iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 2 **Issue:** IV **Month of publication:** April 2014

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

Personalized Web Search

Miss. Ashna Sayyad[1], Miss. Akshada Gaikwad[2], Miss. Rupali Khule[3], Miss. Bhavana Deoker[4]

B.E, IT, Amrutvahini College Of Engineering, Sangamner, Maharashtra, India [1, 2, 3, 4]

Abstract: ***The amount of information on the World Wide Web is growing rapidly, and in this information explosion era the number of new users inexperienced in the art of web research. While searching what users lack is not scale of information but how to obtain information that meets user's search intention rapidly and accurately. The main problem current search engines are facing is to meet the user's need. In this paper, we combined two basic strategies (1) user interest and (2) collaborative filtering to reorder and to make the search results more personalized, we implemented the approach using multi-agents technology.***

Keywords: ***Personalized web search, Reordering, User interest, Collaborative filtering, Multi agents.***

1. INTRODUCTION:

The information floods everywhere on the World Wide Web nowadays. What users lack is not the scale of the information but the way to obtain the needed information rapidly and accurately. Search engine is one of the most used ways by the users to get vast amount information on web. But obviously, it has the problem that no matter how different the user's search intentions are the search results referred to the same query keeps the same ignoring the difference of users on their personalized features such as interests, hobbies, needs and intentions. For example, two users searching for the 'jaguar' may have different perspective however; search engines are not able to differentiate the meaning of the word in terms of the user's interest. The best way to solve the problem is to reorder the search results according to the users' personalization features. The basic idea behind personalized reordering is to discover individual characters of different users. The information a user truly needed typically contains the following features that are (1) It should be appropriate to the user's experience, interest and knowledge background and (2) It should be considered to be useful by most other users. For the first feature, we built a model called user interest model considering the user's experience, interest and knowledge background to provide search results based on individual needs. For the second one, we implemented collaborative filtering according to the other users' evaluation to the search results.

In this paper, we implemented user interest model based on user's interest and further interest value is calculated for each search result items. After this, we calculated the recommendation value in collaborative filtering based on other users' evaluation of the search result items. Finally, we reordered the search results combining strategies, the interest value and recommendation value. We implemented this personalized web search using multi-agents. The second section introduced the related work of personalized web search. The third section described our approach and the process to implement the method using multi-agent. Some probable results are shown section 4. Section 5 is about conclusion and future work. The final part is acknowledgements and references.

ISSN: 2321-9653

2. RELATED WORK

Lots of research has been done on personalized web search. A common way is to achieve personalized web search by content analysis including query expansion[1], web content and user profile analysis[3], user behavior analysis[4] and result processing[2], etc.

Also the research has been done on analyzing the hyper link structure of the web [5]. Another common way to achieve personalized web search is based on user interest and collaborative filtering.

There are various personalized systems. Based on recommendation technology, they can be assigned as systems based on information filtering and collaborative filtering and rules. Systems based on rules are simple and direct. But it's hard to keep the rules appropriate and it will be harder to manage the system as the rules increase. System based on content filtering can reflect users' individual information well.

The static keywords based interest model build the users' interest models with interest keywords that the users feedback, and it is relatively simple (i.e. iGoogle) [6]. The shortcomings are that the description of user interest is based on the user's feedback information and it cannot recognize the interest points and changing processes. The keyword space vector model is a kind of dynamic interest models (i.e. Amalthaea System) [7]. Its shortcoming is that the accuracy of this model is affected by the ambiguity of the keywords seriously. Another dynamic interest model is based on semantic web (i.e. InfoWeb) [8] which can be regarded as an improvement of weighted keyword vector model, and it's still not accurate enough. Alexander Pretschner from German and Susan Gauch from American build a weighted concept hierarchy model of user interest [9] based on individual user's browse history on web, which is a concept based model and improved the personalized web search system's performance. However it needs to build a hierarchy concept structure in advance and its process of computation is complicated. Another dynamic interest model is the interest model based on probability [10]

In this paper, we will combine both the content filtering that is based on user interest and the collaborative filtering to share neighboring users' profiles in our approach. Specifically, content filtering will be achieved by user interest model.

