineering Institute, Xidian University, Xi'an, China
shuilingyi@126.com; qshli@mail.xidian.edu.cn; yslin@mail.xidian.edu.cn

Abstract*—Metasearch engine integrates search results from multiple underlying search engines, improving recall ratio in the big data environment. Multi-agent system is an important way to implement metasearch engine. Great progress has been made in this area, however the previous studies are still short of personalization level. To improve the precision ratio, this paper proposes a personalized metasearch engine which of Agent-based architecture. According to click-through data, the metasearch engine has the ability to schedule the appropriate search engines based on the expertness model, merge all of results into a single list by taking user interest into account, and provide personalized recommendation. Experimental results show that the proposed personalized metasearch engine performs better on precision. It is feasible to provide the required search results more effectively.

Keywords: metasearch engine; multi-agent system; personalized search

I. INTRODUCTION

With the widespread use of Internet, information manifests an explosive growth. A search engine is a tool that helps users to find useful information on the Internet. However, the number of web documents is daily increasing, consisting more than 10 billion web documents distributed on millions of servers [1]. An individual search engine only indexes a small coverage of web pages., Furthermore, the overlapping documents among different individual search engines is very low. Therefore, metasearch engine has been proposed to solve this problem. By combining multiple search engines, metasearch engine has the potential to extend the information retrieval coverage, providing convenience to users [2]. Although metasearch engine improves recall ratio in the big data environment, it is still limited by the precision ratio. The reason is that most of the existing metasearch engines are content-oriented, they use the similarity between the query and returned documents to search, ignoring the perspectives of users. To reach high precision, a personalized metasearch engine which takes user interest into consideration is needed. Generating schedule strategy and providing returned results according to user interest is very helpful to meet user's requirements.

An intelligent agent is an autonomous entity which observes and acts upon an environment, it is proactive and perceptible. A multi-agent system is a computerized system composed of multiple interacting intelligent agents within an environment [3]. Modeling metasearch engine based on multi-agent system has notable advantages. User interest changes overtime, metasearch engine which based on multi-agent system has the ability to perceive the change of user interest actively because of the characteristics of agent. It is helpful to analyze user intent, schedule underlying search engines and merge results more flexibly according to search context. Multi-agent system improves the adaptability of metasearch engine.

In this paper, we implement a metasearch engine based on multi-agent architecture to improve the personalization level, named "IM search". Agent is utilized to mine user interest from the click-through data, generate schedule strategy, merges results returned by underlying search engines and recommends results which user might be interested in. DCG@N [4] is used to evaluate the proposed metasearch engine.

II. PROPOSED ARCHITECTURE

The proposed multi-agent architecture is composed of seven kinds of agents. **Interface Agent** is responsible for interacting with users, including receiving queries from users and displaying the returned results to users. **Search Agent** is utilized to evaluate the expertness of underlying search engines and generate the schedule strategy, the offline search engines will be never invoked. It is also in charge of managing the SE Agent. **SE Agent** is responsible for communicating with underlying search engines. Particularly, each SE Agent corresponds to a search engine. The SE Agents detect the states of underlying search engines. If some search engines are offline, SE Agents perceive the offline state and send the signal to Search Agent. **UserInterest Agent** analyzes the click-through data and obtains user interest. **User Group Agent** clusters users into different groups according to the query data. **ResultRecommended Agent** provides recommended results for a specific query that user might be interested in. **ResultMerge Agent** merges all of returned results into a single list.

