

Ontology-based User Modeling for Chinese Crop Germplasm Resources Information System

Lina Chen^{1, a}, Haiping Si^{2, b, *}, Yongsheng Cao^{3, c}

¹School of Information Technology Shangqiu Normal University Henan, China

²College of Information and Management Science Henan Agricultural University Zhengzhou, China

³Institute of Crop Science Chinese Academy of Agricultural Sciences Beijing, China

^adxn1126@163.com, ^b pingsss@126.com, ^c caoyongsheng@caas.cn

*Corresponding author

Keywords: User Modeling; Ontology; Personalized Service; Crop Germplasm Resources Information System.

Abstract: Personalized service has achieved great success in the field of e-commerce, it is also used to agriculture information service for improving service efficiency, and a key feature in developing successful personalized service is to construct user model that can represent accurately a user's interests. This paper presents an ontology-based dynamic user modeling approach to construct user model for Chinese crop germplasm resources information system (CGRIS). The proposed user model relies on individual register information, domain ontology and weblog, and this paper mainly deals with the problem of constructing user ontology and how user model is updated dynamically by weblog with time going. The user model is a user-centered design, and reflects the individual needs of the users. Due to the powerful knowledge representation formalism and associated inference mechanisms of ontology, the ontology-based user model provides an effective mechanism to provide targeted service for the users of Chinese crop germplasm resources information system.

1. Introduction

With the explosive growth of web information in many areas, people are often overloaded by information, and it is becoming more and more difficult to find relevant information. One possible approach to solve this problem is personalized recommendation by focusing on individual user needs. A user model that reflects the actual user needs is required for the successful personalization recommendation. Traditional user modeling methods are divided into explicit user modeling and implicit user modeling according to user access to information, the former gets user interests and preferences directly, and the latter does this by inference [1]. The user model is constructed using a variety of techniques including neural network [2], keyword statistics [3], ontology [4], tag [5] and social network [6]. Ontology expresses complex conceptual relations and provides a common framework that allows information to be shared and reused, ontology-based user modeling is the use of ontology to structure user model, so it reflects the actual user needs, and we focus on the role of semantically enhanced user modeling for personalized recommendation service. However, traditional ontology-based user modeling methods don't attach importance to ontology concepts, and these user models don't solve effectively the change of user interests with time going, so that these user models don't satisfy with personalized recommendation.

In this paper, we propose an ontology-based user modeling approach to construct agriculture information system user model. Firstly, we construct user ontology which aims to discover interesting from user register information as the input of ontology-based user modeling, due to the user information is more accurate, so is the initial user model. Secondly, we construct domain ontology and extend it, the domain ontology is constructed by object-oriented approach and extension is done by using common ontology such as WordNet and weblog. This paper places

special emphasis on the link between the user ontology and the domain ontology, the user ontology constructs the different characteristics of users and their relationships, and the domain ontology defines the specific concepts and their relationships, we construct initial user model by mapping the user ontology and the domain ontology. Finally, we focus on how the user model is updated dynamically by weblog, giving the formula how to calculate user interest degree. The ontology-based user model considers both different users' needs and time characteristics, the user model is updated dynamically.

The remaining part of this paper is structured as follows. Section 2 includes a brief literature review of ontology and user model. Section 3 describes technical details of the ontology-based user modeling method, including the method of domain ontology construction, the process of user ontology construction, and how the user model is updated. Section 4 applies the user modeling method to construct user model in CGRIS and in section 5 we conclude.

2. Ontology and User Model

2.1 Ontology

The term "ontology" originally comes from philosophy, it attempts to describe the existence of beings in the world, and later people apply ontology in many domains including knowledge engineering, natural language processing, knowledge representation, intelligent information integration and knowledge management. We agree with the definition as "An ontology is a formal, explicit specification of a shared conceptualization" given by Studer in 1998[7]. The ontology represents and structures the different knowledge sources, existing knowledge sources are mapped into the domain ontology and semantically enriched, this semantically enriched information enables better knowledge indexing and searching processes. It provides a shared vocabulary, that is, it is some objects and their attributes and mutual relationships in specific areas, it describes the relationships between information and information better, to implement reuse and sharing of information. In information system, user access behavior models can also be represented as ontology, and man-machine and machine-machine can communicate each other through ontology [8].

