Automated Alignment and Extraction of Bilingual Ontology for Cross-Language Domain-Specific Applications

Jui-Feng Yeh, Chung-Hsien Wu, Ming-Jun Chen and Liang-Chih Yu
Department of Computer Science and Information Engineering
National Cheng Kung University, Tainan, Taiwan, R.O.C.
{jfyeh, chwu, mjchen,lcyu}@csie.ncku.edu.tw

Abstract

In this paper we propose a novel approach for ontology alignment and domain ontology extraction from the existing knowledge bases, WordNet and HowNet. These two knowledge bases are aligned to construct a bilingual ontology based on the co-occurrence of the words in the sentence pairs of a parallel corpus. The bilingual ontology has the merit that it contains more structural and semantic information coverage from these two complementary knowledge bases. For domain-specific applications, the domain specific ontology is further extracted from the bilingual ontology by the island-driven algorithm and the domain-specific corpus. Finally, the domain-dependent terminologies and some axioms between domain terminologies are integrated into the ontology. For ontology evaluation, experiments were conducted by comparing the benchmark constructed by the ontology engineers or experts. The experimental results show that the proposed approach can extract an aligned bilingual domain-specific ontology.

1 Introduction

In recent years, considerable progress has been invested in developing the conceptual bases for building technology that allows knowledge reuse and sharing. As information exchangeability and communication becomes increasingly global, multilingual lexical resources that can provide transnational services are becoming increasingly important. On the other hand, multi-lingual ontology is very important for natural language processing, such as machine translation (MT), web mining (Oyama et al. 2004) and cross language information retrieval (CLIR). Generally, a multi-lingual ontology maps the keyword set of one language to another language, or compute the co-occurrence of the words among languages. In addition, a key merit for multilingual ontology is that it can increase the relation and structural information coverage by aligning two or more language-dependent ontologies with different semantic features.

Over the last few years, significant effort has been made to construct the ontology manually according to the domain expert's knowledge. Manual ontology merging using conventional editing tools without intelligent support is difficult, labor intensive and error prone. Therefore, several systems and frameworks for supporting the knowledge engineer in the ontology merging task have recently been proposed (Noy and Musen 2000). To avoid the reiteration in ontology construction, the algorithm of ontology merging (UMLS http://umlsks.nlm.nih.gov/) (Langkilde and Knight 1998) and ontology alignment (Vossen and Peters 1997) (Weigard and Hoppenbrouwers 1998) (Asanoma 2001) were invested. The final ontology is a merged version of the original ontologies. The two original ontologies persist, with aligned links between them. Alignment usually is performed when the ontologies cover domains that are complementary to each other. In the past, domain ontology was usually constructed manually according to the knowledge or experience of the experts or ontology engineers. Recently, automatic and semi-automatic methods have been developed. OntoExtract (Fensel et al. 2002) (Missikoff et al. 2002) provided an ontology engineering chain to construct the domain ontology from WordNet and SemCor.

Nowadays vast investment is made in ontology construction for domain application. Finding the authoritative evaluation for ontology is becoming a critical issue. Some evaluations are integrated into the ontology tools to detect and prevent the mistakes. The mistakes that might be made in developing taxonomies with frames are described in (Gómez-Pérez 2001). They defined three mainly types of mistakes: Inconsistency, Incompleteness, and redundancy. To deal with these mistakes and carry out the validation and verification of ontology, some ontology checkers, validators and parsers were developed. These tools provide the efficacious appraisal of correctness when developing the new ontology. However, they are disappointing in ontology integration, especial when the original ontologies are well defined. For other approaches (Maedche and Staab 2002), the similarity measures are proposed in the earlier stage of the evaluation. The evaluation consists two layers: lexical layer and conceptual layer. In lexical layer, the edit distance is integrated into the lexical similarity measure. The measure is defined as:

$$SM\left(L_i,L_j\right)\equiv\max\left(0,\frac{\min\left(\left|L_i\right|,\left|L_j\right|\right)-ed\left(L_i,L_j\right)}{\min\left(\left|L_i\right|,\left|L_j\right|\right)}\right)\in\left[0,1\right] \qquad (1)$$

where $SM\left(\cdot\right)$ denotes the lexicon similarity function, $ed\left(\cdot\right)$ is the Levensthein edit distance function defined in (Levensthein. 1966). L_i and L_j are the words within the lexicons of the ontologies. The conceptual layer focuses on the conceptual structures of the ontologiesm namely taxonomic and nontaxonomic relations.