The multi-agent architecture works as shown in Figure1: a user submits a query to the InterfaceAgent. The Interface Agent sends the query to the Search Agent and ResultRecommended Agent. The Search Agent receives the query, generates the schedule strategy and sends the query to the SE Agents which will be scheduled. The SE Agents communicate with underlying search engines to complete the search task, and then pass on the returned results to the ResultMerge Agent. Meanwhile, the Interface Agent also passes on user information to the UserGroup Agent and UserInterest Agent. The

DOI: 10.18293/SEKE2018-082

UserGroup Agent gets the information of group to which the current user belongs based on the Query data, and sends it to the ResultRecommended Agent. After getting the query and user group information, the ResultRecommended Agent generates the recommended results and sends them to the InterfaceAgent. According to user information, the UserInterest Agent obtains user interest factor and passes it on to the ResultMerge Agent. Then the ResultMerge Agent merges all of results returned by SE Agents into a single list, and returns the list to the InterfaceAgent for displaying. The UerInterest Agent analyzes the query data to obtain user interest. After obtaining all of results, the Interface Agent displays results to users.

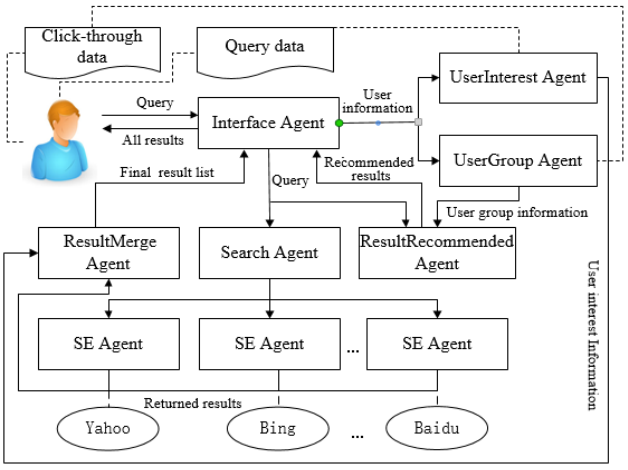

Figure 1. Working diagram of the proposed multi-agent architecture

III. User Interest Model

UserInterest Agent is designed to obtain user interest based on query data. Our previous work [5] describes how to obtain user interest model .

IV. Schedule Strategy

Search Agent is utilized to generate the schedule strategy. In order to evaluate the ability of underlying search engines, According to [6], the expertness model is constructed. When user submits a query, metasearch engine obtains the topic to which the query belongs, then based on the expertness model, the appropriate underlying search engines are selected to complete the search task.

V. Result Merging Method

ResultMerge Agent is utilized to merging the results returned by underlying search engines. When the underlying search engines returns results, the metasearch engine conducts word segmentation for each document and identifies the topic to which document belongs. The topic of user interest is matched with the topic of each returned document so that the user interest factor about the document can be obtained. The final score assigned to a document with user interest can be found in our previous work [5].

VI. Result Recommendation

Providing result recommendation will help users to find useful information with less effort. ResultRecommended Agent is responsible for recommending user wanted results. Both explicit information and implicit information are used to generate recommendation. This paper obtains explicit information based on the registration of a user, which is helpful to find the similar users in the system. The click-through data is the implicit information. From the click-through data, the relevant documents for a query will be found. The recommendation is generated among similar users. The details can be found in our previous work [7].

VII. Experimental Results

In this section, the performance of the proposed methods is discussed. “IM search” is a WWW metasearch engine which combines the search engines “Youdao”, “Baidu”, “Bing”, “Yahoo”, and “Sogou”. Two users with different interests are designed to log in IM Search, User1 and User2. For the same query “lincoln”, the results pages for User1 and User2 are shown in Figure 3 and Figure 4 respectively. It is obvious that these two users have different retrieval results. What’s more, IM Search recommends some results for User1. The recommended results are at the top of the result list.