3. OUR APPROACH TO PERSONALIZED WEB SEARCH

A. Architecture:

Below is the framework of personalized web search approach as figure 1. Users send queries at the search entrance of the tool. First, we get the search results referring to the user's query without any filtration. Then we calculate the interest value of each searching item based on the user's interest model and the recommendation value of each item based on collaborative filtering. Then we combine these two values according to some strategies and as the second step we reorder the search results. Furthermore, the reordered search results are presented to the requested user and user's clicking actions would be gathered as logs. Finally, the user's logs are analyzed to update the interest model and collaborative filtering mode l [11].

ISSN: 2321-9653

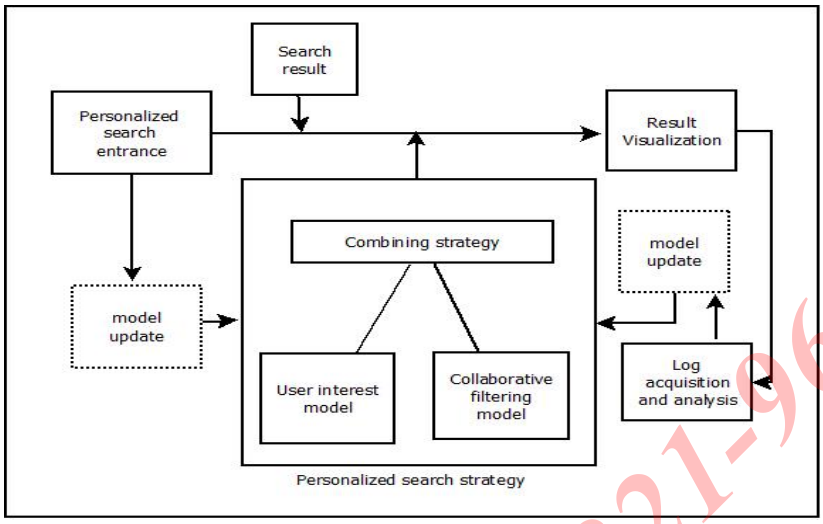


Figure1. Framework of personalized web search

B. *User Interest Model:*

1) Description of User Interest Model:

There are typically three kinds of user interest model that are, implicit dynamic model, explicit dynamic model and static model. User interest model contains formalized description of user's interest information. We adopted one of the explicit dynamic model that is weighted keyword vector model, in our paper.

The weighted keyword vector model consists of following formula:

$Interest_i = \{(k_1,w_1),\ldots\ldots,(k_n,w_n)\}$ (1)

Where, $Interest_i$ is the interest model of user u_i, k_i is the i-th keyword which can be either extracted from the user's logs, queries and typed in by the user in advance and w_i is the weight of keyword k_i which shows how interested the user is in k_i. The weight w_i can also be called as interest value .

2) *Update User Interest Model:*

In this paper, we will update the user interest model dynamically. When user u_i send a query k_i, it will first find out whether the keyword k_i is in user's interest model. If the item (ki, w_i) is in $Interest_i$, a unit score is added to w_i Otherwise a new item (k_i, w_i) will be added into $Interest_i$ where w_i is the default value.

For any user u_i, the interest value w_i of $Interest_i$ will decrease according to the Ebbinghaus Curve.

Assume that w_{i_pre} is the interest value before decrease and w_{i_new} is the interest value after decrease.

$$W_{i\_new} = w_{i\_prev} \times \lambda \quad (2)$$

$$\lambda = e^{log_2(t-t')/30} \quad (3)$$

Where λ is the attenuate coefficient, t is the current time and t' is the time when interest value was lastly updated. The item (k_i, w_i) will be removed from $Interest_i$ if w_{i_new} is less than the threshold.

3) Compute the User Interest Value:

According to the user's query, our tool will get back some search results. Then for each result r_i, we will compute the user's interest value to it. The algorithm is shown below.

ISSN: 2321-9653

Input: u_k*'s interest model* $Interest_k$ *and result* r_j
Output: Interest value I_{kj} *of* u_i *to* r_j
Process: $I_{kj} = 0$;
For each $(k_i, w_i) \in Interest_k$
if r_j *contains* k_i
$I_{kj} = I_{kj} + w_i$
return I_{kj}

C. *Collaborative Filtering Model*

Collaborative filtering system worked by collecting the human judgments (known as rating) for items in a given domain and matching together people who share same information needs or same tastes. User of collaborative filtering system shares their analytical statements.