In this paper, WordNet and HowNet knowledge bases are aligned to construct a bilingual universal ontology based on the co-occurrence of the words in a parallel corpus. For domain-specific applications, the medical domain ontology is further extracted from the universal ontology using the island-driven algorithm and a medical domain corpus. Finally, the axioms between medical terminologies are derived. The benchmark constructed by the ontology engineers and experts is introduced to evaluate the bilingual ontology constructed using the methods proposed in this paper. This paper defines two measures, taxonomic relation and non-taxonomic relation, as the quantitative metrics to evaluate the ontology.

The rest of the paper is organized as follows. Section 2 describes ontology construction process and the web search system framework. Section 3 presents the experimental results for the evaluation of our approach. Section 4 gives some concluding remarks.

2 Methodologies

Figure 1 shows the block diagram for ontology construction. There are two major processes in the proposed system: bilingual ontology alignment and domain ontology extraction.

2.1 Bilingual Ontology Alignment

In this approach, the cross-lingual ontology is constructed by aligning the words in WordNet to their corresponding words in HowNet.

The hierarchical taxonomy is actually a conversion of HowNet. One of the important portions of HowNet is the methodology of defining the lexical entries. In HowNet, each lexical entry is defined as a combination of one or more primary features and a sequence of secondary features. The primary features indicate the entry's category, namely, the relation: "is-a" which is in a hierarchical taxonomy. Based on the category, the secondary features make the entry's sense more explicit, but they are non-taxonomic. Totally 1,521 primary features are divided into 6 upper categories: Event, Entity, Attribute Value, Quantity, and Quantity Value. These primary features are organized into a hierarchical taxonomy.

First, the Sinorama (Sinorama 2001) database is adopted as the bilingual language parallel corpus to compute the conditional probability of the words in WordNet, given the words in HowNet. Second, a bottom up algorithm is used for relation mapping.

In WordNet a word may be associated with many synsets, each corresponding to a different sense of the word. For finding a relation between two different words, all the synsets associated with each word are considered (Fellbaum 1998). In HowNet, each word is composed of primary features and secondary features. The primary features indicate the word's category. The purpose of this approach is to increase the relation and structural information coverage by aligning the above two language-dependent ontologies, WordNet and HowNet, with their semantic features.

