A Novel Chinese Domain Ontology Construction Method for Petroleum Exploration Information

Jike Ge and Zushu Li
College of Automation, Chongqing University, China
Email: gjkweb@126.com; zushuli@cqu.edu.cn

Taifu Li
School of Electrical & Information Engineering, Chongqing University of Science and Technology, China
Email: litaifuemail@126.com

Abstract—**Ontology is playing an important role in knowledge management and sharing, both users and system can communicate with each other using a common understanding knowledge of a domain. This study proposes a context-based ontology construction method for extracting petroleum exploration domain information from unstructured Chinese text documents. The proposed mechanism of domain ontology construction includes four steps. First, domain documents preprocessing aims to separates the text into sentences, including a Chinese Part-of-Speech (POS) Tag and a Chinese corpus extract from the HowNet. Next, the concept clustering based on the fuzzy c-means aims to cluster concepts and instances from documents. In third step, context extraction aims to obtain the contexts. Finally, domain ontology construction aims to generate a petroleum exploration Chinese domain ontology. Experimental results show that the proposed approach can effectively construct Chinese domain ontology from unstructured text documents. This study implements a context-based ontology construction mechanism that can automatically mine domain concepts out of domain document, thereby reducing cost and burden that would be incurred in a manual construction process.**

Index Terms—**Ontology construction, Context, Petroleum exploration, Concept clustering**

I. INTRODUCTION

Large amounts of petroleum operational data are routinely collected and stored in the archives of many organizations. But, data in different organizations are complex in nature and often poorly organized and duplicated, and exist in different formats. Petroleum businesses face the problems of information overload. Effectively utilizing these massive volumes of data is becoming a major challenge for this type of industry. Exploration is one of the key operations of petroleum industry. Nimmagadda et al. [1] describe significance of exploration entity in oil and gas companies. They demonstrate the need of warehousing and data mining technologies in the petroleum companies.

Ontology is a knowledge model which defines concepts, attributes and relations in a specific domain with explicit specifications that feature interoperability between human and machine, and it can solve the problems of ambiguity in knowledge sharing and reuse. Due to its strengths in enhancing knowledge representation, sharing and reusing [2], ontology has found widely applications in areas like knowledge management [3], database design [4], information retrieval and information integration [5][6], and bioinformatics[7][8], and et al. Ontology is an essential part of many applications, and the development and application of ontology technology opens a new way toward knowledge sharing and reusing. Supported by ontology, both the user and the system can communicate with each other using a common understanding knowledge of a domain.

For the importance of ontology, ontology construction has been regarded as a significant issue. Ontology construction is a lengthy, costly and controversial works [9][10]. Hence, many studies for semi-automatic or automatic ontology construction methods have emerged [11-16]. For instance, Zhang and Jiang[11] proposed an approach to semi-automatically constructing domain ontology based on Chinese word partition and data mining. The proposed method is proved to be effective in constructing domain ontology, and also assured the quality of the ontology at a certain level. Hou, Ong, Nee and Zhang [13] proposed a named GRAONTO graph-based approach for automatic construction of domain ontology from domain corpus. In this method, first, each document in the collection is represented by a graph. After the generation of document graphs, random walk term weighting is employed to estimate the relevance of the information of a term to the corpus from both local and global perspectives. Next, the MCL (Markov Clustering) algorithm is used to disambiguate terms with different meanings and group similar terms to produce concepts. Next, an improved gSpan algorithm constrained by both vertices and informativeness is exploited to find arbitrary latent relations among these concepts. Finally, the domain ontology is output in the OWL format. Evaluation experiments show that GRAONTO is a promising approach for domain ontology construction. Nimmagadda and Dreher [15] Integrate ontologically structured data in a warehousing environment, and make it has more flexibility and consistency in attribute mapping and interpretation during data mining stage. Li

doi:10.4304/jcp.7.6.1445-1452

and Ko [16] used hierarchical clustering algorithm to bottom-up construction of ontology for diabetes diet care.

Shih, Chen and Chu [17] propose a concept relation exploration approach that combines the characteristics of middle-out and top-down approaches in a process that resembles snowflakes crystallization. Based on the crystallizing concept exploration approach, this study implements an ontology construction mechanism that can automatically mine domain concepts out of domain document, determine relations between concept, and construct the domain ontology accordingly.

However, regardless the theories and technologies being used, automatic ontology construction have always involved three major construction processes: document preprocessing, concepts extraction, and concept relations exploration [18][19]. Document preprocessing refers to filtering out noises in documents to retain meaningful terms; concepts extraction refers to extracting domain concepts out of vocabulary; and concept relations exploration refers to mining relations between concepts and organizing them to finish the ontology construction process. In the processing of ontology construction, relations between concepts and the ways concepts are organized by their relations influence the ontology structure, which in turn affects the accuracy of domain knowledge. Consequently, concept relations exploration is the most important process of ontology construction.

