An Interaction Framework of Service-oriented Ontology Learning

Jingsong Zhang
Dept. of CSE,
Shanghai Jiao Tong University
800 Dongchuan Road,
Shanghai, 200240, P.R. China
+86-21-34204485
jasun_zhang@163.com

Yinglin Wang
Dept. of CSE,
Shanghai Jiao Tong University
800 Dongchuan Road,
Shanghai 200240, P.R. China
+86-21-34204415
ylwang@sjtu.edu.cn

Hao Wei
Dept. of Informatics,
University of Fribourg
Boulevard de Pérolles 90
Fribourg 1700, Switzerland
+41 76 619 47 56
hao.wei@unifr.ch

ABSTRACT

Ontology plays a very important role in supporting knowledge-based applications. In cloud computing, ontology learning technology is facing new challenges in dealing with heterogeneous data sources from different domains and researchers, which may contain various particular concepts and relations. Traditional ontology learning frameworks usually focus only on the extraction of concepts and taxonomic relations from the multi-structured corpus. However, former researches rarely studied the interactions during ontology learning process among different researchers. Lack of interactions among people who build ontology in different domains may cause inconsistent ontology. Besides, lack of incentive during the ontology building process will also result in low efficiency. To address these challenges, this paper specifies a novel solution to perform ontology learning. The solution includes a service-oriented ontology interaction framework, a service-oriented ontology learning strategy. It shows that it advances ontology learning to a higher level of performance and portability with a number of experiments in demo system.

Categories and Subject Descriptors

I.2.6 [**ARTIFICIAL INTELLIGENCE**]: Learning − Concept learning, Knowledge acquisition

General Terms

Theory, Experimentation

Keywords

Ontology, Ontology Learning, Ontology Interaction, Service-oriented Framework, Cloud Computing

1. INTRODUCTION

The continuing progress in network technologies and data storage has made possible the digitization and dissemination of huge amounts of documents, making it more and more difficult for the user to successfully search and retrieve information both in the Web and in a digital document collection, personal or otherwise [1]. Ontology, as a set of concepts and their interrelationships in a specific domain, is widely accepted that they can facilitate text understanding and automatic processing of textual resources. In addition, it has been proven to be a useful tool in the area of knowledge-based applications such as information integration and fusion, semantic web, knowledge discovery, intelligent information retrieval, and so on. Ontology is defined as a formal, explicit specification of a shared conceptualization [2]. Ontology learning is a subtask of information extraction in natural language processing. The goal of ontology learning is to automatically or semi-automatically extract relevant concepts and relations from a given structured, semi-structured or unstructured corpus including some other kinds of data sets to form an ontology set. Ontology is a conceptualization and should have three features: explicit, formal and shared. *Conceptualization* is a general model and is independent of any specific environment by extracting related concepts from real world; *Explicit* means that the concepts and their constraints are defined clearly in ontology; *Formal* indicates that the representation of ontology can be processed by computer (computer-readable); *Shared* signify that the ontology is common recognition knowledge and common concept sets in related domains.

Cloud computing provides computation, software, data access, and storage services that do not require end-user knowledge of the physical location and configuration of the system that delivers the services [3]. Service is the key feature of cloud computing, because cloud computing distributes a wide variety of requirements of users as a service rather than a specific product. Before the cloud computing era, storage resources, computing resources, software resources, data resources and management resources distributed in different Internet nodes in all parts of the world. These multiple resources could not be effectively integrated and utilized across heterogeneous systems, semantic web applications, and other common information pages. In cloud computing, these multi-resources are constructed as business service model such as SaaS(Software as a service), PaaS(Platform as a service), IaaS(Infrastructure as a Service), MSP(Media Studio Pro), etc. Therefore, the ontology, as data and shared resources, should be represented and reconstructed as service form, and at the same time, the research of service-oriented ontology learning will likely become one of research topics of ontology learning in future.

Athman Bouguettaya et al. [4] proposed an ontological framework for sharing and accessing government databases. The framework allows both FSSA case managers and citizens to access rich information from e-government databases. Jens Lehmann et al. [5] described how to adapt a semiautomatic method for learning OWL class expressions to the ontology

CIKM'12, October 29–November 2, 2012, Maui, HI, USA.

engineering use case and how to extend an existing learning algorithm for the class learning problem. In [6], a performance-oriented approach was presented and ontology was used for constructing formal and machine-understandable conceptualization of the performance-oriented learning environment. In addition, Kaihong Liu et al. [7-9] proposed and described some methods, algorithms or systems of ontology learning from specific details. For now, ontology learning frameworks are only in the microscopic point of view and focus only on concept and taxonomic relation extraction from the given corpus. However, they are all lack of interaction in ontology learning process between different researchers and fairly difficult to maintain ontology consistency among different domains. Besides, lack of incentive during the ontology building process will also result in low efficiency. In addition, isolation and lack of interaction ontology learning methods cannot meet date processing of service-oriented in cloud computing. In this paper, we present a novel an interaction framework of service-oriented ontology, a service-oriented ontology learning strategy.