Assume U is the set of all the users. For each $u_i \in U$, R_i is the set of the resources (search results) that u_i had clicked. So we obtained the matrix M as figure 2.

items / users	r_1	….	r_n
u_1	M_{11}	….	M_{1n}
….	….	M_{ij}	….
um	Mm1	….	Mmn

Figure 2. Recommendation matrix

In the matrix, $u_i \in U$ and $r_i \in R$ where M_{ij} is the recommendation value of user U_i to item r_j. In this paper we take the count of clicks as the recommendation value. If user U_i has never clicked item r_j, M_{ij} will be set as zero. Then the similarity of two users is computed based on the cosine similarity:

$$Sin\left(\vec{ui},\vec{uj}\right) = cos\left(\vec{ui},\vec{uj}\right) = \frac{\vec{ui}\cdot\vec{uj}}{|\vec{ui}| \times |\vec{ui}|} \quad (4)$$

Where, u_i is the vector (M_{i1}, M_{i2}, ... , M_{in}) and u_j is the vector (M_{j1}, M_{j2}, ... , Mjn).

For each user u_i, we compute the set of its nearest neighbors as C_i and its recommendation value R_{ij} to item r_j based on C_i. Then we record the value R_{ij} in the i-th row and j-th column of matrix M_{out}.

$$Rij = \begin{cases} Mij \text{ if } Mij \neq 0 \\ Pij \text{ if } Mij = 0 \end{cases} \quad (5)$$

Pij=

$$\frac{\sum_{c \in c1} \sin(ui,c) \times Mcj}{\sum_{c \in c1} |\sin(ui,c)|} \quad (6)$$

When user u sends a search query, we first get the search results. And for each result, we check the corresponding recommendation value in the matrix M_{out}. If r_j is not contained in matrix Mout its recommendation value will be set as zero. Then we update the matrix M when user u has clicked any search results and update the matrix Mout in an iteration way.

D. *Personalized Web Search Strategy*

After the previous steps, for each search result r_j, we have obtained the interest value I_{ij} and recommendation value R_{ij} of user u_i to r_j. Then we compute the corresponding reordering value $Score_{ij}$ as following.

$$Score_{ij} = \alpha \times I_{ij} + (1 - \alpha) \times R_{ij} \quad (7)$$

Finally, we reorder all the search results according to the descending order of $Score_{ij}$ and show user u_i the final results. The α in formula (7) is computed in an iteration way. By analyzing of the user logs in period T, we found the best α value (α_T) of T. And in period T+1, we use α_T to provide services.

ISSN: 2321-9653

E. Design and implementation of the Approach Based on Multi-Agent:

In this paper, we used a Agent System which contain 5 agents, these are user interaction

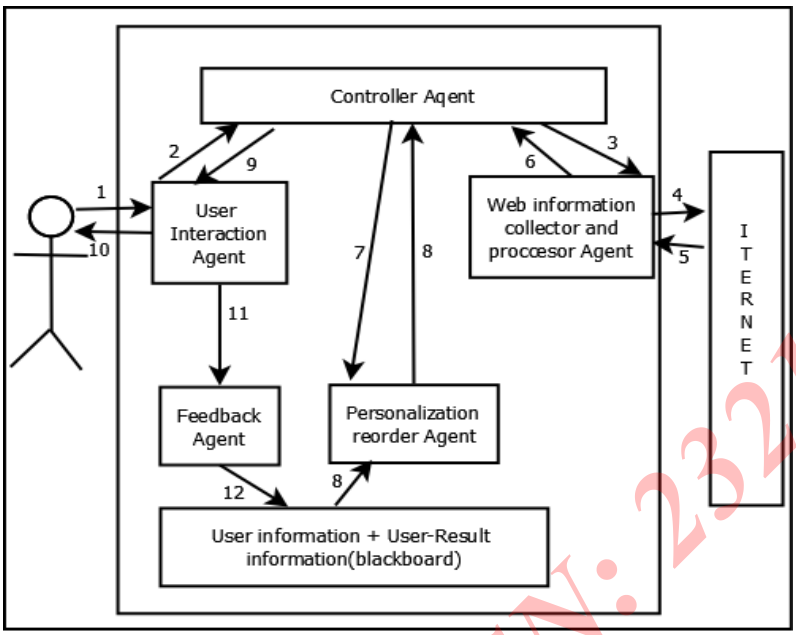


Figure 3. Multi-Agent Technique

agent, controller agent, web information collector and processor agent, personalization reordering agent and feedback agent. This system used distributed planning based centralized scheduling.