The existing concept relation exploration processes mainly follow the three approaches: top-down, bottom-up and middle-out [20]. Each of these approaches has its own strengths and weaknesses. A bottom-up approach identifies first the most outstanding concepts and generalizes them into more abstract concepts. However, a bottom-up approach finds it hard to spot commonality between related concepts. A top-down approach starting at the top can result in choosing and imposing arbitrary high level categories. A middle-out approach identifies the core of basic terms, and then specifies and generalizes them. The approach, by contrast, strikes a balance in terms of the level detail and requires less re-work, which also leads to less overall effort.

This study proposes a context-based ontology construction method for extracting petroleum exploration domain ontology from unstructured Chinese text documents. The proposed approach includes the steps of (i) domain documents preprocessing, (ii) concept clustering based on the fuzzy c-means, (iii) context extraction, and (iv) domain ontology construction. According to the context-based ontology construction approach, the proposed approach of ontology construction mechanism in this study that can automatically mine domain concepts out of domain document, determine relations between concepts, and construct the domain ontology accordingly, thereby reducing cost and burden that would be incurred in a manual construction process.

II. Ontology Construction Methods

So far, various ontology construction approaches have been presented recently [12-14]. The first contributions to ontology building methods are due to [21-24], representing the basis for many subsequent proposals. Gruber's seminal work discusses some basic ontology design criteria [22]. Some related to the quality of ontology building methodology (clarity and ontological commitment) and some related to the quality of the built ontology (coherence, extendibility, and minimal encoding bias). Gruninger and Fox provide a skeletal methodology for ontology building based on CQs [23], while Uschold and King [24] present a method based on four main activities: identification of the purpose of the ontology, building activity, evaluation, and documentation.

De Nicola, Missikoff and Navigli [25] propose an ontology building methodology that capitalizes the large experience drawn from a widely used standard in software engineering. The methodologies and the results of its adoption in the context of the Athena EU Integrated Project are also discussed. Ensan and Wu [26] attempt to study and investigate ontology development and maintenance frameworks from a domain-centric point of view. By frameworks they mean the structures which have been designed to allow ontology engineers and domain experts to develop and maintain domain ontologies. Such frameworks usually specify particular phases for developing ontologies and provide implemented components for each phase. Their purpose is to analyze the suitability of a framework for developing ontologies which can fulfill the necessities of a specific domain. Lee, Jiang and Hsieh [27] presented a meeting scheduling system based on the personal ontology and the fuzzy meeting scheduling ontology. They also presented some approaches for Chinese text processing, for instance, a episode-based ontology construction mechanism to extract domain ontology from unstructured text documents [10].

Following current studies of ontology construction [28] [11-14], a generic ontology construction procedure shows as Fig. 1, the processing of ontology construction include a domain document set as its input and domain ontology as its output, and involves three processes – document preprocessing, concept extraction, concept relation exploration.

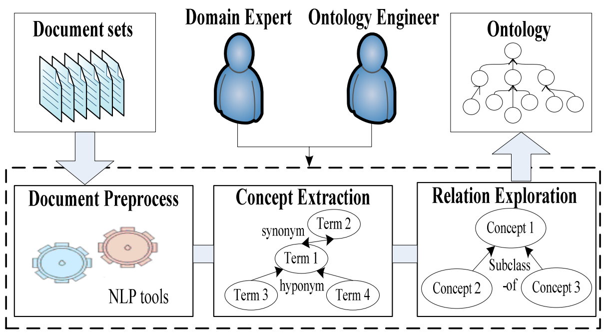

Figure 1. Generic ontology construction procedure

(1) Document preprocessing

The document preprocessing extracts meaningful terms by filtering out worthless symbols and words. The processes of word extraction, tokenization and part-of-speech analysis must be employed to filter out noise in the documents. Currently, several natural language process tools or resources are available for facilitating the document preprocess, including OpenNLP [29], CKIP [10], WordNet [30] and HowNet [31].

(2) Concept extraction

Concept extraction aims to identifying concepts using a domain thesaurus or topic map, based on lexical relations and groups of synonyms between terms. However, documents may have problems associated with inconsistent names, which may result in synonyms and homonyms, cause semantic disambiguation, and compromise the accuracy of concept extraction. Most previous studies in this field have relied on a domain thesaurus to solve the problem of semantic disambiguation by determining hyponyms and synonyms in the vocabulary [32]. However, if the vocabulary in a thesaurus is insufficient or unable to cover all domain concepts, then the extracted concepts may not adequately convey domain knowledge, reducing ontology accuracy.