The rest of this paper is organized as follows: section 2 introduces an interaction framework of service-oriented ontology, describes the service-oriented ontology learning model, and exhibits the database designing; section 3 exhibits the experimental results of our approach; section 4 makes a short conclusion to the work we have done and introduces our future work.

2. INTERACTION FRAMEWORK OF SERVICE-ORIENTED ONTOLOGY

2.1 Ontology Services Interaction Interface

In cloud computing, the traditional models and frameworks of ontology learning are fairly difficult to meet the needs of maintaining ontology consistency among different domains. We propose a solution called interface framework of ontology services to solve the challenge that the traditional ontology learning approaches cannot meet the actual requirements and real application. In this solution (see Figure 1), the interaction framework of Service-oriented ontology includes ontology learning part, service interface and user services. The detailed analyses of each part are as follows:

Ontology learning: In a specific method or algorithm perspective, ontology can be extracted and integrated from a given corpus. The ontology set of preliminary learning is likely to be rough and rich noise. For instance, the ontologies themselves may be incomplete, uncertain, redundant and contradictory. Therefore, it is necessary to adopt the ontology integrated approach to complement incompletion, reduce uncertainty, remove redundancy and resolve contradiction. At the same time, the result sets of ontology integration also help for ontology learning processes. Ontology itself also has a life cycle which includes the production, growth and demise in different domains or different times, though, some parts of ontologies have a longer life cycle span, while the other parts are shorter. So, the goal of ontology evolution is to manage the production, growth and demise of the ontology and maintain the ontology sets keeping robust. The traditional methods or algorithms can be used in these processes, but the traditional ontology learning processes themselves can not meet the requirements of service-oriented date processing. So we propose a novel ontology learning interaction model can be seen in Figure 2.

Ontology learning, integration and evolution are interdependent and mutually reinforcing. After several improving cycles, the relatively perfect ontologies can be added

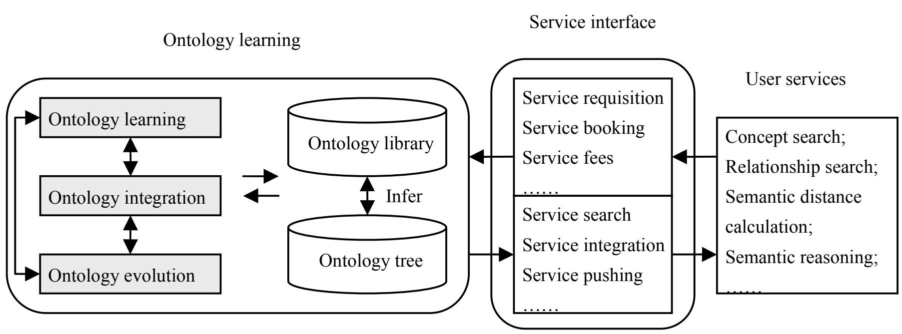

Figure 1. Interaction framework of Service-oriented ontology

to ontology library for supporting ontology retrieval and inference services. According to the four relationships: whole and part relationship (part-of), integrated relationship (kind-of), instance relationship (instance-of) and property relationship (attribute-of), the relation chains can be inferred and be built to relatively independent ontology trees. Both ontology library and ontology tree are useful resources for both researchers and ordinary users in the service-oriented cloud computing environment.

Service interface: Ontology resources generally include ontology library and ontology trees and they can be seen as service resources, but it would be difficult to provide effective services if they only used in some domains. The service interface of the framework is the interaction platform between ontology services providers and end-users or other researchers. In Figure 1, end-users or other researchers can submit their query requirements and query bookings in the service interface. From the counterpart, the service interface can search some most relevant information according to the requirements of the end-users or the researchers, and integrate the information to an(some) optimal result(s), and push it(them) to the users.

User services: User services departments usually include concept search, relationship search, semantic distance calculation, semantic inference and so on. These services not only can be some terminal services to the end-users but also can be a middle service to the researchers or business providers. For some researchers, they can do a further research and application using the ontology services. The middle service business providers can also restructure and repackage the services to some new services. The middle service business providers can be seen as one web service node in cloud computing.

2.2 Service-oriented Ontology Learning Model

The traditional ontology learning methods most focus on the expert methods or (semi)automatic methods in the interior or some domains rather than interaction of cross-domains and cross research teams which will inevitably lead to some low-quality ontologies including incompletion, uncertainty, redundancy and contradiction. In this model, we propose a novel ontology learning approach to address the challenges. In Figure 2, we adopt interaction learning approach among different teams and domains to build ontology. In one specific approach, we combine automatic and expert method to construct it in the interior of one domain or team.