Studer emphasizes that constructing ontology for a domain requires a profound analysis, revealing the relevant concepts, attributes, relations, constraints, instances and axioms of that domain [7]. Currently, ontology library has been constructed in many fields, such as common ontology WordNet, gene engineering ontology, fisheries ontology, plants ontology and so on, however, there is no standard method to construct ontology because different fields care about different problems. A set of ontology languages such as OWL, KAON, RDF/RDFS, recommended by the World Wide Web Consortium facilitate the implementation of ontology-based applications, these languages enable expression and implementation of ontology-based conceptual models in a computational form.

2.2 User Model

A user model is defined as explicit or implicit representation of properties about individual user that includes preferences, knowledge and attributes for a domain [9]. User model is an integrated prototype of actual users, and it is a description of user information demand and individual behavior that are relatively stable in some period. It is basis and core to construct the right user model to implement personalized service, and the quality of user model is directly related to the quality of personalized service.

The current user modeling methods are divided into handmade modeling, sample user modeling and automatic user modeling by modeling way. Handmade modeling method is relatively simple, but it always needs input data manually, it has more workload, and its effect is entirely dependent on the user model. Sample user modeling method is to construct user model according to the related sample, which adds an additional burden. Automatic user modeling method is that the user model is constructed automatically based on users' historical behavior, the user model is more accurate, but the users' historical behavior is difficult to grasp, hence, it is difficult to construct [10]. User model is also divided into static model and dynamic model by update way, static model is more accurate, but

it can't reflect the actual user intentions better, dynamic model reflects the user needs within the nearest time, but it is more difficult to construct. User model is also divided into explicit model and implicit model according to information sources, generally, static user knowledge is gotten by explicit model, and dynamic user knowledge is gotten by implicit model.

3. Ontology-based User Modeling Construction Method

There are three main steps to construct ontology-based user model including domain ontology construction, user ontology construction and user model extension and update.

3.1 Domain Ontology Construction

The domain ontology describes the domain knowledge in terms of concepts and relationships between various concepts. There are many typical ontology construction methods including TOVE, IDEF5, skeleton, and seven-step methods [11], which provide a good methodology basis for ontology construction. But ontology construction is done for specific application domain, so it is impossible to copy a method for all domains. This paper uses the semi-automatic ontology construction method of combining structured data and unstructured data that we previously proposed [12]. According to this approach, the initial domain ontology is constructed by extracting concepts and their relationships from relational database based on object-oriented approach and from unstructured web data, then the semantic of these concepts is enriched by online dictionaries such as WordNet to improve the initial ontology. The method makes the database resources fully utilized, the complementary role of unstructured data considered, the cost of ontology construction reduced, and the ontology more perfected.

3.2 User Ontology Construction

Modeling user behavior is an ongoing challenge in many application domains which requires a better knowledge of the user in order to offer him/her enhanced service. There are many elements that need to be considered when constructing a user model for a specific application domain, most of the studies of internet user behavior are based on statistics and user traces, in this case, user behavior characterization takes into account statistics of user login information and web browsing strategies in order to predict trends or make recommendations to the users. The user ontology structures the characteristics of the users in concepts, subconcepts, properties and their relationships. When a user accesses information system, he generally registers basic information on website, the explicit information reflects the actual user needs and it is more accurate, so we construct the initial user ontology by user register information to improve the accuracy of recommendation system.

User ontology is expressed as: $U_{ontology} = \{U_{type}, U_{info}, U_{concept}, U_{interest}\}$.

U_{type} represents the user type, different type of users cares for different kinds of information, so users are grouped by the information type that they care about.

$U_{info} = \{U_{name}, U_{ID}, U_{dept}, U_{phone}, U_{addr}, U_{mail}\}$, U_{info} describes the user's information, it comes from user register information in information system. User name, ID number, department, telephone number, address and mailbox are included in this set, it is a description about user basic information.

$U_{concept} = \{S_{concept}, S_{relation}, S_{similarity}\}$, $U_{concept}$ represents the user concept set, it includes concepts set, semantic relationships set between the concepts, and conceptual similarity which are obtained from the domain ontology mappings. Similarity is introduced to tackle the mismatch between keywords input by users and keywords in user model, by calculating conceptual similarity, keywords that are semantic same or similar are found, and so the result is returned. Because ontology can be seen as the concept tree structure, the similarity of two concepts are mapped to the similarity of the two nodes in concept tree, so conceptual similarity is calculated by the node distance, the formula is expressed as: $S(C_i, C_j) = \frac{2 * \text{depth}(\text{LCA}(C_i, C_j))}{\text{depth}(C_i) + \text{depth}(C_j)}$, where C_i, C_j

represents any two concepts, $LCA(C_i, C_j)$ represents the shortest common ancestor of C_i, C_j , $depth(C_i)$ represents the depth of C_i in ontology tree.