(3) Concept relation exploration

Concept relation exploration refers to the mining of complete relations between concepts, and organizing the concepts to construct an ontology. Extracting relations between concepts is critical. Most studies in this field use top-down, bottom-up and middle-out approaches for exploring concepts [20]. Top-down: concept exploration begins at the most general concept, defined as the root concept, from which specialization is conducted downwards [25]; Bottom-up: concept exploration begins with the most specific concepts, identifying the bottom-level concepts, and then proceeds upwards to cover more general concepts [33]; Middle-out: the most outstanding concepts are identified and defined as middle-level concepts, from which generalization and specialization are conducted upwards and downwards, respectively.

However, all of these approaches have specific strengths and weaknesses. A bottom-up approach yields a very high level of detail, making the identification of commonality between related concepts difficult. A top-down approach that begins at the top can choose and impose arbitrary high-level concept. A middle-out approach, in contrast, strikes a balance between the level of detail and the amount of re-working, and it requires less effort overall. Therefore, integrating these models in a manner that eliminates the flaws of the individual models should greatly improve the effectiveness of concept relation exploration.

III. Context-based Domain Ontology Construction

This study integrates the generic ontology construction procedure, the concept of context and data mining methods, and proposes a context-based ontology construction model for constructing petroleum exploration domain ontology from Chinese domain documents.

A. The Definition of the Context

The concept of the context in this paper was defined as follows. A context c is formally defined as a triple (V, ≤, g), where V denotes a set of nodes; ≤ denotes a partial order on V, and g: V→ E denotes a mapping that associates each node with an event type E. The interpretation of a context is that the events in g(V) have to occur in the order described by ≤. A context c is parallel if the partial order ≤ is a trivial order (i.e., x not ≤ y for all x, y ∈V such that x ≠ y). Conversely, a context c is serial if the partial order≤ is a total order (i.e., x ≤ y or y ≤ x for all x, y ∈V). Informally, a context is a partially ordered collection of events occurring together. Context can be directed acyclic graphs. For example, consider context α, β, and γ in Fig. 2.

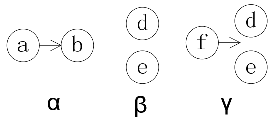

Figure 2. An example of a context

Context α is a serial context, and can only occur in a sequence that includes events of type *a* and *b* in that order. Context β is a parallel episode that does not impose constraints on the order of *d* and *e*. Context γ is a non-serial and non-parallel context, and can occur in a sequence in which occurrences of *f* precede occurrences of *d* and *e*, which may occur in any order.

B. Construction Process for Domain Ontology

This section applies the concept of the context assisting the construction process of Chinese petroleum exploration domain ontology from unstructured text documents. Additionally, the FCM algorithm [34] is adopted to cluster the concepts of Chinese terms. Fig. 3 displays the flowchart of the context-based Chinese petroleum exploration domain ontology construction process, which includes four processes, namely Document Preprocessing, Concept Clustering, Context Extraction, and Domain ontology construction, which are described below.

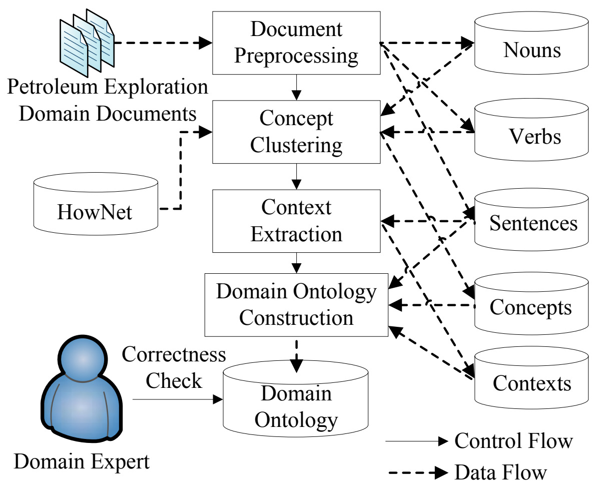

Figure 3. The flowchart of Context-based petroleum exploration Chinese domain ontology construction

(1) Document preprocessing

This process aims to separates the text into sentences, including a Chinese Part-of-Speech (POS) Tag and a

Chinese corpus extract from the HowNet. HowNet [31] is an on-line common-sense knowledge base unveiling inter-conceptual relations and inter-attribute relations of concepts as connoting in lexicons of the Chinese and their English equivalents. The corpus and dictionary provide adequate Chinese POS knowledge to analyze the features of the terms for semantic concept clustering.

