New Generation Computing, 28(2010)113-120
Ohmsha, Ltd. and Springer

Building Web Knowledge Flows based on Interactive Computing with Semantics

Xiangfeng LUO[1], Jie YU[1], Qing Li[1,2],
Fangfang LIU[1] and Zheng XU[1]
1. School of Computer Engineering and Science, Shanghai University, CHINA
2. City University of Hong Kong, CHINA
{luoxf,jieyu,ffliu,xuzheng}@shu.edu.cn,
itqli@cityu.edu.hk

Received 20 February 2009
Revised manuscript received 14 April 2009

Abstract Web personalized services alleviate the burden of information overload by providing right information which meets individual user's needs. How to obtain and represent knowledge needed by users is a key issue. This paper presents *Web Knowledge Flow* (WKF) to represent the specific knowledge on Web pages and a model of Interactive Computing with Semantics (ICS) to provide a feasible means of generating WKF. *Objective WKF* (OWKF) and *Real-time WKF* (RWKF) are firstly proposed to satisfy staged and real-time user interests. Secondly, the generation algorithm of WKF is proposed based on Semantics Link Network. Thirdly, "interactive point" is introduced to detect the moment of user interests change to ensures the dynamics of WKF. Experimental results demonstrate that ICS can effectively capture the change of user interests and the generated WKF can satisfy user requirements accurately.

Keywords: Web Knowledge Flow, Interactive Computing with Semantics, Semantics Link Network, User Interest.

§1 Introduction

Web personalized services applications, such as e-Commerce alleviate the burden of information overload by providing right information which meets individual user's needs. How to obtain and represent knowledge needed by individual user is a key issue. Knowledge flow is a model to represent relationship between knowledge entities and can be used in organizing Web resources. Many efforts have been done on knowledge flow area and some achievements have been gained.[1,2] However, existing methods have some deficiencies. They are insensitive to user interests change and can not provide specific topics. In addition, they are independent of user's interaction, which make Web personalized services unable to provide topics which can meet user real-time requirements effectively. Aiming at solving these deficiencies, this paper presents *Web Knowledge Flow* (WKF) to represent knowledge on Web pages, along with a model of Interactive Computing with Semantics (ICS) to provide a feasible means of generating WKF that can meet user real-time requirements. In particular, we first propose *Objective WKF* (OWKF) and *Real-time WKF* (RWKF) to satisfy staged user interests and real-time interests. Next, we propose the generation algorithm of OWKF and RWKF based on Semantics Link Network (SLN)[3] with the help of ICS. The notion of "interactive point" is then introduced to detect the moment of user interests change, so as to ensure the dynamics of WKF. Consequently, WKF provides a feasible means of representing knowledge on Web pages for Web personalized services. For example, in a WKF-powered e-Commerce system, a user can obtain the specific goods which meet his/her requirements without wasting time. Therefore, WKF has wide application and a great impact on providing qualified Web personalized services.

The rest of this paper is organized as follows. Section 2 introduces the definition of WKF. Section 3 introduces the model of Interactive Computing with Semantics (ICS). The building method of WKF is given in section 4. Experimental results are discussed in section 5. In the last section, some conclusions are given.

§2 Web Knowledge Flow, WKF

In order to find specific resources in large number of disordered Web resources with (relatively) high performance, it is indispensable to well organize these resources. In our approach, Web resources are organized through Semantic Link Network (SLN),[3] which can link textual and multimedia causal relations. Based on SLN, WKF is built by Interactive Computing with Semantics. The definition of WKF is given as follows.

Definition 2.1 (Web Knowledge Flow, WKF)
Web Knowledge Flow (WKF) is a sequential link with rich semantics, which is activated by user's demands and changes with the demands. The nodes in WKF represent the Web resources which meet user's specific requirements.
A WKF can be represented by a two-tuple $F = (V, R)$ where $V = < v_1, v_2, ..., v_n >$ is an n-tuple. Each node is represented by an Element Fuzzy Cognitive Map (E-FCM),[4] the order of these nodes is in accordance to user requirement; $R = < r_1, r_2, ..., r_n >$ defines the semantic relationships between nodes, $r_i = (rt, w)$ where rt defines the relationship type and w defines the relationship degree.