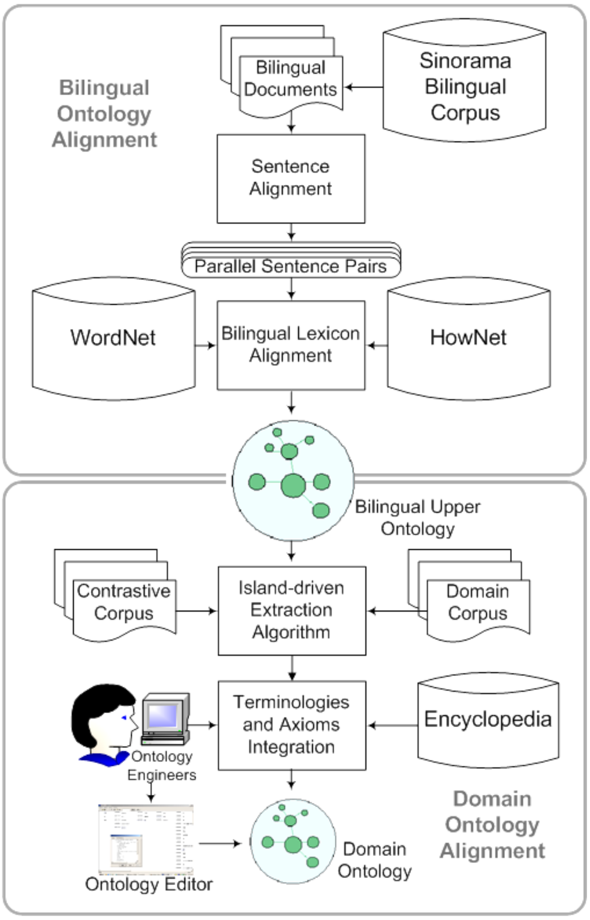


Figure 1 Ontology construction framework

The relation "is-a" defined in WordNet corresponds to the primary feature defined in HowNet. Equation (2) shows the mapping between the words in HowNet and the synsets in WordNet.

Given a Chinese word, CW_i , the probability of the word related to synset, $synset^k$, can be obtained via its corresponding English synonyms, $EW_j^k, j=1,...,m$, which are the elements in $synset^k$. The probability is estimated as follows.

$$
\begin{aligned}
&\Pr(synset^k \mid CW_i)\\
&= \sum_{j=1}^{m} \Pr(synset^k, EW_j^k \mid CW_i)\\
&= \sum_{j=1}^{m} (\Pr(synset^k \mid EW_j^k, CW_i) \times \Pr(EW_j^k \mid CW_i))
\end{aligned}
\tag{2}
$$

where

$$
\begin{aligned}
&\Pr\left(synset^k \mid EW_j^k, CW_i\right)\\
&= \frac{N\left(synset_j^k, EW_j^k, CW_i\right)}{\sum_l N\left(synset_j^l, EW_j^k, CW_i\right)}
\end{aligned}
\tag{3}
$$

In the above equation, $N\left(synset_j^k, EW_j^k, CW_i\right)$ represents the number of co-occurrences of CW_i , EW_j^k and $synset_j^k$. The probability $\Pr\left(EW_j^k \mid CW_i\right)$ is set to one when at least one of the primary features, $PF_i^l\left(CW_i\right)$, of the Chinese word defined in the HowNet matches one of the ancestor nodes of synset , $synset_j^k\left(EW_j\right)$ except the root nodes in the hierarchical structures of the noun and verb; Otherwise the probability $\Pr\left(EW_j^k \mid CW_i\right)$ is set to zero.

$$
\begin{aligned}
&\Pr\left(EW_j \mid CW_i\right)\\
&= \begin{cases}
1 \; if & \left(\bigcup_l PF_i^l\left(CW_i\right) - \{entity, event, act, play\}\right) \cap \\
& \left(\bigcup ancestor(\bigcup_k synset_j^k(EW_j)) - \{entity, event, act, play\}\right) \neq \varnothing \\
0 & otherwise
\end{cases}
\end{aligned}
\tag{4}
$$

where {enitity,event,act,play} is the concept set in the root nodes of HowNet and WordNet.

Finally, the Chinese concept, CW_i , has been integrated into the synset , $synset_j^k$, in WordNet as long as the probability, $\Pr(synset^k \mid CW_i)$, is not zero. Figure 2(a) shows the concept tree generated by aligning WordNet and HowNet.

2.2 Domain ontology extraction

There are two phases to construct the domain ontology: 1) extract the ontology from the cross-language ontology by the island-driven algorithm, and 2) integrate the terms and axioms defined in a medical encyclopaedia into the domain ontology.

2.2.1 Extraction by island-driven algorithm

Ontology provides consistent concepts and world representations necessary for clear communication within the knowledge domain. Even in domain-specific applications, the number of words can be expected to be numerous. Synonym pruning is an effective alternative to word sense disambiguation. This paper proposes a corpus-based statistical approach to extracting the domain ontology. The steps are listed as follows:

Step 1 Linearization: This step decomposes the tree structure in the universal ontology shown in Figure 2(a) into the vertex list that is an ordered node sequence starting at the leaf nodes and ending at the root node.