The preserved terms used in this study are Na (common noun), Nb (proper noun), Nc (location noun), Nd (time noun) and various classes of verbs (VA, VB, VC, VD, VE, VF, VG, VH, VI, VJ, VK, VL). The filtered terms are stable noun, quantity noun, direction noun, pronoun, adjective, adverb, preposition, conjunction, particle, and interjection.

(2) Concept clustering

This process aims to cluster concepts and instances from documents. For selecting important terms for Concept Clustering, the nouns with the highest TF×IDF values are preserved and adopted, where TF is the term frequency and IDF is inverse document frequency [35]. In this process, the POS was selected as the concept similarity factors for analyzing the Chinese terms and calculating the concept similarity between any two Chinese terms based on the features of the Chinese language. Each node of the tagging tree denotes a Chinese POS tag defined by HowNet. The path length between two nodes is adopted to calculate the concept similarity in POS between any two Chinese terms. Each node of the tagging tree in Fig. 4 denotes a Chinese POS tag defined by HowNet.

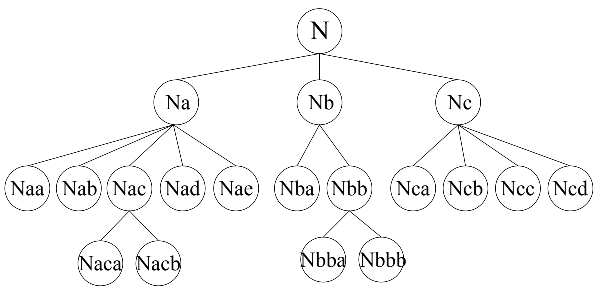

Figure 4. POS tagging tree of HowNet

The path length between two nodes is used to calculate the concept similarity w_{pos} in POS between any two Chinese terms. The value of w_{pos} is large when the path distance of any two Chinese terms is short. For instance, the two terms "计算机(computer)" and "软件(software)" have POS values of "Nab" and "Nad" respectively, the path distance between them is 2 (Nab→Na→Nad). The concept similarity w_{pos} is calculated as follows:

Calculating the concept similarity in POS algorithm

Input: All terms $(t_1,t_2,\ldots,t_n)$ selected from TF×IDF Selection

Output: Concept similarity w_{pos} in POS between any two Chinese terms

Step 1 Build a HowNet tagging tree

Setp 2 For all terms $(t_1,t_2,\ldots,t_n)$

Generate a term pair (t_a,t_b) 1≤a<b≤n

Path=length of path between two POS tags of (t_a,t_b) in HowNet tagging tree

$w_{pos}=(LB-path)/LB$ /* LB denotes the maximum path length in concept structure tree*/

Next, the fuzzy c-means (FCM) clustering algorithm [34][36] is adopted for concept clustering. FCM introduces the concepts of fuzzy logic to classic K-means, and it is one of the best known methods in fuzzy clustering. Fuzzy clustering allows an entity to belong to more than one cluster with different degrees of accuracy, while hard clustering assigns each entity exactly to one of the clusters. Thus, fuzzy clustering is suitable in constructing domain ontology because some information is not forced to fully belong to any one of the term. Fuzzy clustering methods may allow some information to belong to several terms simultaneously with different degrees of accuracy.

(3) Context extraction

The Document Preprocessing process separates the text into nouns, verbs and sentences, which are then fed into the Context Extraction process to obtain the contexts. This study denotes a term as a triple (*term, POS, index*), where *index* is the position of this *term* in the sentence. A context is extracted if the context occurs within an interval of a given window size, and the context's occurring frequency of the text document set is larger than the defined minimal occurrence value. To increase the accuracy of the context, the punctuation is filtered and the POS of terms with Na, Nb, Nc, Nd and verbs are retained in the sentence. The context extraction algorithm is as follows:

The context extraction algorithm

/*$T<t_1,t_2,\ldots,t_k>$: The set stores the term sequence $t_1, t_2,\ldots,t_k$ occurring in a given sentence.
$T<t_1,t_2,\ldots,t_k>$.cardinality: It denotes the number of item in $T<t_1,t_2,\ldots,t_k>$,and the number of occurrences of the term sequence $t_1,t_2,\ldots,t_k$. t_i.position: denotes the position of t_i in a sentence. sentence_num: The sequence number of a sentence.*/

Input: Sentences, Window_Size, Minimal_Occurence

Output: Contexts

Step 1 Generate Large 1-Sequence

For all terms t_i

Scan all sentences

If t_i appears in this sentence

Record sentence_num in $T<t_i>$

If $T<t_i>$.cardinality ≥ Minimal_Occurence

Add $<t_i>$ to Large 1-Sequence

Step 2 Generate Large 2-Sequence

For all permutations $<t_a,t_b>$ where $<t_a>$ and $<t_b>$ are selected from Large 1-Sequence