Compared with other knowledge flows, WKF has some special characteristics, as listed below: (1) Web-dependant: WKF reflects the flow of knowledge among

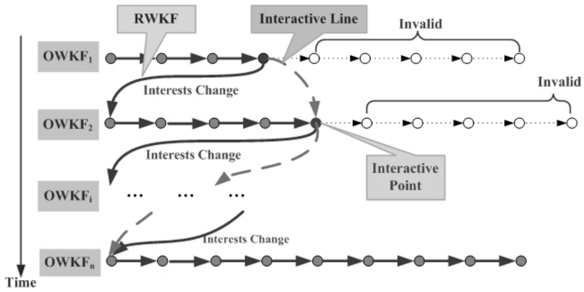


Fig. 1 Relationship between OWKF and RWKF

Web resources, and the generation of WKF is closely related to user's Web activities; (2) Semantics-rich: WKF contains rich semantics and reflects various relations between Web resources; (3) User-specific: The nodes of WKF and the type of semantic relationships are both dependant on user's demand; (4) Dynamics: WKF can adjust its organization according to user demands change. WKF is classified into two kinds: Objective WKF and Real-time WKF which are defined as follows.

Definition 2.2 (Objective Web Knowledge Flow, OWKF)
An *Objective WKF* (OWKF) is activated by user's interests at a certain period. Its validity is restricted by the user behavior and a time range.

Definition 2.3 (Real-time Web Knowledge Flow, RWKF)
A *Real-time WKF* (RWKF) is activated by the change of a user's interests. It grows based on OWKF with the proceeding of that user's browsing behaviors.

Definition 2.4 (Interactive Point, IP)
An *Interactive Point* is a node in OWKF and indicates the occurance of user's interests change. It activates a new OWKF and a new section of RWKF. IPs of OWKFs compose an interactive line which indicates the process of interests change.

Relationship between OWKF and RWKF is shown in Fig. 1. Due to user's different browsing stages, there are several OWKFs and one RWKF. Not all nodes of each OWKF will be visited. The validity of an OWKF ends when an IP occurs. The growth of RWKF is based on the generation of a new OWKF.

§3 Interactive Computing with Semantics, ICS

Our model of Interactive Computing with Semantics (ICS) is used to provide WKF by analyzing user's behaviors on the Web. ICS is built based on the theory of Interaction Machine[5)] which can model Web browsing process.

Definition 3.1 (Model of Interactive Computing with Semantic, ICS)
Model of Interactive Computing with Semantics is a tetrad $M = (S, I, O, F)$, where

- S is an enumerable set of states describing users' browsing states, each of which reflects a user's recent interests;

- I is an enumerable set of input states, which describes user's browsing behaviors;
- O is an enumerable set of output states, which describes the generated WKF;
- $F : S \times I\text{->} S \times O$ is a computable function with semantics of WKF.

An input represents a user's new behavior on the Web. And it will cause interactive computing. A state can be represented by the user profile[6] which records a user's browsing history and indicates his/her real-time interests. Therefore, the same input state may correspond to different output states, because of the different stages of user's interests reflected by the user profiles.

The properties of ICS are: 1) Interleaving of inputs and outputs. A computation function involves exchange of information between users and the Web. It is the iterative computation that generates WKF; 2) Semantics dependence. ICS is implemented with SLN which involves abundant semantics. The computation method is also based on the semantics of topic content; 3) History dependence. Not only an input state but also user's browsing history affects the result of ICS. Same inputs can generate different outputs due to users' browsing history.

§4 Generation of WKF based on ICS

Key essentials in generating WKF which are detection of IP and operator of ICS are discussed first. Then building process of WKF is introduced.

4.1 Implementation of ICS

An IP indicates the change of a user's interest which can be reflected from the situation that user browses a new topic which is not recommended by an OWKF. Based on this, we use "similarity degree" between topics that user newly browses and the one in OWKF to obtain an IP. Let c_n denote the topic user newly browses, $sd(v_i, v_j)$ denote similarity degree between topics, V denote state set in the current OWKF, the necessary condition for an IP to occur is

$$\exists v_i \in V, sd(v_i, c_n) < sthr \tag{1}$$

where $sthr$ is the threshold of similarity degree.

The occurrence of an IP indicates that a new OWKF which can satisfy the user's new interests should be built. Based on the situation that if a user is interested in one topic, he/she will be probably interested in the associated ones, the nodes in OWKF are the topics which have association relationship[7] with the topics of IP. Let V_n denote state set in newly-generated OWKF, c_n denote the topic of IP, $ad(v_i, v_j)$ denote association degree between topics, the following condition should be satisfied:

$$\exists v_i \in V_n, ad(v_i, c_n) < athr \tag{2}$$

where $athr$ is the threshold of association degree.