Step 2 Concept extraction from the corpus: The node is defined as an operative node when the Tf-idf value of word W_i in the domain corpus is higher than that in its corresponding contrastive (out-of-domain) corpus. That is,

$$
\begin{aligned}
&operative\_node(W_i)\\
&= \begin{cases}
1, & if \;\; Tf-idf_{Domain}(W_i) > Tf-idf_{Contrastive}(W_i)\\
0, & Otherwise
\end{cases}
\end{aligned}
\tag{5}
$$

where

$$
\begin{aligned}
&Tf-idf_{Domain}(W_i)\\
&= freq_{i,Domain} \times \log \frac{n_{i,Domain} + n_{i,Contrastive}}{n_{i,Domain}}\\
&Tf-idf_{Contrastive}(W_i)\\
&= freq_{i,Contrastive} \times \log \frac{n_{i,Domain} + n_{i,Contrastive}}{n_{i,Contrastive}}
\end{aligned}
$$

In the above equations, $freq_{i,Domain}$ and $freq_{i,Contrastive}$ are the frequencies of word W_i in the domain documents and its contrastive (out-of-domain) documents, respectively. $n_{i,Domain}$ and $n_{i,Contrastive}$ are the numbers of the documents containing word W_i in the domain documents and its contrastive documents, respectively. The nodes with bold circle in Figure 2(a) represent the operative nodes.

Step 3 Relation expansion using the island-driven algorithm: There are some domain concepts not operative after the previous steps due to the problem of sparse data. From the observation in ontology construction, most of the inoperative concept nodes have operative hypernym nodes and hyponym nodes. Therefore, the island-driven algorithm is adopted to activate these inoperative concept nodes if their ancestors and descendants are all operative. The nodes with gray background shown in Figure 2(a) are the activated operative nodes.

Step 4 Domain ontology extraction: The final step is to merge the linear vertex list sequence into a hierarchical tree. However, some noisy concepts not belonging to this domain ontology are operative. These nodes with inoperative noisy concepts should be filtered out. Finally, the domain ontology is extracted and the final result is shown in Figure 2(b).

After the above steps, a dummy node is added as the root node of the domain concept tree.

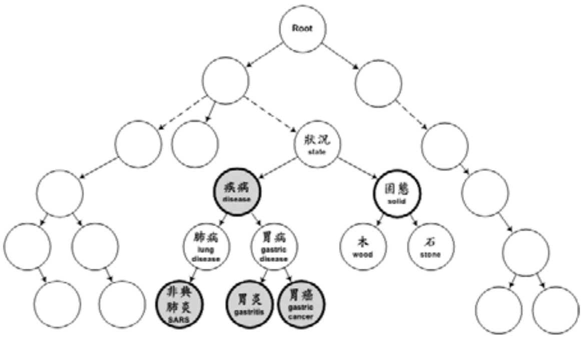


Figure 2(a) Concept tree generated by aligning WordNet and HowNet. The nodes with bold circle represent the operative nodes after concept extraction. The nodes with gray background represent the operative nodes after relation expansion.

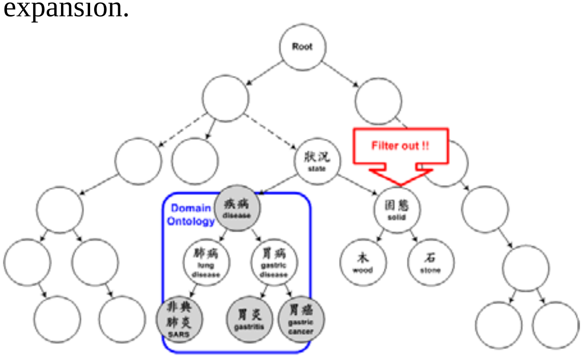


Figure 2(b) The domain ontology after filtering out the isolated concepts

2.2.2 Axiom and terminology integration

In practice, specific domain terminologies and axioms should be derived and introduced into the ontology for domain-specific applications. There are two approaches to add the terminologies and axioms: the first one is manual editing by the ontology engineers, and the other is to obtain from the domain encyclopaedia.