Uinterest shows user interest degree which is a set of user interest degree, and it is calculated by concept visiting frequency. When system register information is counted, the visiting frequency of each type of information is calculated, and the preliminary user interest degree of certain information is also obtained, it is expressed as: $I = \{\sum_{i=1}^n I(C_i)\} = \{\sum_{i=1}^n frequency(C_i)\}$.

Typically, ontology is represented as tree structure, each node of the tree represents a user interest point, so that the domain ontology is reused. Meanwhile, different weights are given depending on the user's interest, so the initial user model is constructed by mapping user ontology and domain ontology [13].

3.3 User Model Extension and Update

Initial user model only contains the basic concept set and basic user information, and is not able to satisfy the personalized service demand. Due to implicit user behavior can be inferred from the data extracted from the log files that record the traces of users, it is a true user portrayal of user intention, and it reflects the actual user demand better. Large amounts of information such as keywords, URL address, click order, absolute visiting time (the ratio of time to document length) are extracted from the weblog [14], by adding some new concepts to the initial user model, initial user model is extended and enriched.

In addition, different users have different semantic search keywords in search information, so the initial user model is also extended by synonym. Common ontology library WordNet is used for synonym expansion in this paper, it is an online English dictionary that is developed by cognitive science laboratory at Princeton University, the words are composed a “word net” according to their meanings, and are composed a set of synonyms depending on speech, each set represents a basic vocabulary concept. WordNet has strong ability to express the relationship between words, so it improves search integrity by synonym mapping extension.

Initial user model is static, but the actual user demand always changes with time going. Initial user model is dynamically updated using weblog because it tracks the user's search trace. User interest degree is calculated as: $I = Inew * \frac{Mdate}{M} + Iold * (1 - \frac{Mdate}{M})$, $Inew$ represents the new interest degree obtained by weblog, $Iold$ represents the old interest degree, $Mdate$ represents the time interval between current time and the nearest time that users visited the system, M represents the parameter which is used to adjust the user interest degree, it is a time interval that user model updated in this paper. So, the user interest degree is calculated by old interest degree, new interest degree and impact factor. The user model after being updated has dynamic characteristics, so it reflects the user demand better.

Above all, the process of user modeling is shown in Figure 1.

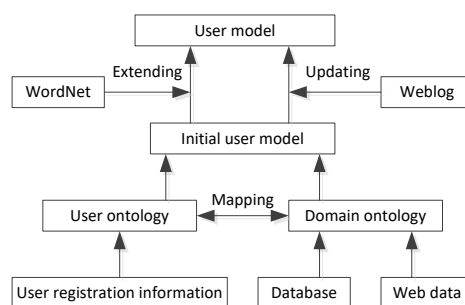


Fig. 1. Process of User Modeling

4. Application of User Modeling in CGRIS

4.1 CGRIS

Now China has more than 420,000 copies of crop germplasm resources, about 200 categories, and the work of crop germplasm resources generally includes resources collection, protection, identification and distribution. CGRIS has been providing information share and public service since it was established in 1980s, it has nearly 700 databases, 1,350,000 records, 230GB data, this system provides some information service including germplasm resources information inquiry, seed information, scientific knowledge, germplasm resources protection and distribution online for biotechnology researchers, breeders, agricultural technology popularization personnel, farmers and students, and it has achieved good service results [15]. Ontology-based knowledge representation and user modeling for crop germplasm resources domain provide an effective mechanism to improve the quality for CGRIS, the user model offers personalized service for different users and supports the tracking of service efficiency, so it is very important to construct CGRIS user model according to the characteristics of crop germplasm resources.

