Personalized Retrieval of Spatial Information Combining User Profile with Query Request

Yu Xia[1*], Jun Gong[1], Xinyan Zhu[2]
[1]School of Geography and Environment, Jiangxi Normal University, Nanchang, China
[2]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China
[*]Corresponding author, e-mail:geoxy@126.com

***Abstract*—Conventional information retrieval doesn't meet users' personalized requirements. Furthermore, personalized retrieval in recent few years focuses on mainly personalization of web information retrieval regardless of the difference between traditional information and spatial information. In this paper, a personalized retrieval strategy of spatial information is further put forward, which combines user profile with query request, and consists of following stages: Firstly, original data collected by system is cataloged into different levels, and useful information is refined from the levels; Secondly, user profiles are constructed on the basis of refined information; Thirdly, the search for spatial information from query conditions is implemented and the candidates are got accordingly; Lastly, utility degrees of the candidates aren't identical for different users in spite of the same query conditions, and so the utility degrees are achieved based on user profiles and personalized retrieval results are available after descending the candidates by utility degrees.**

Keywords-personalized retrieval; spatial information; user profile;

I. INTRODUCTION

Information Retrieval usually refers to the process of getting specific information from a certain amount of information database according to the needs of users which is a cognitive process essentially [1]. The classic definition of information retrieval was given in 1960s by the father of modern information retrieval, G. Salton: Information retrieval is the area for research on the structure of information, organization, storage, search and retrieval [2]. As Fig. 1 shows, the principle of information retrieval is to select the results that matched requirement expression and feature database. An information retrieval system can be seen as quad formally R = (S, F, Q, ρ), where $S =(S_1 ,S_2 ,\cdots, S_s)$, $F = (F_1 , F_2 , \cdots, F_t)$, $Q = (Q_1 , Q_2 , \cdots,Q_n)$, ρ: $Q \times S \rightarrow R$, S indicate the information sets of information source, F are feature sets of information source, Q are the sets of requirement expressions, ρ is the function for matching requirement expressions and information, and R are retrieval results.

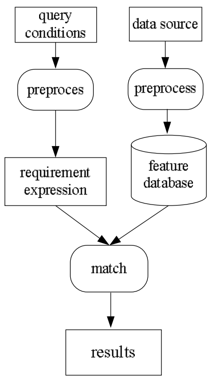


Figure 1. The principle of information retrieval

The following problems have been usually revealed in traditional information retrieval system since recent years. Firstly, it is difficult to accurately express users' information needs, and user requirements can not be achieved and expressed effectively. Secondly, for different users, even if their preferences are different remarkably, the identical results are returned only when the subscription or query conditions are identical. Obviously, users' personalized requirements are not met in this service mode.

To overcome the shortcomings, personalized information retrieval has got many attentions in recent years [3-9].

In fact, after the appearing of internet in 1960s, the emergence of WWW and widespread application promote the extreme expansion of online information. The flood of information leads to the problem of information overload when bring people facilitate. In this context, rapid and accurate access to personalized information resources has become an urgent need to address the problem.

Accordingly, many works on personalized information retrieval have been done [4-5, 8-12]. Personalized information retrieval is based on user characteristics and preferences of the search, returns the search results related to the user-specific information needs. The user profile model is the foundation and basis of personalized information retrieval. Personalized meta-search engine system Inquirus2 designed by NEC Research Institute is considered as a model of personalized information retrieval. The difference with general meta-search engines is to increase the description of user profile, and the user's information needs are expressed not through query keywords, but through both query keywords and user profiles. [13-14].

Project supported by National Natural Science Foundation of China (No.41001222) and by Scientific Research Foundation for JXNU Ph.D (No.3057).

978-1-61284-848-8/11/$26.00 ©2011 IEEE

II. Personalized Retrieval of Spatial Information

A. The Implementation Schema of Personalized Spatial Information Retrieval

The implementation schema of personalized spatial information retrieval is as Fig. 2 shows. The system combines query conditions provided by users with user profiles constructed by the system, and takes the combinations as retrieval conditions. Accordingly, related results can be got which are delivered to users subsequently. For the normal query results have different utility degrees for different users, so personalized retrieval re-ranking the results based on the utility degrees which are got from user profiles.