For all sentences with both $<t_a>$ and $<t_b>$

if t_b.position>t_a.position and t_b.position-t_a.position ≤ Window_Size

Record sentence_num in $T<t_a,t_b>$

if $T<t_a,t_b>$.cardinality ≥ Minimal_Occurence

Add $<t_a,t_b>$ to Large 2-Sequence

(4) Domain ontology construction

After obtaining the contexts, the terms are mapped to the result of the Concept Clustering to tag the concept name. Then, the Attributes, operations and associations are extracted from context according to the morphological information of the Chinese term and the Chinese syntax. Finally, the domain ontology construction algorithm shows as follows:

Domain ontology construction algorithm

Input: Context with the concept name
Output: Constructed domain ontology
Step 1 For all contexts c_i
If the number of terms in c_i is 2
If the first term t_1 is an instance and the POS of the second term t_2 is Nouns or VH
The second term t_2 is an attribute of this instance t_1.
If the first term t_1 is an instance and the POS of the second term t_2 is VA
The second term t_2 is an operation of this instance t_1.
If the number of terms in c_i is 3
If the first term t_1 and the third term t_3 are instances, and the POS of the second term t_2 is a transitive verb, status transitive verb or Nouns
The second term t_2 is an association of the instance t_1 and t_3.
Step 2 Output domain ontology

Following the above mentioned procedure, we can construct generic domain ontology. Generally, in generic ontology construction, if the stages of concept similarity computation and context extraction suffer from the problems of incomplete extraction of concepts and poor context of relations, respectively, it would preventing the constructed ontology from effectively describing domain knowledge.

IV. Experimental Results

A. Experiment Vvaluation Design

In this study, a total of 165 experiment documents were selected accordingly and then verified by experts as the basis for petroleum exploration Chinese ontology construction. Next, the constructed ontology was compared with ontology constructed by experts according to the indictors of Precision, Recall, and F measures to evaluate the accuracy rate of ontology created by proposed context-based ontology construction. Precision refers to the percentage of accurate concepts extracted, as shown in Eq. (1), whereas Recall refers to percentage of expert-defined concepts extracted, as shown in Eq. (2). Precision and Recall can be summarized into another metric known as the F measure (Eq. (3)).

$$precision = \frac{|\{Retrieved\ Concepts\} \cap \{Expertsdefined\ Concepts\}|}{|\{Retrieved\ Concepts\}|} \quad (1)$$

$$Recall = \frac{|\{Retrieved\ Concepts\} \cap \{Experts\ defined\ Concepts\}|}{|\{Experts\ defined\ Concetps\}|} \quad (2)$$

$$F = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3)$$

Furthermore, in order to verify the performances and features of context-based petroleum exploration Chinese domain ontology construction method, two experiments were included in this study: (1) analyzing the difference in using different document sets for the context-based petroleum exploration Chinese domain ontology construction; and (2) evaluating the impact of precision and recall when adjusting minimal occurrence and window size on the petroleum exploration Chinese domain ontology construction.

B. Document Sets Evaluation

Experiment document set: petroleum exploration Chinese domain ontology construction is based on domain documents, and the number of documents being used may make a difference to ontology thus generated. Short documents contain fewer concepts, and the ontology constructed from such document samples also contain less concepts, which may result in incomplete concepts or the missing of important concepts in the domain. Longer documents, on the other hand, contain more concepts, and the ontology thus constructed also contains more concepts, which may lead to the problem of excessive concepts. Consequently, in this study experiment documents was divided into two sets of abstract and full-text samples to analyze the difference in ontology constructed with different samples in the processing of the petroleum exploration Chinese domain ontology construction.

This Experiment uses abstract documents and full-text documents as experiment samples to conduct petroleum exploration Chinese domain ontology with four group different documents which extracting from the 165 experimental documents, and the results is shown as Table 1.

TABLE I. The Results of Document Sets Evaluation

Abstract				Full-text			
No.	concepts	branch	depth	No.	concepts	branch	depth
1	10	6	3	1	56	25	3
2	8	4	3	2	48	22	3
3	9	5	3	3	51	23	3
4	11	8	3	4	60	30	3

In the experimental results, “concepts” stands for the total number of ontology concepts; “branch” stands for the number of ontology’s leaf; “depth” stands for the number of ontology layers. However, in a constructed ontology based on abstract documents, both the number of concepts and its branch are relatively low. This is because the number of terms contained in the abstract contents is limited, and it leads to a low number of mined concepts. On the contrary, a constructed ontology based on full-text documents obviously has more concepts as well as wider branch compared to its Abstract counterpart.

This is because a full-text is longer, contains more concepts, thus resulting in more mined concepts and an ontology with more concepts that are more complete and comprehensive.