4.2 User Modeling for CGRIS

Ontology tool and ontology representation: Protégé is being developed at Stanford University, it is a free, open source ontology editor and ontology development tool based on Java language, and it is used to build concepts, relationships, attributes and instances. Protégé is one of the most widely used ontology editing tools at present, and it also includes deductive classifiers to validate that models are consistent and to infer new information based on the analysis of ontology. Protégé 4.3 is used to construct ontology in this paper, and Graphviz is used to visualize the ontology. Of all ontology languages, the Web Ontology Language (OWL) are one of the most common language that is used to construct ontology in semantic web, so we use OWL to construct ontology which is used to represent user information and crop germplasm resources information.

Crop germplasm resources ontology: According to the proposed method of ontology construction, the process of crop germplasm resources ontology construction has two steps including initial ontology construction from relational database and the semantic improvement by web data. We get concepts and relationships such as collection, protection, distribution and attribute from structured relationships database, and get concepts and relationships such as knowledge, seed information and exchange from web data. Figure 2 represents the crop germplasm resources ontology model.



Fig. 2. Crop Germplasm Resources Ontology Model

User ontology construction: According to the previously description about user ontology, Utype represents different user interests, the users of CGRIS are divided into different types according to their interests, each only care about a certain resource information, user interest is classified according to the resource type. For example, scientists who engage in rice research care mainly about rice information which includes rice seed information, rice science knowledge, rice protection etc., while other type of resource is not cared about by them. Uinfo represents the basic user information that comes from user register information in CGRIS, it constructs the basic user node. Uconcept

represents the concepts set interest users, relationships set between the concepts, it is consistent with domain ontology concepts and relationships, so it is easy to construct mapping between user ontology and domain ontology, and conceptual similarity is introduced to improve user query efficiency. Uinterest represents user interest degree about some a concept, it is one of the most important factors in user ontology.

The initial user interest may be regarded as zero, this paper uses concepts visited frequency of CGRIS weblog between 2010 and 2012 to calculate the user's initial interest. According to the resource type, the user interest is classified as food crops, fiber crops, oil crops, vegetable, fruit tree and other that don't belong to the former five kinds of resources. The initial interest degree of various resources is calculated as follows: $I(\text{Food}) = 0.23$, $I(\text{Fiber}) = 0.20$, $I(\text{Oil}) = 0.16$, $I(\text{Vegetable}) = 0.12$, $I(\text{FruitTree}) = 0.11$, $I(\text{Other}) = 0.08$.

The initial user model is constructed by mapping user ontology and domain ontology. The concepts set of user ontology is a subset of domain ontology concepts set, and besides concepts and their relationships, user type and the interest degree of concepts is also included in the initial user model.

User model extension and update: To enhance the semantic of these concepts in initial user model, we extend them use the common ontology WordNet, for example, we extend “maize” and “ear” as synonyms of the concept “corn”. The initial user model is also extended by weblog, the weblogs of CGRIS between 2013 and 2014 are analyzed, and only relatively high frequency concept is cared, if the keywords are not involved in initial user model, they are added to the initial user model, for example, chestnut, mulberry and other high frequency keyword are added to the initial user model.

The user information has little change in a shorter period, but it changes after a longer period. We calculate the user interest for a concept according to CGRIS weblog. If the keyword already exists in the initial user model, then new user interest degree is calculated, the ultimate user interest degree of resources is obtained from the initial user interest degree and the new user interest degree. The interest degree of grapes, cucumbers and melons is good example, the ultimate user interest degree is shown as follows: $I(\text{Food}) = 0.26$, $I(\text{Fiber}) = 0.14$, $I(\text{Oil}) = 0.13$, $I(\text{Vegetable}) = 0.14$, $I(\text{FruitTree}) = 0.24$, $I(\text{Other}) = 0.09$. The extended and updated new user model is shown in Figure 3.

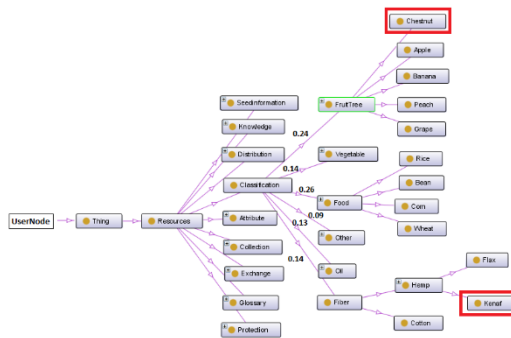


Fig. 3. New User Model

As shown from the user model, the user interest degree of food, vegetable and fruit tree has been greatly increased, this is because these types of resources are inquired by a large number of users in 2013 and 2014. For example, grape, chestnut and pistachio are inquired frequently from CGRIS weblog, in contrast, the inquiry frequency of fiber and oil is reduced.