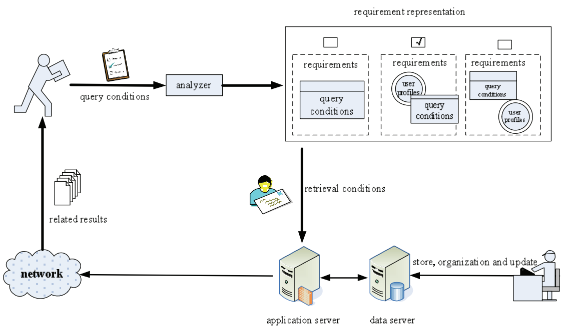


Figure 2. The implementation schema of personalized spatial information retrieval

B. Personalized Retrieval of Spatial Information Combining User Profile with Query Conditions

Supposed the user is called Uc and the user profile model for Uc is called Pro. Then, personalized spatial information retrieval can be expressed as (1) based on user profile models named Pro and query condition called c.

$$inf' = f_{Pro,c}(inf) \tag{1}$$

where inf are spatial information to be deliver to Uc, and inf' are all spatial information candidates available to Uc; $f(\cdot)$ is the delivery algorithm.

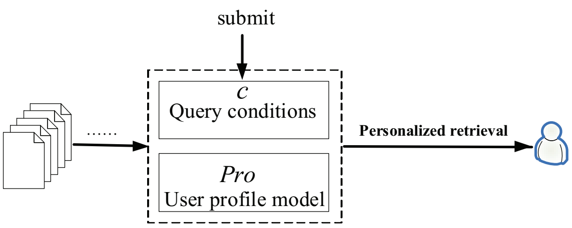


Figure 3. Personalized spatial information retrieval

.

C. The Data Processing Procedure of Personalized Retrieval of Spatial Information Combining User Profile with Query Conditions.

The data processing procedure of personalized retrieval of spatial information is shown in Fig. 4. Firstly, query conditions are provided by users, and are sent to middleware combined with user profiles. Then, the retrieval conditions are produced and then sent to registration center. Consequently, the personalized retrieval results are got and delivered to data receiving agent from data center.

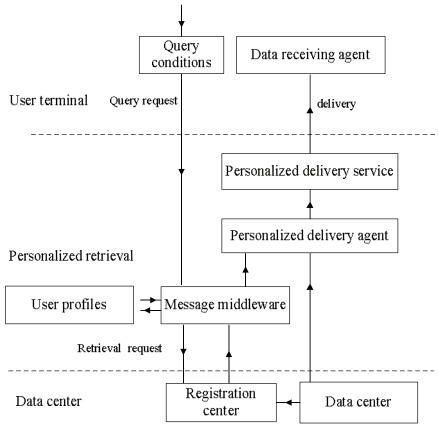


Figure 4. The data processing procedure of personalized retrieval of spatial information

III. The Algorithm on Personalized Retrieval of Spatial Information

Note that personalized retrieval of spatial information consists of three processes: general query, user profile constructing and re-rank the candidates. Firstly, the candidates without priority order are got from general query. Then, the user profiles are constructed based on the queries. The key steps of user profiles constructing are given as follows.

Algorithm: user profiles constructing
Input: query conditions $\{ c_j | j = 1, 2, \cdots, t\}$
Output: user profile C
Step:
1. For every j from 1 to t
   For every i from 1 to s

$$Count(g_{ij}(x_i)) = \begin{cases} 0, if\ g_{ij}(x_i) = \Phi \\ 1, if\ g_{ij}(x_i) \neq \Phi, \end{cases}$$

where $g_{ij}(x_i)$ is the sub-formula of x_i for expression c_j.
2. GetFreq() {
The total count of user query is:

$$d_t = t$$

For every i from 1 to s,
{
The times and frequency of the element x_i from users are following:

$$d_i = \sum_{j=1}^{t} count(g_{ij}(x_i)), i = 1, 2, \dots, s$$

$$p_i = d_i / d_t$$

} }

3. GetWei() {

For each i from 1 to s

{

The unitary item frequency is:

$$p_i' = p_i \Big/ \sum_{i=1}^{s} p_i$$

The weight is：

$$w_i = p_i'$$

} }

4. GetSubIntervalData() {

Observed values of the samples are Z_i (E_i), and the No. k observed value is $Z_{i,k} = [z_{i,k}^-, z_{i,k}^+]$, $k \in E_i = \{1, \cdots, n_i\}$