It is noticeable that in SLN there are many topics which can satisfy the above condition. Therefore, some topics which are called key topic should be selected to compose several candidate OWKFs. Here we use the average association degree between a node and its neighbors in SLN to choose key topics. The average association degree $aad(i)$ of $topic_i$ can be computed by

$$aad(i) = \frac{\sum_{j=1}^{n_i} ad(i,j)}{n_i} \quad (3)$$

where $ad(i,j)$ denotes the association degree between $topic_i$ and its neighbor $topic_j$ in SLN; n_i denotes the number of $topic_i$'s neighbor.

When key topics are selected, several candidate OWKFs are generated. To evaluate these OWKFs, we define effectiveness value which reflects OWKF's ability to provide abundant semantics with limited nodes. The effectiveness value of a candidate OWKF cf_i is

$$ef_i = \frac{\sum_{j=1}^{l_i-1} r_j \cdot w}{l_i} \cdot \frac{l_i}{max(l_1, l_2, \cdots, l_m)} \quad (4)$$

where $r_j.w$ denotes the association degree between the j^{th} node and $(j+1)^{th}$ node; l_i denotes the length of cf_i; and m denotes the number of candidate OWKFs.

4.2 Generation of Web Knowledge Flow

Based on the discussions above, the main steps of building a WKF are outlined as follows:

1) Identify an interactive point according to user profile;
2) Obtain topics from similar SLN based on user behaviors;
3) Start a new OWKF by closing the old OWKF and growing RWKF;
4) Choose the key topics via formula (3), and generate candidate OWKFs;
5) Evaluate candidate OWKFs, select the final one via formula (4), and get a new section of RWKF;
6) If the user proceeds with his/her Web activities, then go to step 1), else finish.

§5 Experiments

To demonstrate the usefulness of the proposed method, we traced users' browsing activities on the database of Web of Science, and OWKF and RWKF are generated to guide their browsing. In order to provide proper WKFs, we have built several SLNs which correspond to the fields of Computer architecture, education, and Web service (SLNs in Fig. 2 are only partly shown due to space of paper). User profile model that we use is Vector Space Model,[8)] viz: $u = < kw_1, kw_2, ..., kw_n >$, where kw_i denotes the weight of i^{th} keyword.

As a case study, after tracing a particular user browsing 20 papers, from user profile we obtain the keywords which have relatively high weights are *computer architecture, processor, register, memory, power*, and *design*, which indicates that the user is interested more in the field of computer architecture. Consequently, the corresponding candidate OWKFs in the 1^{st} stage are generated as

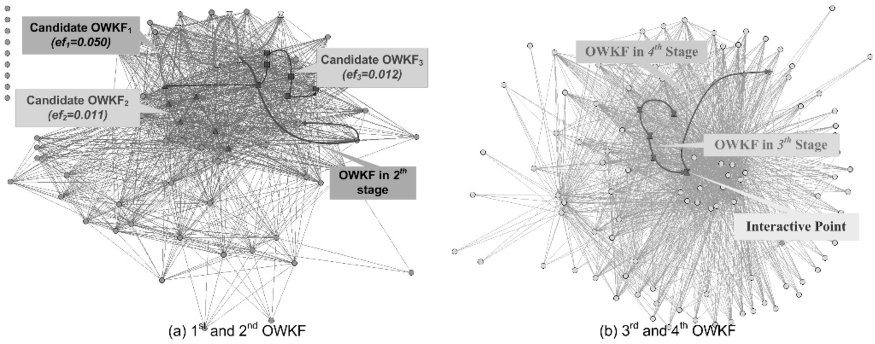


Fig. 2 OWKF of Different Browsing Stage

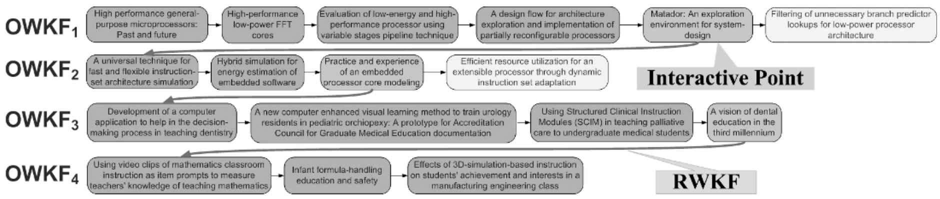


Fig. 3 Topics of each OWKF and RWKF

shown in Fig. 2(a). Based on the effectiveness values, the 2^{nd} one is chosen and recommended to the user. After reading four papers, the user searches for new papers in which the keyword *"simulation"* has the highest weight value. Thus, an IP is detected and corresponding OWKF is generated as shown by the red arrowed line in Fig. 2(a). In the whole process of this user's browsing, there are four OWKFs generated, which correspond to four browsing stages. At the end of the 3^{rd} stage, the user's interest changes into the filed of education (which has little relationship with the field of computer architecture) and is represented by another associated SLN shown in Fig. 2(b). Although the user's interest is still related to the topics of education in the 4^{th} stage, his interest changes from medicine and biology education to general education method. The generated OWKFs are shown in Fig. 2(b). Topics of all the OWKFs in the whole process the user's browsing and the generated RWKF are illustrated in Fig. 3.