For medical domain, we obtain 1213 axioms derived from a medical encyclopaedia about the terminologies related to diseases, syndromes, and the clinic information. Figure 3 shows an example of the axiom. In this example, the disease "diabetes" is tagged as level "A" which represents that this disease is frequent in occurrence. And the degrees for the corresponding syndromes represent the causality between the disease and the syndromes. The axioms also provide two fields "department of the clinical care" and "the category of the disease" for medical information retrieval or other medical applications.

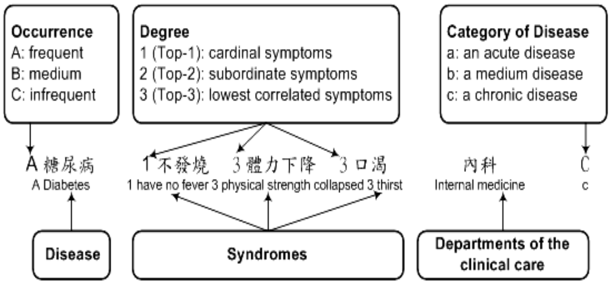


Figure 3 One example of the axioms

3 Evaluation

For evaluation, a medical domain ontology is constructed. A medical web mining system is also implemented to evaluate the practicability of the bilingual ontology.

3.1 Conceptual Evaluation of Ontology

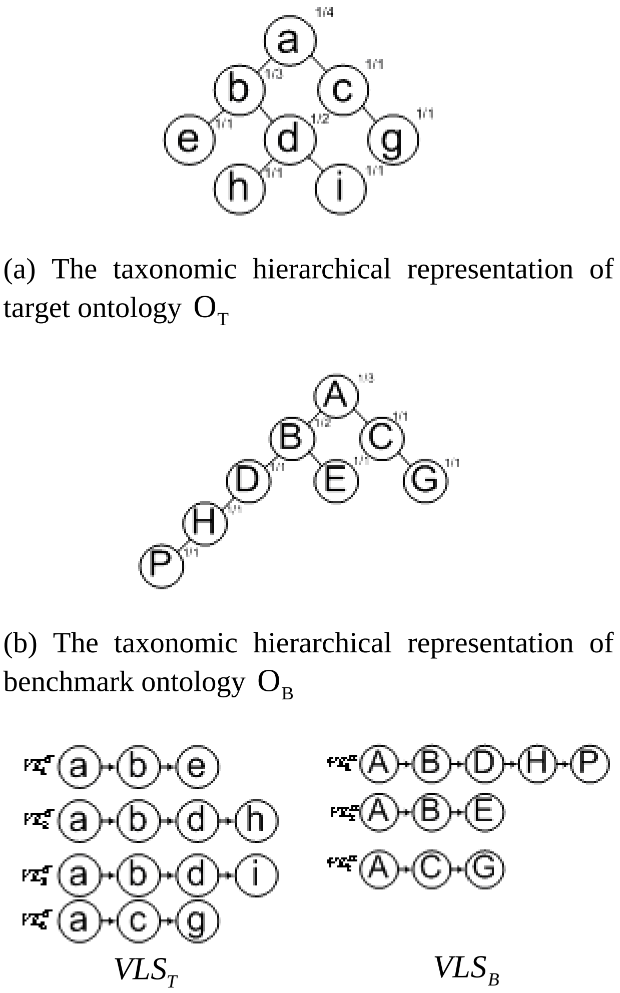
The benchmark ontologies are created to be the test-suites of reusable data which can be employed by ontology engineers or constructer for benchmarking purposes. The benchmark ontology was constructed by the domain experts including two doctors and one pharmacologist based on UMLS. The domain experts have integrated the Chinese concepts without changing the contents of UMLS

Evaluation of ontology construction adopted the two layer measures: Lexical and Conceptual layers (Eichmann et al. 1998). The evaluation in the conceptual layer seems to be more important than that in the lexical layer when the ontology is constructed by aligning or merging several well defined source ontologies. There are two conceptual relation categories for evaluation: Taxonomic and non-Taxonomic evaluations.