For every i from 1 to s

{

The minimum space for observed values of the samples is:

$$I_i = [\min_{k \in E_i}(z_{i,k}^-), \max_{k \in E_i}(z_{i,k}^+)]$$

Then, sorting the up-bound and low-bound sets of every sample in Z_i (E_i) which are $\{z_{i,k}^-, z_{i,k}^+ \mid i = 1, 2, \cdots, n_i\}$, and m_i non-uniform subintervals are following:

$$I_{i,g} = [\xi_{i,g-1}, \xi_{i,g}) \ , \ g = 1, \cdots, m_i - 1 \text{ and}$$

$$I_{i,m_i} = [\xi_{i,m_i-1}, \xi_{i,m_i}]$$

} }

where, $I_{i,g}$ is No. g subinterval of No. i item, and m_i is the number of subintervals.

5. GetProb() {

From every i from 1 to s

{

$$f_{i,g} = \frac{1}{n} \sum_{k \in E_i} I_k \left(Z_{i,k} \cap I_{i,g} \right)$$

} }

Where $I_{i,g}$ is the appeared frequency of No. i subinterval respectively.

6. GetFeaVal() {

For every i from 1 to s

{

$$s_{i,g} = \left(f_{i,g} - \left(2 \times \min(f_{i,g}) - \max(f_{i,g}) \right) \right) \Big/ \left(2 \times \left(\max(f_{i,g}) - \min(f_{i,g}) \right) \right)$$

} }

where $s_{i,g}$ is the feature value after linear transformation for the appeared frequency of subintervals.

7. GetIntenPro() {

For every i from 1 to s

{ $V_i = \left\{ < I_{i,g}, s_{i,g} > \mid g = 1, 2, \cdots, m_i \right\}$

Store(X_i, W_i, V_i)

}

Accordingly, the user profile is got and described formally as $C = \{< x_i, w_i, v_i > \mid i = 1, 2, \cdots, s\}\}$.

After the user profile is constructed, the utility degrees of the candidates for the user can be achieved based on the user profile and will be used for re-ranking the candidates. Consequently, personalized retrieval results with priority order for users are got.

IV. Experiments and Result Analysis

A. Personalized Retrieval Experiments

An experimental system for personalized spatial information retrieval combining user profile with query request has been implemented. The personalized retrievals of spatial information for user A and user B are shown respectively as Fig. 5 and Fig. 6.

Figure 5. Personalized spatial information retrieval results based on query conditions and user profile for user A

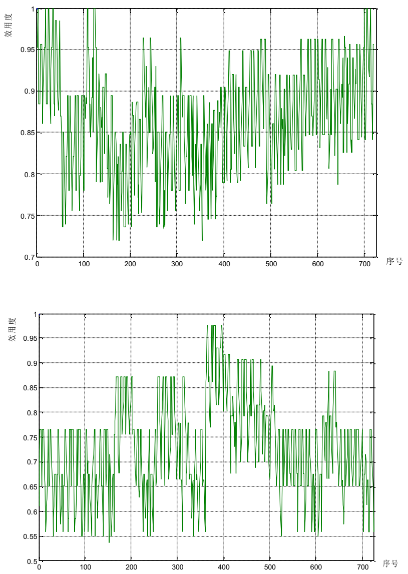
Figure 6. Personalized spatial information retrieval results based on query conditions and user profile for user B

Note that Fig. 5 is personalized retrieval results for user A, and first candidates are got from query conditions, and then the utility degrees of these candidates are calculated based on user profiles. Accordingly, the candidates are re-ranked based on utility degrees and so personalized retrieval results are achieved. Fig. 6 is personalized retrieval results for user B. The result identified with "48df1dfb-b6g9h7ac7-cc61-023534f8

9674" is appeared with highlighted text in Fig. 5 and Fig. 6 at the same time. However, the utility degrees are different for different users. Accordingly, personalized retrieval results are not the same position for the two users.