C. Precision and Recall Evaluation

This experiment aims to investigate the difference of precision and recall between the proposed petroleum exploration Chinese domain ontology construction method and domain experts. The results of Experiment are shown in Table 2,

TABLE II. THE EXPERIMENT FOR PRECISION AND RECALL

Min	Win	concepts	branch	depth	precision	recall	F-measure
3	10	40	32	3	0.65	0.15	0.24
4	10	41	32	3	0.60	0.18	0.28
5	10	45	33	3	0.62	0.22	0.32
6	10	48	35	3	0.57	0.37	0.45
7	10	58	40	3	0.55	0.43	0.48
8	10	67	56	3	0.48	0.43	0.45
9	10	80	70	4	0.41	0.47	0.44
10	10	112	96	4	0.36	0.52	0.43

* The Min and Win are the minimal occurrence and window size, respectively

As shown in Table 2, when minimal occurrence gradually increased, concepts, branch and recall also increased while precision decreased. On the whole, high minimal occurrence led to an increase in F measure, and low minimal occurrence led to a decrease in F measure.

Finally, the result of petroleum exploration Chinese domain ontology construction with optimal F-measure (F = 0.48, minimal occurrence = 7, Precision = 0.55 and Recall = 0.43) was derived, and the constructed ontology is shown as Fig. 5. The architecture of petroleum exploration Chinese domain ontology shows as Fig. 6.

Figure 5. The constructed petroleum exploration domain ontology

V. Conclusions and Future Work

This study proposes a petroleum exploration Chinese domain ontology construction model that features and integration of middle-out and top-down approaches, The technologies like context mechanism, natural language process, word segmentation for Chinese and clustering were employed to develop an ontology construction mechanism from unstructured Chinese text documents. Experimental results indicate that the proposed approach can successfully construct the petroleum exploration Chinese domain ontology. The quality of ontology derived from proposed ontology construction method might not be as good as what can be constructed by experts, still it can serve as an primitive ontology that assists experts in collecting and organizing concepts related to domain knowledge and the relations between these concepts. Based on such a primitive ontology, experts only have to increase or decrease certain parts of the concepts and make minor adjustments to their relations to obtain ontology with a better fitness to the domain. The proposed ontology construction mechanism can be used to accelerate the process of constructing domain ontology and reduce the cost for purely manual construction of domain ontology.

However, for some special cases, such as a domain with rapid changing terms and concepts or with complex semantics, it is very difficult to construct appropriate domain ontology.

Future work will include efforts to improve the precision of the proposed method. The proposed approach will also be applied to other languages with semantic corpus or semantic dictionaries.

Acknowledgment

The work described in this study is partially supported by the Science and Technology Research Project of Chongqing Municipal Education Commission of China under Grant No. KJ101413 and the Natural Science Foundation Project of CQCSTC under Grant No. 2010BB2285, and the Research Foundation Project of Chongqing University of Science & Technology under Grant No. CK2010B06.

References

[1] S. L. Nimmagadda, H. Dreher, and A. Rudra, "*Data Warehouse Structuring Methodologies for Efficient Mining of Western Australian Petroleum Data Sources*," The 3rd IEEE International Conference on Industrial Informatics, pp. 611-616, 2005.

[2] N. Lammari, and E. Metais, "*Building and maintaining ontologies: A set of algorithms*," Data and Knowledge Engineering, vol. 48, pp. 155-176, 2004.

[3] C. Brewster, and K. O'Hara, "*Knowledge representation with ontologies: Present challenges–Future possibilities*," International Journal of Human-Computer Studies, vol. 65, pp. 563-568, 2007.

[4] V. Sugumaran, and V. C. Storey, "T*he role of domain ontologies in database design: An ontology management and conceptual modeling environment*," ACM Transactions on Database Systems, vol. 31, pp. 1064-1094, 2006.