3.1.1 Evaluation of the taxonomic relation

Step1 Linearization: This step decomposes the tree structure into the vertex list as described in Section 2.2. The ontology, $\mathrm{O_T}$, and the benchmark, $\mathrm{O_B}$ are shown in the Figure 4(a) and 4(b), respectively. After this linearization, the vertex list sets: VLS_T and VLS_B are obtained as shown in Figure 4(c), where $VLS_T = \{VL_1^T, VL_2^T, VL_3^T, VL_4^T\}$ and

$$VLS_B = \{VL_1^B, VL_2^B, VL_3^B\}.$$



(a) The taxonomic hierarchical representation of target ontology $\mathrm{O_T}$

(b) The taxonomic hierarchical representation of benchmark ontology $\mathrm{O_B}$

(c) The taxonomic vertex list set representation of target ontology and benchmark ontology

Figure 4 Linearization of ontologies

Step 2 Normalization: Since the frequencies of concepts in the vertex list set are not equal, the normalization factors are introduced to address this problem. For the target ontology, the factor vectors for normalization is

$NF^T = \left\{ nf_1^T, nf_2^T, nf_3^T, nf_4^T, nf_5^T, nf_6^T, nf_7^T, nf_8^T \right\}$,

and for the benchmark ontology is

$NF^B = \left\{ nf_1^B, nf_2^B, nf_3^B, nf_4^B, nf_5^B, nf_6^B, nf_7^B, nf_8^B, nf_9^B \right\}$

where nf_i^o is the normalization factor for the i-th concept of the ontology O. It is defined as the reciprocal of the frequency in the vertex list set.

$$nf_i^O = \frac{1}{\left|\text{the vertex lists contain the concept}_i \text{ in ontology O}\right|}$$

Step 3 Estimation of the vertex list similarity: Therefore, the pairwise similarity of these two vertex lists of the target ontology and benchmark ontology can be obtained using the Needleman/Wunsch techniques shown in the following steps:

Initialization: Create a matrix with m+1 columns and n+1 rows. m and n are the numbers of the concepts in the vertex lists of the target ontology and the bench mark ontology, respectively. The first row and first column of the matrix can be initially set to 0. That is,

$$Sim(m,n) = 0,\ if\ \ \mathrm{m} = 0\ or\ \mathrm{n} = 0 \tag{6}$$

Matrix filling: Assign the values to the remnant elements in the matrix as the following equation:

$$Sim(V_m^{T_i}, V_n^{B_j}) = \max \begin{cases} Sim(m-1,n-1) + \frac{1}{2}\left(nf_{m-1}^{T_i} + nf_{n-1}^{B_j}\right) \times Sim_{lexicon}(V_{m-1}^{T_i}, V_{n-1}^{B_j}), \\ Sim(m-1,n)) + \frac{1}{2}\left(nf_{m-1}^{T_i} + nf_{n}^{B_j}\right) \times Sim_{lexicon}(V_{m-1}^{T_i}, V_{n}^{B_j}), \\ Sim(m,n-1) + \frac{1}{2}\left(nf_{m}^{T_i} + nf_{n-1}^{B_j}\right) \times Sim_{lexicon}(V_{m}^{T_i}, V_{n-1}^{B_j}) \end{cases} \tag{7}$$

There are some synonyms belonging to the same concept represented in one vertex. So the lexicon similarity can be described as

$$Sim_{lexicon}(V_{m-1}^{T_i}, V_n^{B_j}) = \frac{\left|\text{Synonyms defined in the } V_{m-1}^{T_i} \text{ and } V_n^{B_j}\right|}{\left|\text{ Synonyms defined in the } V_{m-1}^{T_i} \text{ or } V_n^{B_j}\right|} \tag{8}$$

Traceback: Determine the actual alignment with the maximum score, $Sim(V_m^{T_i}, V_n^{B_j})$, and therefore the pairwise similarity will be defined as the following equation:

$$Sim\left(VL_i^T, VL_j^B\right) \equiv \arg\max Sim(V_m^{T_i}, V_n^{B_j}) \tag{9}$$

Step 4 Pairwise similarity matrix estimation: The pairwise similarity matrix is obtained after $p \times q$ times for Step3. p ,q are the numbers of the vertex list of target ontology and benchmark ontology. Each element of the pairwise similarity matrix as Equation (10) is obtained using Equation (9).