Automatic method: The resource of ontology extraction is usually rich in a lot of noise such as interference symbol, mark, non-standard format, etc. Therefore, the resource called corpus should be pretreated through the given template or rule to standard Xml or text documents. And then, ontology concept, relationship and axiom can be extracted from the standard documents by utilizing Natural Language Processing (NLP) technologies. In the model, we utilize the upper ontology and domain keywords to improve the representation of ontology extraction. Rule-based and statistical-based approaches are available in the ontology concept, relationship and axiom extraction processes. At last, the results of optimization build a primary ontology set.

Expert method: Domain experts can build ontologies by their knowledge in the specific domains. However, almost all experts are not familiar with all knowledge even though in their own domains not to mention other domains, but most experts can offer same domain keywords or terms to help the building of primary ontology which is fairly helpful to improve the recall and precision in NLP. In addition, it also reduces significantly the time consumption of algorithm processing.

Ontology optimization: In cloud computing, a lot of teams do the similar ontology research or application in the same domain or different domains which will lead to low efficiency of work and some noise-rich ontologies. Actually, the primary ontology sets can be shared among different researchers and teams to reduce duplication work. So some alone ontology sets from respective researchers and teams can compose a large and shared ontology set. However, the set was built in different teams which may have different level of experts and different ontology learning approaches or algorithms, so it is inevitable that the set is noise-rich in some aspects such as incompletion, uncertainty, redundancy and contradiction. Utilizing the syntax, grammar, rules and statistics approaches to integrate the sets to improve the level of ontology representation.

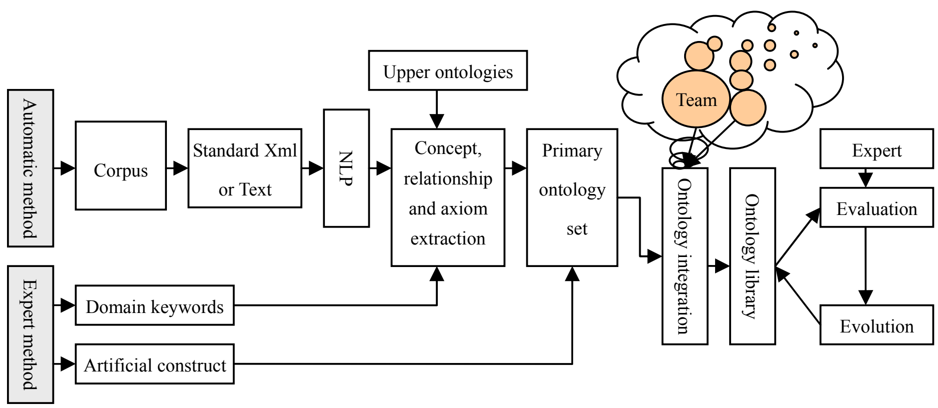

Figure 2. Service-oriented Ontology Learning Model

2.3 Database Design

The ontology stored in the database is one of key storage types, having itself advantages such as efficient reading and writing by computer. In the service-oriented ontology learning model, ontologies are from different nodes in cloud, so all information of ontology contributor should be stored as some fields' context in database which is convenient to maintain the ontology and calculate the reward of the ontology contributors. The details of ontology concept fields can be seen in Table 1.

Table 1. The fields of ontology concept

Field name	Data type	Length	Allow empty	Remark
No_id	int	8	No	Key, Randomly generated
concept_key	nvarchar	10	No	Ontology keywords
Concept	nvarchar	300	Yes	Ontology concept
source_text	nvarchar	1000	Yes	Source context
source_file	nvarchar	100	Yes	Source file name
domain_name	nvarchar	50	Yes	Domain Name
souce_id	nvarchar	10	Yes	Ontology builder id
review_id	nvarchar	10	Yes	Ontology reviewer id
use_times	int	8	Yes	pushing times

In Table 1, *No_id* field is generated randomly and is the key in this database table; *concept_key* is ontology keywords; concept is the detail description of ontology; *source_text* is the context text of ontology, and it can be a paragraph text; *source_file* is the file name including the ontology text; *source_id* is the person who is constructor of ontology;

domain_name is the domain name or industry name of the ontology; *review_id* is the reviewer (a person) of the ontology; *use_times* is the times of the ontology pushing.