It can be seen that an OWKF can accurately capture a user's interests within a time duration. RWKF can meet user's requirement during the whole process of Web activities, even if the user's interests may cross different fields.

§6 Conclusions

How to obtain and represent knowledge needed by an individual user is a key issue in personalized Web services. In this paper, we have proposed Web Knowledge Flow (WKF) to represent knowledge on the Web along with a model of Interactive Computing with Semantics (ICS) to provide a feasible means of generating WKF. Our contributions mainly include 1) The notions of Objective WKF(OWKF) and Real-time WKF(RWKF) are proposed to satisfy user interests; 2) OWKFs and RWKFs are generated based on SLN with ICS which takes the semantics of user's behavior into account; 3) The notion of interactive point is introduced to detect the moment of user interests change. Our initial experiment demonstrates that ICS can effectively capture the change of user interests and the generated WKF can greatly satisfy user requirements.

It can be seen that the proposed method has a brilliant perspective in the application of Web personalized services.

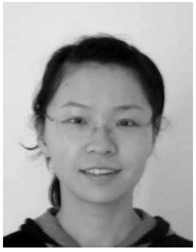
Acknowledgements

Research work is supported by the key basic research program of Shanghai Science and Technology Committee (project no. 09JC1406200), the National Science Foundation of China (grants 90612010 and 60402016), the National Basic Research Program of China (973 project no. 2003CB317008) and Shanghai Leading Academic Discipline Project (project no. J50103).

References

1) Nissen, M. E., "An Extended Model of Knowledge Flow Dynamics," *Communications of the Association for Information Systems*, pp. 251-266, 2002.
2) Zhuge, H., Guo, W., Li, X., Ding, L., "Knowledge Energy in Knowledge Flow Networks," in *Pro. Int. Conf. Semantics, Knowledge and Grid*, 2005.
3) Zhuge, H., "Autonomous semantic link network model for the Knowledge Grid," *Concurrency and Computation: Practice and Experience*, *7*, *19*, pp.1065-1085, 2007.
4) Luo, X. F., N. F., "Semantic Representation of Scientific Documents for the e-Science Knowledge Grid," *Concurrency and Computation: Practice and Experience*, *20*, *7*, pp. 839-862, 2008.
5) Wegner, P., "Interactive foundations of computing," *Theoretical Computer Science*, *192*, *2*, pp. 839-862, 1998.
6) Pazzani, M., et al., "Learning and Revising User Profiles: The Identification of Interesting Web Sites," in *Machine Learning*, pp. 313-331, 1997.
7) Luo, X. F., et al., "Discovery of Associated Topics for the Intelligent Browsing," in *Pro. 1st IEEE Int. Conf. Ubi-Media Computing and Workshops*, pp.119-125, 2008.
8) Raghavan, V. V., et al., "A critical analysis of vector space model for information retrieval," in *Journal of the American Society for Information Science*, *37*, *5*, pp. 279-287, 1986.

Xiangfeng Luo, Ph.D.: He finished his post-doc research in Institute of Computing Technology, Chinese Academy of Science in 2005. He joined the Joint Lab of Next-Generation Internet Interactive Computing in Shanghai University since 2005. His main research interests include knowledge flow, interactive computing, and cognitive semantics.



Jie Yu, Ph.D.: She received her Ph.D. in University of Science and Technology of China in 2007 and then joined the Joint Lab of Next-Generation Internet Interactive Computing in Shanghai University. Her main research interests include semantic computing, interactive computing with semantics and knowledge flow.

Qing Li, Ph.D.: He is a professor at the Department of Computer Science in City University of Hongkong. He received his Ph.D. in University of Southern California in 1988. His main research interests include object modeling, multimedia databases and web services.

Fangfang Liu, Ph.D.: She received her Ph.D. in Fudan University in 2007 and then joined the Joint Lab of Next-Generation Internet Interactive Computing in Shanghai University. Her main research interests include Web service retrieval, composition and knowledge discovery.

Zheng Xu: He received his B. S. degree in Shanghai University in 2007. He is a Ph.D. candidate. His main research interests include Semantic Web, Web search and mining, and interactive computing.