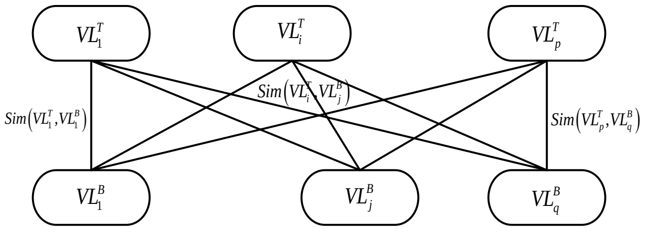

VL_1^T
VL_i^T
VL_p^T
$Sim\left(VL_i^T, VL_j^B\right)$
$Sim\left(VL_1^T, VL_1^B\right)$
$Sim\left(VL_p^T, VL_q^B\right)$
VL_1^B
VL_j^B
VL_q^B

Figure 5 Pairwise similarity between the target ontolgy and benchmark ontology

$$PSM\left(O_T, O_B\right) \equiv \begin{bmatrix} Sim\left(VL_1^T, VL_1^B\right) & \cdots & Sim\left(VL_1^T, VL_q^B\right) \\ \vdots & \ddots & \vdots \\ Sim\left(VL_p^T, VL_1^B\right) & \cdots & Sim\left(VL_p^T, VL_q^B\right) \end{bmatrix}_{p \times q} \qquad (10)$$

Step 5 Evaluation of the taxonomic hierarchy: The whole similarity between target ontology and benchmark ontology can be represented as:

$$Sim_{taxonomic}\left(O_T, O_B\right) = \frac{1}{p}\sum_{i=1}^{p} \underset{1 \le j \le q}{\operatorname{argmax}}\, Sim\left(VL_i^T, VL_j^B\right) \qquad (11)$$

3.1.2 Evaluation of the non-taxonomic relation

Some relations defined in the ontology are non-taxonomic set such as synonym. In fact, the lexicon similarity is applied to measure the conceptual similarity. The lexicon similarity of set can be defined as the following equation:

$$Sim_{lexicon}(V_s^{T_i}, V_t^{B_j}) = \frac{\left|\text{Words defined in the } V_s^{T_i} \text{ and } V_t^{B_j}\right|}{\left|\text{Words defined in the } V_s^{T_i} \text{ or } V_t^{B_j}\right|} \qquad (12)$$

Therefore, the evaluation of the non-taxonomic relation is defined as

$$Sim_{non-taxonomic}\left(O_T, O_B\right) = \frac{1}{p \times q}\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{s}\sum_{t} Sim_{lexicon}(V_s^{T_i}, V_t^{B_j}) \qquad (13)$$

3.1.3 Evaluation Results

Using the benchmark ontology and evaluation metrics described in previous sections, the evaluation results are shown in Table 1.

Table1 the similarity measure between the target ontology and benchmark ontology

Taxonomic relation similarity	0.57
Non-Taxonomic relation similarity	0.68

According to the experimental results, some phenomena are discovered as follows: first, the number of words mapped to the same concept in the upper layer of ontology is larger than that in the lower layer because the terminologies usually appear in the lower layer.

3.2 Evaluation of domain application

To assess the ontology performance, a medical web-mining system to search the desired page has been implemented. In this system the web pages were collected from several Websites and totally 2322 web pages for medical domain and 8133 web pages for contrastive domain were collected. The training and test queries for training and evaluating the system performance were also collected. Forty users, who do not take part in the system development, were asked to provide a set of queries given the collected web pages. After post-processing, the duplicate queries and the queries out of the medical domain are removed. Finally, 3207 test queries using natural language were obtained.