The interaction with user and visualization is performed by the user interaction agent. Together user's queries and show the personalized search results after reordering is a main task. Each task is completed by the corresponding agent. But the controller agent is very important agent allocation and scheduling of tasks is competed by it. The web information collector and processor agent collect the web information according to users' queries and perform the process on it. It is connected to the network during this process. Personalization reorder agent is reorders the search results based on the user's individual information and achieve the personalization. Feedback agent collects the users' actions on the search result page and then adjusts it to improve the personalization service. Each agent is unique and they can work independently .The cooperation and communication of all the agents is a task of controller agent.

The figure 3 shows the process of personalized web search and communication among agents. The user interest model and user-resource information are used by more than one agent. The communication among agents is necessary. In this, we used the "blackboard" method to implement the communication between the agents.

4. RESULT IMPLEMENTATION

For some queries, our approach can obtain better reordering effect than the original order provided by Google.

In the experiment, the users searched many queries. The queries with high search frequency are

ISSN: 2321-9653

Query	Previous Result	Our Result	Improvement
Topic-map	0.501	0.730	60.10%
Memory	0.713	0.533	-17.20%
Dijkstra	0.742	0.722	6.30%
Visualization	0.724	0.451	-30.46%
Neural network	0.940	0.800	0.60%

Table 1. Experimental results

In the experiment, the users searched many queries. The queries with high search frequency are topic-map, memory, Dijkstra, neural lnetwork, visualization. The following table shows experimental results for above queries.

Finally, from the topic-map row in Table 1 we can conclude that, comparing with the order given out by Google API (Previous result) our approach can do some improving by doing reordering based on the values given by our approach.

4) CONCLUSION AND FUTURE SCOPE

We can conclude from the experimental result that our approach can obtain a good result when the query is ambiguous or can be divided into some categories or the search results contain much spam information

When the query is explicit and the search result is precise this approach doesn't works well, so there is need to add various hybrid approaches in order to remove the drawbacks. On the whole, our approach can do some contribution in personalized web search comparing with the order of the search results given out by Google API.

The interest model based on keyword vector is simple and convenient but has some disadvantages because of semantic ambiguity etc. In the future work, we can use the semantic network to improve our interest model. And the collaborative filtering model also has some problems like sparseness, cold boot and difficulties in expansibility which should be solved in further work. In future work, we need to find a method to update database in real time. Moreover, it is possible that, after a long-time using the content of our models can become more rich and the effect of our approach may be improving, but it will still need experimental verifications.

REFERENCES

[1]. P.Chirita, Claudiu S.Firan, W. Nejdl. Personalized Query Expansion for the Web. SIGIR, 2007.

[2]. Oyama S. , Kokubo T. et al. Domain-specific Web search with keyword spices. IEEE Trans. Knowl. Data Engin. 16, I, 17- 27. 2004.

[3]. Xiaogang Han, Zhiqi Shen, et al. Folksonomy-Based Ontological User Interest Profile Modeling and Its Application in Personalized Search [A]. (Eds.): AMT 2010, LNCS 6335, pp. 34-46, 20 I O.

[4]. Cheqian Chen, Kequan Lin, et al. Personalized Search Based on Learning User Click History [Z]. Proc. 9th IEEE Tnt. Conf. on Cognitive lnformatics,20 I O.

[5]. T. Sarlos, et al. To Randomize or Not to Randomize: Space Optimal Summaries for Hyperlink Analysis[Z]. Proc. 15th Int'l World Wide Web Conf. pp. 297-306, 2006.