[5] L. Wang, and M. Li, "*Information retrieval model for the semantic search*," Journal of Computational Information Systems, vol. 3, pp. 1359-1366, 2007.
[6] P. C. Smits, and A. Friis-Christensen, "*Resource discovery in a european spatial data infrastructure*," IEEE Transactions on Knowledge and Data Engineering, vol. 19, pp. 85-95, 2007.
[7] S. K. Bechhofer, R. D. Stevens, and P. W. Lord, "*GOHSE: Ontology driven linking of biology resources*," Journal of Web Semantics, vol. 4, pp. 155-163, 2006.
[8] Y. Liu, Z. Sui, Y. Hu, and Q. Zhao, "*Research on automatic construction of medical ontology*," International Conference on Biomedical Engineering and Computer Science, pp. 1-4, 2010.
[9] W. Zhou, Z. Liu, and Y. Zhao, "*Ontology learning by clustering based on fuzzy formal concept analysis*," In Proceedings of the 31st Annual International Computer Software and Applications Conference, pp. 204-210, 2007.
[10] C. S. Lee, Y. F. Kao, Y. H. Kuo, and M. H. Wang, "*Automated ontology construction for unstructured text documents*," Data & Knowledge Engineering, vol. 60, pp. 547-566, 2007.
[11] D. Zhang, L. Zhang, and H. Jiang, "*Research on Semi-automatic Domain Ontology Construction*," Seventh Web Information Systems and Applications Conference, pp. 115-118, 2010.
[12] X. Zhou, G. Xu, and L. Liu, "*An Approach for Ontology Construction Based on Relational Databases*," International Journal of Research and Reviews in Artificial Intelligence, vol. 1, pp. 16-19, 2011.
[13] X. Hou, S.K. Ong, A.Y.C. Nee, X.T. Zhang, and W.J. Liu, "*GRAONTO: A graph-based approach for automatic construction of domain ontology*," Expert Systems with Applications, vol. 38, pp. 11958-11975, 2011.
[14] J. Kim, P. Kim, and H. Chung, "*Ontology construction using online ontologies based on selection, mapping and merging*," International Journal of Web and Grid Services, vol. 7 pp. 170-189, 2011.
[15] S. L. Nimmagadda, and H. Dreher, "*Petroleum Ontology: an effective data integration and mining methodology aiding exploration of commercial petroleum plays*," The IEEE International Conference on Industrial Informatics, pp. 1289-1295, 2008.
[16] H. Li, and W. Ko, "*Automated food ontology construction mechanism for diabetes diet care*," International conference on machine learning and cybernetics, vol. 5, pp. 2953-2958, 2007.
[17] C. W. Shih, M. Y. Chen, H. C Chu, and Y. M. Chen, "*Enhancement of domain ontology construction using a crystallizing approach*," Expert Systems with Applications, vol. 38, pp. 7544-7557, 2011.
[18] R. Chen, J. Liang, and R. Pan, "*Using recursive ART network to construction domain ontology based on term frequency and inverse document frequency*," Expert Systems with Applications, vol. 34, pp. 488-501, 2008.
[19] M. Y. Dahab, H. A. Hassan, and A. Rafea, "*TextOntoEx: Automatic ontology construction from natural English text*," Expert Systems with Applications, vol. 34, pp. 1474-1480, 2008.
[20] M. Fernández-López, and A. Gómez-Pérez, "*Overview and analysis of methodologies for building ontologies*," The Knowledge Engineering Review, vol. 17, pp. 129-156, 2002.
[21] O. Corcho, M. Fernandez-Lopez, and A. Gomez-Perez, "*Methodologies, tools and languages for building ontologies: where is their meeting point?*," Data Knowledge Engineering, vol. 46, pp. 41-64, 2003.
[22] T. R. Gruber, "*A translation approach to portable ontology specification*," Knowledge Acquisition, vol. 5, pp. 199-220, 1993.
[23] M. Gruninger, and M. S. Fox, "*Methodology for the design and evaluation of ontologies*," Workshop on Basic Ontological Issues in Knowledge Sharing in International Joint Conference on Artificial Intelligence, pp. 1-10, 1995.
[24] M. Uschold, and M. King, "*Towards a methodology for building ontologies*," Workshop on Basic Ontological Issues in Knowledge Sharing in International Joint Conference on Artificial Intelligence, 1995.
[25] A. De Nicola, M. Missikoff, and R. Navigli, "*A software engineering approach to ontology building*," Information Systems, vol. 34, pp. 258-275, 2009.
[26] F. Ensan, and W. Du, "*Towards domain-centric ontology development and maintenance frameworks*," Nineteenth International Conference on Software Engineering & Knowledge Engineering, pp. 622-627, 2007.
[27] C. S. Lee, C. C. Jiang, and T. C. Hsieh, "*A genetic fuzzy agent using ontology model for meeting scheduling system*," Information Sciences, vol. 176, pp. 1131-1155, 2006.
[28] A. Gómez-Pérez, and D. Manzano-Macho, "*An overview of methods and tools for ontology learning from texts*," The Knowledge Engineering Review, vol. 19, pp. 187-212, 2004.
[29] D. Sánchez, and A. Moreno, "*Learning non-taxonomic relationships from web documents for domain ontology construction*," Data and Knowledge Engineering, vol. 64, pp. 600-623, 2008.
[30] G. A. Miller, "*WordNet: A lexical database for English*," Communications of the ACM, vol. 38, pp. 39-41, 1995.
[31] Z. Dong, "*Bigger Context and Better Understanding -- Expectation on Future MT Technology*," Proceedings of the International Conference on Machine Translation & Computer Language Information Processing, pp.17-25, 1999.
[32] K. Tan, H. Han, and R. Elmasri, "*Web data cleansing and preparation for ontology extraction using WordNet*," In Proceedings of the First International Conference on Web Information Systems Engineering, pp. 11-18, 2000.
[33] J. Park, J. Nam, Q. Hu, and H. Suh, "*Product ontology construction from engineering documents*," International conference on smart manufacturing application, pp. 305-310, 2008.
[34] J. C. Bezdek, and R. Ehrlich, "*FCM: the fuzzy c-means clustering algorithm*," Computers & Geosciences, vol. 10, pp.191–203, 1984.
[35] A. Shiri, *Introduction to Modern Information Retrieval*, 2nd ed., Emerald Group Publishing Limited, CA, 2004.
[36] J. Kang, L. Min, Q. Luan, and X. Li, "*Novel modified fuzzy c-means algorithm with applications*," Digital Signal Processing. Vol.19, pp. 309-319, 2009.