Figure 3. The returned pages for User1

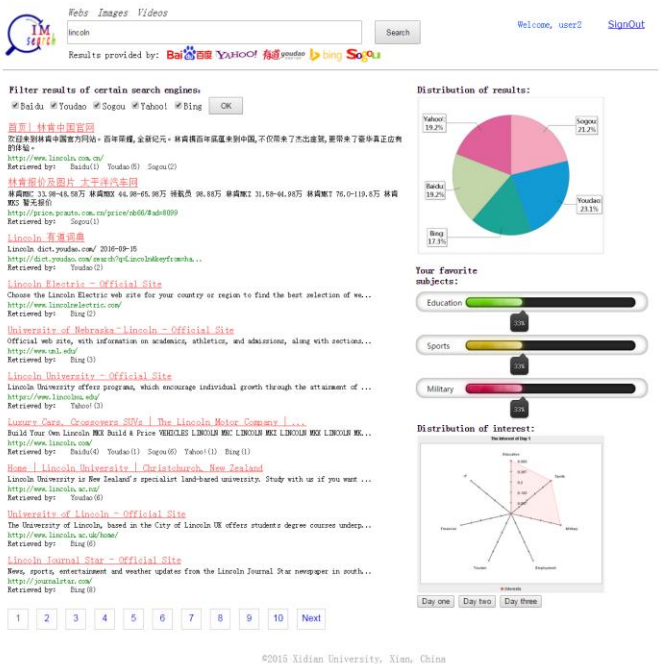

Figure 4. The returned pages for User2

DCG@N (Here, N=10) is utilized to evaluate the precision of the personalization mechanism. Five users are invited to use IM Search for a period of time. Then each user is asked for request 10 queries with login and logout status respectively, and the mean value of DCG@N for the 10 queries is calculated. The results are shown in TABLE. 1. We can see, Users get better results after logging in the system. Because the personalized mechanism only available with login status.

TABLE I. THE MEAN VALUES OF DCG@N

	user1	**user2**	**user3**	**user4**	**user5**
Login	18.12	17.25	17.93	18.57	16.35
Logout	17.72	16.75	16.31	16.52	16.10

VIII. CONCLUSIONS

This paper presents a personalized metasearch engine based on multi-agent architecture. By collecting user's click-through data, the metasearch engine has the ability to mining user interest, schedule the appropriate underlying search engines and obtain the personalized results. According to the group members' behaviors, it can also generate recommended results for users. Experimental Results show that the proposed metasearch engine performs better on precision. But there are still open issues ahead needed to address: 1).User interest is obtained based on the click-through data. But others user behaviors, such as the browsing time, download history are also significant for analyzing user interest. 2).Recommending personalized query words for different users is also necessary.

ACKNOWLEDGMENT

This work is supported by the Projects (JBZ171001) supported by the Fundamental Research Funds for the Central Universities of China, Projects (2017073CG/RC036 (XDKD004)) supported by the Technology Program of Xi'an, Projects (JB171004, BDY221411, K5051223008) supported by the Fundamental Research Funds for the Central Universities of China, Projects (61373045, 61672401) supported by the National Natural Science Foundation of China; Project (315***10101) supported by the Pre-Research Project of the "Thirteenth Five-Year-Plan" of China.

REFERENCES

[1] A. H. Keyhanipour , et al. "Aggregation of web search engines based on users' preferences in WebFusion." Knowledge-Based Systems. vol.20, no.4, pp.321-328, 2007

[2] W. Meng, C. Yu, and K. L. Liu. "Building efficient and effective metasearch engines." Acm Computing Surveys. vol.34, no.1, pp.48-89, 2001

[3] Y. Lin, P. Descamps, N. Gaud, et al. "Multi-Agent System for intelligent Scrum project management." Integrated Computer Aided Engineering, vol. 22, no. 3, pp.281-296, 2015.

[4] K. Zhou, et al. "Learning the Gain Values and Discount Factors of DCG." IEEE Transactions on Knowledge & Data Engineering ,2012.

[5] M. Wang, Q. Li, Y. Lin, et al. "A personalized result merging method for metasearch engine." The International Conferenceon on Software and Computer Applications. pp.203-207, 2017.

[6] Li Q, Wang J, Chu H, et al. "Personalization mechanism in an intelligent agent-based meta-search engine." Scientia Sinica, vol.45, no.5, pp.605-622, 2015

[7] Y. Li , Q. Li, Y. Lin. "A Personalized Result Recommendation Method based on Communities." International Conference on Data Mining, Communications and Information Technology. pp. 6-11, 2017.