B. *Reslults Analysis*

The importance of the item for user A and user B is different in spite of its appearance in both retrieval results. The position for user A is No. 2, but it is No. 699 for user B. So, the utility degrees of all the results for user A and user B are shown in Fig. 7.



(a) (b)

Figure 7. The distribution of utility degrees for two users

Fig. 7 (a) is the distribution of utility degrees of results in normal retrieval for user A and (b) is the distribution of utility degree for user B. Accordingly, there is different distribution of utility degrees of normal retrieval results for different users. For example, the utility degree of the item identified "48df1dfb-b6g9h7ac7-cc61-023534f89674" is 0.745 for user A, while it is 0.976 for user B. So, personalized retrieval overcomes the shortcomings of general retrieval that the distributions of utility degrees for different users are not distinguished, and gets the personalized results in priority order and meets the needs of users more exactly.

REFERENCES

[1] Q. Fang, "A Cognitive Model for Information Retrieval Process," Journal of Information, vol. 7(1), 2007, pp. 54–56.

[2] G. Salton and M. E. Lesk, "Computer Evaluation of Indexing and Text Processing," Journal of the ACM, vol. 15(1), 1968, pp.8-36.

[3] L. Jéribi, "Improving Information Retrieval Performance by Experience Reuse," Electronic Publishing in the Digital Publishing Odyssey: Proc. of an ICCC/IFIP conf. held at the University of Kent, Kenterbury, UK, pp. 83-96, 2001.

[4] G. Jeh and J. Widom, "Scaling Personalized Web Search," Proc. of the International Conf. on World Wide Web, pp.271-279, 2003.

[5] S. Gauch, J. Chaffee, and Alexander Pretschner, "Ontology-Based Personalized Search and Browsing," Web Intelligence and Agent System, vol. 1(3), 2003, pp.219-234.

[6] F. Liu, C. Yu, and W. Meng, "Personalized Web Search For Improving Retrieval Effectiveness," IEEE Transactions on Knowledge and Data Engineering, vol. 16(1), 2004, pp.28-40.

[7] M. Speretta and S. Gauch, "Personalized Search Based on User Search Histories," Proc. of the IEEE/WIC/ACM International Conference on Web Intelligence, France, pp.622-628, 2005.

[8] P. Haase, N. Stojanovic, and J. Volker, "Personalized Information Retrieval in Bibster, a Semantics-Based Bibliographic Peer-to-Peer System," Proc. of the 5th International Conference on Knowledge Management (I-KNOW), Graz, Austria, pp.104-111, 2005.

[9] T. kim, H. Jeon, and J. Choi, "Personalized Information Retrieval using the User History," Proc. of the International Conference on Multimedia and Ubiquitous Engineering, pp.229-232, 2008.

[10] Y. Motomura, K. Yoshida, and K. Fujimoto, "Generative user models for Adaptive Information Retrieval," IEEE International Conf. on Systems, Man and Cybernetics, Nashville, TN, USA, pp.665-670, 2000.

[11] L. Jéribi, "Improving Information Retrieval Performance by Experience Reuse," Electronic Publishing in the Digital Publishing Odyssey: Proc. of an ICCC/IFIP conf. held at the University of Kent, Kenterbury, UK, pp. 83-96, 2001.

[12] G. Li, "The research and design of personalized internet information retrieval system," Ph.D thesis, the Chinese academy of sciences, 2002.

[13] E. J. Glover, S. Lawrence, and W. P. Birmingham, "Architecture of a Metasearch Engine that Supports User Information Needs," Proceedings of the International Conference on Information and Knowledge Management, Kansas City, Missouri, United Sates, pp.210-216, 1999.

[14] E. J. Glover, S. Lawrence, and M. D. Gordon, "Web Search-Your Way," Communications of the ACM, December vol. 44(12), 2001, pp.97-102.