Jike Ge, he was born in Puyang City, Henan Province of China in 1977. In the year of 2009, he received the Ph.D degree in artificial intelligence from Southwest University of China.

He is currently working as a Postdoctoral Research Fellow in the College of Automation, Chongqing University, Chongqing, China, and his cooperation tutor is Professor Zushu Li.

His research area centers on artificial

intelligence methods such as ontology, knowledge discovery, neural networks, genetic algorithms, intelligent control theory and application， and support vector machines.

Zushu Li, he was born in Sichuan province of China in 1945. Professor, He is currently working as a teacher in Chongqing University, China.

His research area includes intelligent control theory and application, intelligent automation, artificial intelligence, Human-emulated intelligence, knowledge discovery.

Taifu Li, he was born in Sichuan province of China in 1971. In the year of 2004, he received the Ph.D. degree in Mechanical Design & Theory from Chongqing University, China.

He is working as a professor in school of Electrical & Information Engineering, Chongqing University of Science and Technology, China.

His research interests include intelligent control, soft sensor, complex system modeling & its optimization.

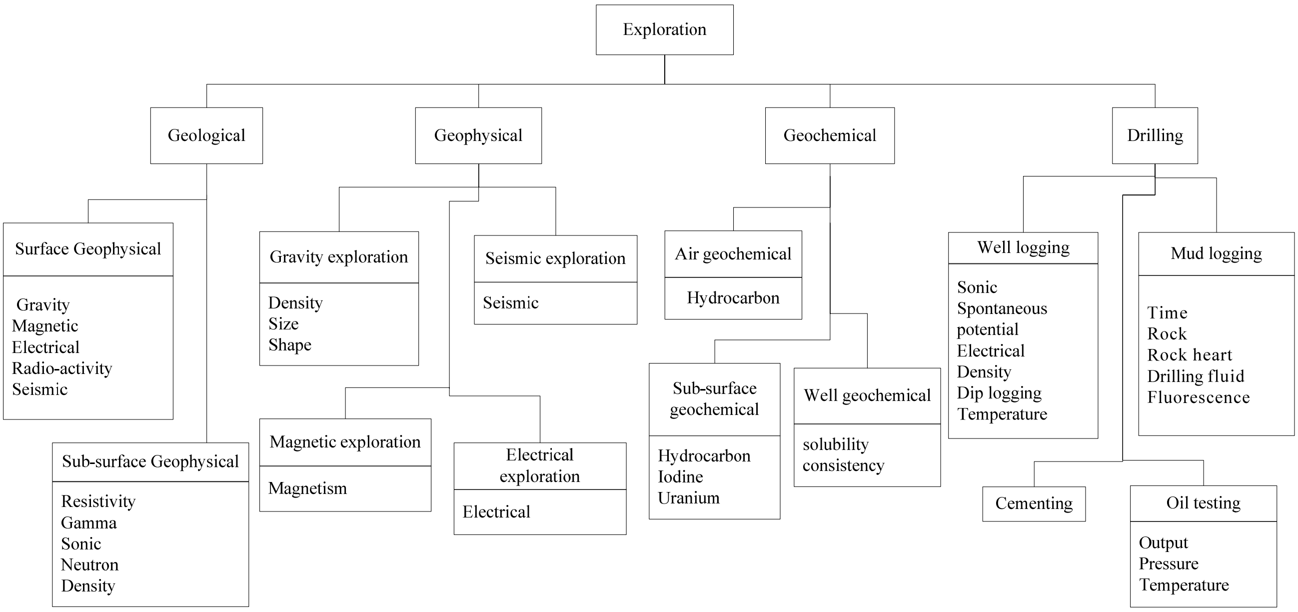

Figure 6. The architecture of petroleum exploration Chinese domain ontology