The baseline system is based on the Vector-Space Model (VSM) and synonym expansion. The conceptual relations and axioms defined in the medical ontology are integrated into the baseline as the ontology-based system. The result is shown in Table 2. The results show that ontology-based system outperforms the baseline system with synonym expansion, especially in recall rate.

4 Conclusion

A novel approach to automated ontology alignment and domain ontology extraction from two knowledge bases is presented in this paper. In this approach, a bilingual ontology is developed from two well established language-dependent knowledge bases, WordNet and HowNet according to the co-occurrence of the words in the parallel bilingual corpus. A domain-dependent ontology is further extracted from the universal ontology using the island-driven algorithm and a domain and its contrastive corpus. In addition, domain-specific terms and axioms are also added to the domain ontology. This paper also proposed an evaluation method, benchmark and metrics, for ontology construction. Besides, we also applied the domain-specific ontology to the web page search in medical domain. The experimental results show that the proposed approach outperformed the synonym expansion approach. The overall performance of the information retrieval system is directly related to the ontology.

Table 2 Precision rate (%) at the 11 points recall level

Recall Level	**0**	**.1**	**.2**	**.3**	**.4**	**.5**	**.6**	**.7**	**.8**	**.9**	**1**
Baseline system	78	73	68	65	60	52	38	30	21	15	11
Ontology based	87	86	82	77	73	71	68	62	51	40	32

References

N. Asanoma, 2001. Alignment of Ontologies: WordNet and Goi-Taikei. WordNet and Other Lexical Resources Workshop Program, NAACL2001. 89-94

D. Eichmann, M. Ruiz, and P. Srinivasan, 1998. Cross-language information retrieval with the UMLS Metathesaurus, Proceeding of ACM Special Interest Group on Information Retreival (SIGIR), ACM Press, NY (1998), 72-80.

D. Fensel, C. Bussler, Y. Ding, v. Kartseva1, M. Klein, M. Korotkiy, B. Omelayenko and R. Siebes, 2002. Semantic Web Application Areas, the 7th International Workshop on Applications of Natural Language to Information Systems (NLDB02).

F. C. Fellbaum, 1998. WordNet an electronic Lexical Database, The MIT Press 1998. pp307-308

A. Gómez-Pérez, 2001. Evaluating ontologies: Cases of Study IEEE Intelligent Systems and their Applications: Special Issue on Verification and Validation of ontologies. Vol. 16, Number 3. March 2001. Pags: 391-409.

I. Langkilde and K. Knight, 1998. Generation that Exploits Corpus-Based Statistical Knowledge. In Proceedings of COLING-ACL 1998.

V. Levensthein, 1966. Binary codes capable of correcting deletions, insertions, and reversals. Soviet Physics–Doklady, 10:707–710.

A. Maedche, and S. Staab, 2002. Measuring Similarities between Ontologies. In Proceedings of the 13th European Conference on Knowledge Engineering and Knowledge Management EKAW, Madrid, Spain 2002/10/04

M. Missikoff,, R. Navigli, and P. Velardi, 2002. Integrated approach to Web ontology learning and engineering, Computer, Volume: 35 Issue: 11 . 60 –63

N. F. Noy, and M. Musen, 2000. PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment, Proceedings of the National Conference on Artificial Intelligence. AAAI2000. 450-455

S. Oyama, T. Kokubo, and T. Ishida, 2004. Domain-Specific Web Search with Keyword Spice. IEEE Transactions on Knowledge and Data Engineering, Vol 16,NO. 1, 17-27.

Sinorama Magazine and Wordpedia.com Co., 2001. Multimedia CD-ROMs of Sinorama from 1976 to 2000, Taipei.

P. Vossen, and W. Peters, 1997. Multilingual design of EuroWordNet, Proceedings of the Delos workshop on Cross-language Information Retrieval.

H. Weigard, and S. Hoppenbrouwers, 1998. Experiences with a multilingual ontology-based lexicon for news filtering, Proceedings in the 9th International Workshop on Database and Expert Systems Applications. 160-165