5. Conclusion

This paper proposes an approach to construct user model for CGRIS by user register information, common ontology and weblog. Three key factors are taken into account in the process of the user model construction. Firstly, the conceptual similarity is introduced to the user model, and it is calculated to find the closest keywords, so the closest result is returned, the mismatch contradiction

of entered keywords and resources are solved effectively. Secondly, the initial user model is extended by common ontology WordNet and weblog, synonyms, different semantic words and keywords that was not in the user model are extended, to the user model meet higher level demand. Finally, the user model is regularly updated by weblog, so the user model represents not only the original user interest degree, but also the user's interest change with time going. Regularly updating of user model has smaller workload than frequently calculation, and it meets the user needs of CGRIS.

Ontology-based conceptual representations have shortcoming such as scalability, persistency, reliability, and transactions which are in classical data-base driven applications, and they are typically not available in ontology-based systems. However, the work is an effective attempt to construct service model for CGRIS, and it is also a useful exploration and promising direction for further information service.

Acknowledgements

This study was supported by Henan Science and Technology Research Project (182102110434), Henan Education Project (2018-JSJYZD-024), Shangqiu Normal University Ph.D. Research Project (4000203).

References

- [1] Shen X, Tan B, Zhai C X. Implicit user modeling for personalized search[C]//Proceedings of the 14th ACM international conference on Information and knowledge management. ACM, 2005: 824-831.
- [2] Jennings A, Higuchi H. A personal news service based on a user model neural network [J]. IEICE Transactions on Information and Systems, 1992, 75(2): 198-209.
- [3] Zukerman I, Albrecht D W. Predictive statistical models for user modeling [J]. User Modeling and User-Adapted Interaction, 2001, 11(1-2): 5-18.
- [4] Razmerita L, Angehrn A, Maedche A. Ontology-based user modeling for knowledge management systems [M]//User modeling 2003. Springer Berlin Heidelberg, 2003: 213-217.
- [5] Carmagnola F, Cena F, Cortassa O, et al. Towards a tag-based user model: How can user model benefit from tags? [M]//User Modeling 2007. Springer Berlin Heidelberg, 2007: 445-449.
- [6] Abel F, Gao Q, Houben G J, et al. Analyzing user modeling on twitter for personalized news recommendations[M] // User Modelling, Adaption and Personalization. Springer Berlin Heidelberg, 2011: 1-12.
- [7] Studer R, Benjamins V R, Fensel D. Knowledge engineering: principles and methods [J]. Data & knowledge engineering, 1998, 25(1): 161-197.
- [8] Xing Jiang, Ah-Hwee Tan. Learning and inferencing in user ontology for personalized semantic web search [J]. Information Sciences, 2009: 2794-2808.
- [9] Andrejko A, Barla M, Bieliková M. Ontology-based user modeling for web-based information systems[M]//Advances in Information Systems Development. Springer US, 2007: 457-468.
- [10] Hongguang Fu, Li Liu, Xiuqin Zhong. Semi-automatic construction of plane geometry ontology based-on WordNet and Wikipedia [J]. Journal of University of Electronic Science and Technology of China, 2014, 43(4): 575-580.
- [11] Gao G, Liu Y S, Wang M, et al. A query expansion method for retrieving online BIM resources based on Industry Foundation Classes [J]. Automation in Construction, 2015, 56: 14-25.

- [12] CHEN Li-Na, FANG Wei, SI Hai-Ping, and CAO Yong-Sheng. Ontology Construction of Crop Germplasm Resources. *ACTA AGRONOMICA SINICA*, 2016, 42(3): 407-414.
- [13] Pretschner A, Gauch S. Ontology based personalized search[C], Proceedings of the 11th IEEE International Conference on Tools with Artificial Intelligence. Chicago, USA, 1999:391-398.
- [14] Anissa Mimi, Safia Nait Bahloul. Protect user anonymity in query log[C], 2010 International Conference on Machine and Web Intelligence, ICMWI 2010 - Proceedings, 2010: 421-425.
- [15] Yongsheng Cao, Wei Fang. Establishment and application of national crop germplasm resources infrastructure in China [J]. *Biodiversity Science*, 2010, 18 (5): 454-460.