3. EXPERIMENTAL RESULTS AND ANALYSIS

The development of service-oriented ontology learning system is a fairly complex and hard task. In this paper, for evaluating the ontology learning approaches, we choose some documents from advanced manufacturing technology domain as a test set to extract concept. These documents sources come from a real aircraft design agency and they include aircraft design manual, the component of aircraft structural design, physical terms of aircraft design and so on. We adopted the automatic methods of Figure 2 to extract primary concept. The main processes are as follows: the standard documents generation from corpus to text, segmentation of paragraphs and sentences, Chinese vocabulary segmentation, and concept extraction. We use feature-based terms extraction, template-based extraction, syntax-based extraction, anomaly exclusion and other NLP methods to do concept extraction.

From Table 2, we can see that most concepts include "是", however, they also include more noise because the recall rate (84.00%) is higher and the precision rate (66.14%) is lower. From the overall average statistics we can know that recall rate is 87.86% and the precision rate is 76.38% which is respectively higher than 76.92% and 60.61% in [10]. The recall rate and precision rate will be considerably improved if we adopt experts' participatory approach in concept extraction processes.

Table 2. The recall and rpecision rate of concept extraction.

Feature words	Number of concepts	Number of extraction	Number of appropriate	Recall rate	Precision rate
是(shi)	100	127	84	84.00%	66.14%
称(cheng)	29	30	27	93.10%	90.00%
指(zhi)	15	12	12	80.00%	100.00%
Null	29	30	29	100.00%	96.67%
Overall average	173	199	152	87.86%	76.38%

4. CONCLUSION AND FUTURE WORK

Ontology learning research has faced the challenge that does not meet the requirement of applications based on knowledge, and especially, traditional ontology learning models and frameworks are fairly difficult to maintain ontology consistency among different domains. This paper proposes a novel interaction framework of service-oriented ontology learning method including interaction framework of service-oriented ontology, service-oriented ontology learning model, database designing and incentive mechanism. These frameworks and methods considerably support the development of core-serviced theory in cloud computing. In addition, the framework had been built in our laboratory since last year. Some frame has been built and some function can be available such as ontology conception extraction, experts artificial maintaining and so on. In addition, the system was tested in complex product design processes and we got some considerable results. In the future, we will aim at the building of overall system and the representation of recall and precision by our model and approaches.

5. ACKNOWLEDGMENTS

Funding was provided by the National Natural Science Foundation of China (NSFC No.60773088), the National High-tech R&D Program of China (863 Program No. 2009AA04Z106), and the Key Program of Basic Research of Shanghai Municipal S&T Commission (No. 08JC1411700).

6. REFERENCES

[1] Katifori, A., Halatsis, C., Lepouras, G., Vassilakis, C., and Giannopoulou, E. 2007. Ontology visualization methods-A survey. ACM Comput. Surv. 39, 4, Article 10 (October 2007), 43 pages DOI = 10.1145/1287620.1287621 http://doi.acm.org/10.1145/1287620.1287621

[2] Studer R, Benjamins VR, Fensel D. Knowledge Engineering: Principles and Methods, IEEE Transactions on Data and Knowledge Engineering, 25(1-2): pages 161-199, 1998.

[3] cloud_computing, http://en.wikipedia.org/wiki/Cloud_computing.

[4] Athman Bouguettaya, Mourad Ouzzani, Ahmed Elmagarmid, Brahim Medjahed.Database middleware for distributed ontologies in state and federal family & social services.In Proceedings of the 2004 annual national conference on Digital government research, ACM International Conference Proceeding Series: Vol. 262.pages1-2,2004.

[5] Jens Lehmann, Sören Auer, Lorenz Bühmann, Sebastian Tramp. Class expression learning for ontology engineering. Web Semantics: Science, Services and Agents on the World Wide Web: 9(2011), pages 71–81.

[6] Haiyang Jia, Minhong Wang, Weijia Ran, Stephen J.H. Yang, Jian Liao, Dickson K.W. Chiu.Design of a performance-oriented workplace e-learning system using ontology. Expert Systems with Applications:38(2011),pages 3372–3382.

[7] Kaihong Liu, William R. Hogan, Rebecca S. Crowley.Natural Language Processing methods and systems for biomedical ontology learning.Journal of Biomedical Informatics:44 (2011),pages163–179

[8] Francesco Colace and Massimo De Santo.Ontology for E-Learning: A Bayesian Approach. IEEE TRANSACTIONS ON EDUCATION, VOL. 53, NO. 2, MAY 2010

[9] Kuo-Kuang Chu, Chien-I Lee, Rong-Shi Tsai. Ontology technology to assist learners' navigation in the concept map learning system. Expert Systems with Applications: 38 (2011) 11293–11299

[10] Yufang Zhang, Fen Yang, Zhongyang Xiong, Xiaoli Chen.Study on context-based domain ontology concept extraction and relation extraction. ApplicationResearch of Computers:2010,27(1), pages 74-76