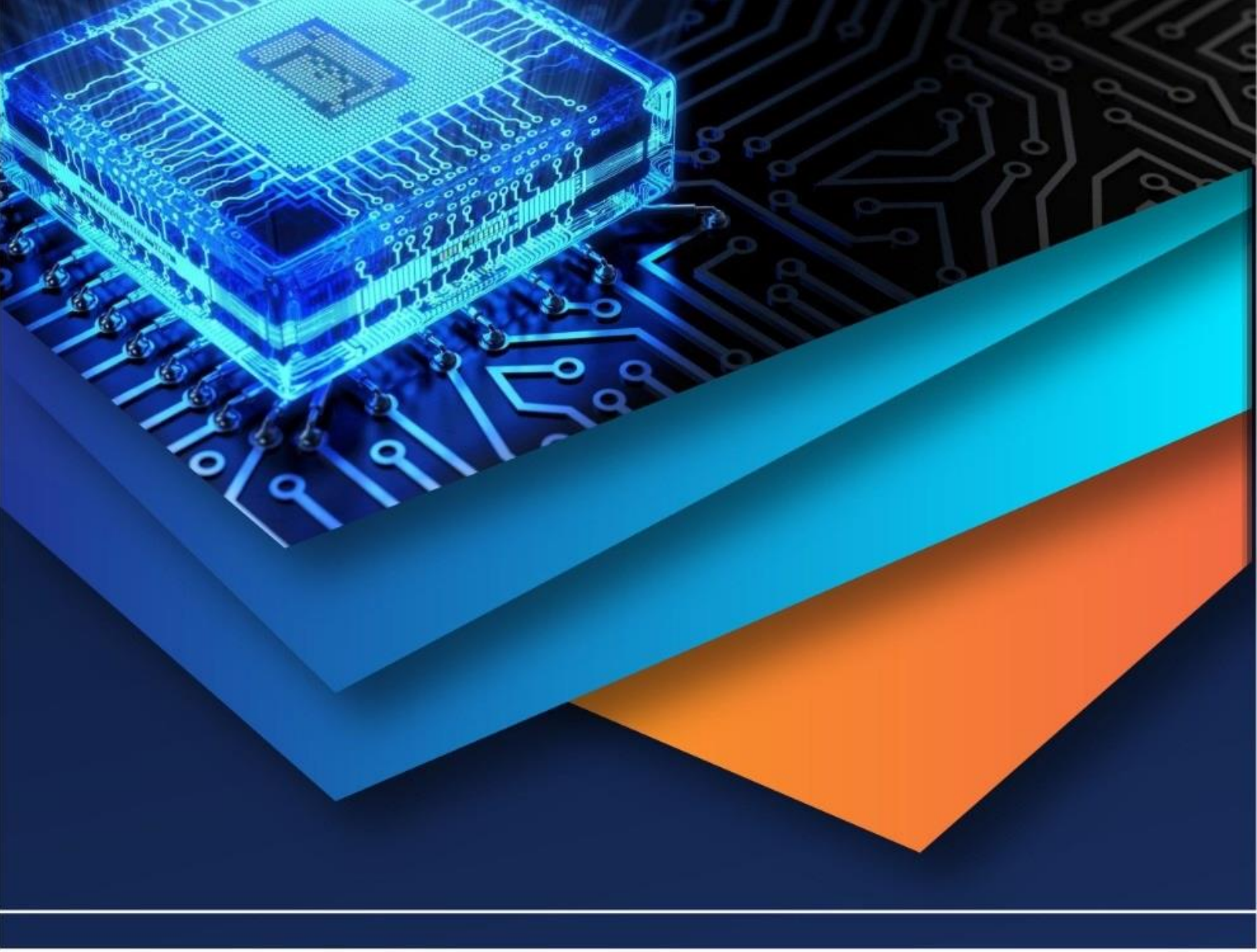
[6]. http://www.google.com. hklig?hl=zh-CN.

[7]. Gauch S, Speretta M, Aravind C, et al. User profiles for personalized information access[J]. The Adaptive Web Methods and Strategies of Web Personalization, 2007: 54-89lxii+ 762.

[8]. Gentili G, Micarelli A, Sciarrone F. InfoWeb: An adaptive information filtering system for the cultural heritage domain. Applied Artificial Intelligence, 2003, 17 (8-9): 715-744.

[9]. Alexander Pretschner,Susan Gauch.Ontology Based Personalized Search[C]. Proc.llth IEEE Intl.Conf.on Tools with Artificiallntelligence. Chicago.November,1999.

[10]. Hao Wang. Research on Change Pattern of Interest in Learning for e-Learner[D].Xi'an:Xi'an Jiaotong University. 2011.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY