Using Semantic Structure to Improve Chinese-English Term Translation

Guiping Zhang[1], Ruiqian Liu[1], Na Ye[1], and Haihong Huang[2]

1 Knowledge Engineering Research Center, Shenyang Aerospace University, Shenyang, China;
2 Chinese COMAC Shanghai Aircraft Design and Research Institute, Shanghai, China;
zgp@ge-soft.com, liuruiqian99@163.com
yena_1@126.com, Huanghaihong@comac.cc

Abstract. This paper introduces a method which aims at translating Chinese terms into English. Our motivation is providing deep semantic-level information for term translation through analyzing the semantic structure of terms. Using the contextual information in the term and the first sememe of each word in HowNet as features, we trained a Support Vector Machine (SVM) model to identify the dependencies among words in a term. Then a Conditional Random Field (CRF) model is trained to mark semantic relations for term dependencies. During translation, the semantic relations within the Chinese terms are identified and three features based on semantic structure are integrated into the phrase-based statistical machine translation system. Experimental results show that the proposed method achieves 1.58 BLEU points improvement in comparison with the baseline system.

Keywords: dependency analysis; semantic analysis; term translation;

1 Introduction

With the rapid development of science and technology, international technical exchanges are more and more frequent. As the carrier of scientific and technological concepts, the translation of technical terms has attracted much attention. It can not only be applied to cross-language information retrieval, but also to bilingual dictionary compilation.

Some researchers use internet or comparable corpora to mine bilingual terms[1, 2]. These methods cannot deal with out-of-vocabulary (OOV) terms, which is the focus of this paper. Previous work on OOV term translation mainly makes use of the characteristics of terms to extract some new features and combine them into the traditional statistical machine translation (SMT) framework. The new features include the character similarity between Japanese and Chinese[3], the part-of-speech (POS) sequence similarity between Chinese and Korean[4], the morphological correspondence between English and Chinese[5], the phonic correspondence between English and Japanese katakana[6].

We can see that current approaches mainly take advantage of bilingual linguistic information to improve term translation accuracy, and the semantic relations within the term are seldom considered. However, such semantic information plays an important role for lexical reordering and word selection in term translation. For example, "雨伞(umbrella) 自动(automatic) 装袋(bagging) 机(machine)" is translated into "automatic umbrella bagging machine" by the NiuTrans SMT system[7]. But if we know that "自动(automatic)" depends on "装袋(bagging)", then the words in the translation should be reordered and adopts the preposition structure as "automatic bagging machine for umbrella". Another example is "切削(cutting) 工具(tool)". This term is translated into "cut tools" by NiuTrans, but through semantic analysis we can see that the semantic relationship between "切削(cutting)" and "工具(tool)" is property-host instead of patient-event. So this term should be translated into "cutting tools".

With the development of machine translation, semantic relations have been used as an additional feature to improve the translation performance[8]. However, there are too many types of semantic relationships in the translation of sentences and it is difficult to define them in a uniform framework. In contrast, the semantic relationship types within terms are much less and more specific. Besides, the head word of a term is always the last word, which also reduced the difficulty of analyzing terms. Therefore the dependency and semantic analysis result of terms can be more accurate than sentences, and the term semantic structure will be helpful for term translation.

In this paper, we propose a Chinese-English term translation method based on the term semantic analysis. First we use words, part of speech (POS) tags, word distances, word contexts and the first sememe of a word in HowNet[9] as features to train a dependency analysis model by SVM. The model is used to identify dependencies embedded inside a term. A CRF model is used afterwards to incorporate the dependencies and acquire the semantic structure of the Chinese term. Next, three semantic-based features (lexical reordering feature, word selection feature and POS selection feature) are extracted and integrated into the phrase-based SMT translation system. Experimental results show that our method is effective.

This paper is organized as follows: in section 2 we give the framework of the Chinese-English term translation system. In section 3 we introduce the Chinese term semantic analysis method. In section 4 the three features based on semantic structure for term translation are described. Experimental results are shown in section 5. Finally, we draw conclusions in section 6.

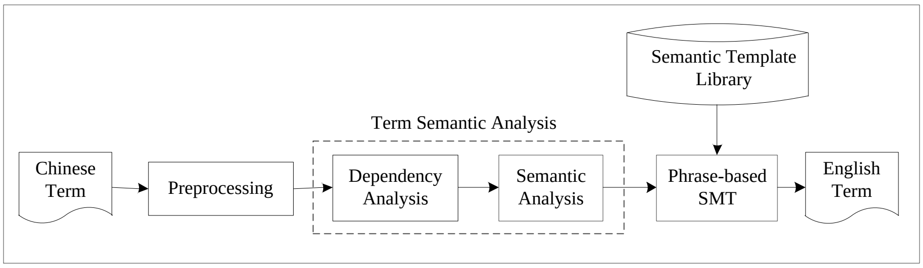

Fig.1. The framework of the Chinese-English term translation system

2 System Framework

The framework of the Chinese-English term translation system is shown in Figure 1. In the pre-processing stage, Chinese word segmentation and POS tagging are performed. Then the labelled term is input to the term semantic analysis module, which is divided into two parts. Firstly, the dependency analysis module identifies the dependency relationships within the term as shown in figure 2.

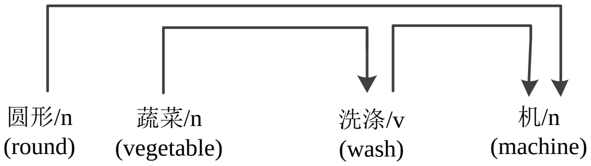

Fig.2. Term dependency analysis result

Secondly, the semantic analysis module identifies the semantic relationship between each two dependent words as shown in Figure 3.

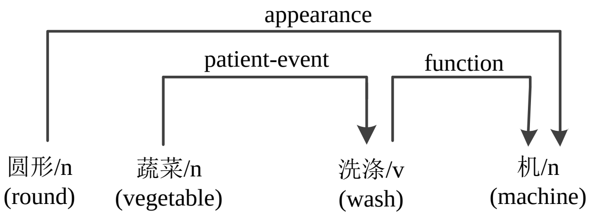

Fig.3. Term semantic analysis result

After semantic analysis, the term is input to a phrase-based SMT module for translation. On the basis of the semantic structure within the term, three additional features for lexical reordering, word selection and POS selection are extracted through matching the semantic template library. Then the new features are integrated into the SMT model together with the traditional machine translation features to achieve the optimal translation.

3 Term Semantic Analysis

In order to use semantic structure to improve Chinese-English term translation, we should analyze the Chinese terms in semantic level. However, existing dependency analysis tools like Stanford parser[1] do not perform well on Chinese terms. This is because Chinese terms have some special characteristics such as (1) there are many out-of-vocabulary words in terms; (2) most terms are noun phrases without head verbs; (3) there is few appropriate corpora designed for terms. Due to the above reasons, a general parser is incapable of dealing those dependencies in a term. In order to solve these problems, we propose a term semantic analysis approach in this paper.

3.1 Dependency Analysis Based on SVM

Through analyzing the characteristics of Chinese terms, we select the following characteristics to train the SVM based dependency analysis model:

(1) Basic features: We use words, POS tags, distance between words and context information as basic features.

(2) Mutual information: Mutual information is used as a measure of correlation degree among words in a term. Total 643,908 terms are computed for mutual information.

(3)The first sememe in HowNet: This feature is generated according to Liu[10], presenting the semantic categories of a word. First sememe of two words can be used to compute their semantic relations.

During dependency analysis, the SVM model is used to analyze each two words in the term and returns a real value. This value indicates the probability that there is dependency between the two words. Therefore, each word w_i in the term is scanned from left to right and the word with highest value returned by SVM is selected as the word that w_i depends on. According to the dependency axiom, there should be no cross dependency, so the algorithm performs backtracking until there is no intersection.

3.2 Semantic Analysis Based on CRF

Through analyzing the characteristics of technical terms, this paper defined 14 types

[1] http://nlp.stanford.edu/software/

of semantic relationships as shown in Table 1 and two syntactic level relationships which include the structure of "之(of)" and the structure of "与(with)". In order to reflect the semantic relationships between words inside a term, we further divide the relationship type "property-host" into seven subtypes, namely "measurement", "appearance", "situation", "nature", "quantity", "category" and "function".

Table 1. Definition of the semantic relationships

No.	Type	No.	Type
1	Agent-Event	8	Usage
2	Patient-Event	9	Negation
3	Property-Host	10	Name
4	Material-Product	11	Location
5	Overall-Part	12	Continuity
6	Suffix	13	Degree
7	Mode	14	Object

We then choose CRF as our tool to analyze term semantics in this paper. The selected features are shown in table 2.

Table 2. Features for dependency analysis

Feature	Description
WD1	modifier
WD2	modified word
POS1	POS of the modifier
POS2	POS of the modified word
WD1_ATOM	first sememe of the modifier
WD2_ATOM	first sememe of the modified word
PRE_WD	word that modifies the modifier
PRE_POS	POS of the word that modifies the modifier
UN/UNKN	whether the modifier is the head

4 Term Translation based on Semantic Structure

Traditional statistical machine translation methods use the source-channel model, which is shown in formula 1. It allows an independent modelling of target language model $Pr(e_1^I)$ and translation model $Pr(f_1^J \mid e_1^I)$.

$$\hat{e}_1^I = \underset{e_1^I}{argmax}\left\{Pr(e_1^I \mid f_1^J)\right\} \qquad (1)$$

$$= \underset{e_1^I}{argmax} \left\{ Pr(e_1^I) \cdot Pr(f_1^J \mid e_1^I) \right\} \tag{2}$$

An alternative to the classical source-channel approach is the direct modeling of the posterior probability $Pr(e_1^I \mid f_1^J)$. Using a log-linear model[11], we obtain formula 3:

$$Pr\left(e_1^J \mid f_1^J\right) = \exp\left(\sum_{m=1}^{M} \lambda_m h_m \left(e_1^I, f_1^J\right)\right) \cdot Z\left(f_1^J\right) \tag{3}$$

Here, $Z\left(f_1^J\right)$ denotes the appropriate normalization constant. As a decision rule, we obtain formula 4:

$$\hat{e}_1^I = \underset{e_1^I}{\arg\max} \left\{ \sum_{m=1}^{M} \lambda_m h_m \left(e_1^I, f_1^J\right) \right\} \tag{4}$$

New features can be added to it freely, so in this paper, we propose three new features based on the semantic structure of the Chinese term to improve the lexical reordering, word selection and POS selection in the term. This section will describe the features in detail.

4.1 Lexical Reordering Feature

This feature aims at modelling the influence of semantic relationships on the order of words within the scope of a phrase. Suppose there is a Chinese phrase *c* and the corresponding English phrase *e*, we calculate the lexical reordering probability between *c* and *e*. This probability is estimated through the semantic relationships of the word pairs. If there are *N* dependent word pairs, then we examine the word alignment and extract the pairs with straight order to construct set R_1 and extract the pairs with inversed order to construct set R_2. And the lexical reordering probability can be computed with formula 5.

$$p_{lex\_od}(c,e) = \begin{cases} \alpha_1 \\ \dfrac{\sum_{i=1}^{|R_1|} p_i(w_1,w_2,r) + \sum_{i=1}^{|R_2|} (1 - p_i(w_1,w_2,r)) + \beta(N - |R_1| - |R_2|)}{N} \end{cases} \tag{5}$$

where w_1 and w_2 are the two dependent words, and $p_\mathrm{i}(w_1,w_2, r)$ is the probability that w_1 and w_2 has straight order when they have a semantic relationship of *r*. The third component in the numerator refers to the number of words whose dependent word is out of the scope of the current phrase. β is a scaling factor and is empirically set to 0.1. If c has only one word, then we set the probability to a constant α_1.

The value of $p_{\mathrm{i}}(w_1,w_2, r)$ is stored in the lexical reordering template as follows:

$$(w_1|w_2)\ \#\ r \qquad p(w_1, w_2, r)$$

We use maximum likelihood estimation to compute $p(w_1, w_2, r)$ as formula 6 shown.

$$p(w_1,w_2,r)=\frac{c_t(w_1,w_2,r)}{c(w_1,w_2,r)} \tag{6}$$

where $c_t(w_1,w_2,r)$ is the frequency that w_1 and w_2 has straight order, and $c(w_1,w_2,r)$ is the total frequency. Through the formula we get the lexical reordering template as shown in figure 4. If the word pair "储物(w_1)|箱(w_2)" has the semantic relationship of "function", then the corresponding English part of the word pair has the probability of 83.81 percent to be in straight order.

（储物|箱）#function	0.8381
（处理|用）#suffix	0.8333
（底盘|修理）#agent-event	0
（服装|尺）#category	0.1667

Fig.4. Examples of the lexical reordering template

4.2 Word Selection Feature

This feature aims at modeling the influence of semantic relationships on word selection. Suppose there is a Chinese word pair $(w_1|w_2)$, we calculate the probability of their corresponding translations $(t_1|t_2)$ if they have the semantic relationship of r. The probability is stored in the word selection template as follows:

$$(w_1|w_2)\ \#\ r\ \#\ (t_1|t_2) \qquad p(w_1, w_2, r)$$

The template shows that if $(w_1|w_2)$ has the semantic relationship r then they have the probability of $\mathrm{p}(w_1, w_2, r)$ to be translated to $(t_1|t_2)$. The probability is also computed by maximum likelihood estimation. We give some examples of word selection template in figure 5.

（储物|箱）# function # (storage|box)	0.721
（储物|箱）# agent-event # (for storing|box)	0.653
（服装|尺）# category # (for clothes|spline)	0.75
（服装|尺）# category # (dress|ruler)	0.25

Fig.5. Examples of the word selection template

According to traditional phrase based statistical machine translation method, the word pair "储物|箱" has higher probability to be translated into "storage|box". However, with the above templates, if we know that the semantic relationship between "储物" and "箱" is "agent-event", then we know that the word pair is more likely to be translated into "for storing|box".

Suppose there is a Chinese phrase *c* and the corresponding English phrase *e*, we calculate the word selection probability between *c* and *e*. This probability is estimated through the word pairs with semantic relationships. If in the Chinese phrase *c*, word pair ($w_1|w_2$) has semantic relationship *r*, and its candidate English translation is ($t_1|t_2$) which matches a word selection template, then we use the probability in the template as the translation probability of the word pair. We add all the *N* probabilities which match a word selection template as the word selection probability of the phrase *c*. Formula 7 shows the calculation. If *c* has only one word, then we set the probability to a constant α_2.

$$p_{word\_sl}(c,e)=\begin{cases}\alpha_2 \\ \dfrac{\sum_{i=1}^{|N|}p(w_1,w_2,r)}{|N|}\end{cases} \tag{7}$$

4.3 POS Selection Feature

Because many words in technical terms do not appear frequently in the corpus, so it may be difficult to match the template if we follow the template style in section 4.2. To solve this data sparseness problem, we use POS sequence instead of the word itself in the POS selection template. Suppose there is a Chinese word pair ($w_1|w_2$), we calculate the probability of their corresponding POS ($POS_1|POS_2$) if they have the semantic relationship of *r*. The probability is stored in the POS selection template as follows:

$$(w_1|w_2)\ \#\ \mathrm{r}\ \#\ (POS_1|POS_2) \qquad p(w_1,\ w_2,\ r)$$

The template shows that if ($w_1|w_2$) has the semantic relationship *r* then they have the probability of $p(w_1,\ w_2,\ r)$ to be translated to ($POS_1|POS_2$). The probability is also computed by maximum likelihood estimation. We give some examples of POS selection template in figure 6.

（储物|箱）# function # (JJ|NN)	0.785
（储物|箱）# function # (NN|NN)	0.215
（储物|箱）# agent-event # (IN NN|NN)	1
（处理|水）# patient-event # (VB|NN)	0.5

Fig.6. Examples of the POS selection template

The POS selection template shows that if "储物|箱" has the semantic relationship of "function" then the probability of its corresponding English POS sequences being "JJ|NN" is 0.785, and "NN|NN" is 0.215.

Suppose in the Chinese phrase *c*, word pair ($w_1|w_2$) has semantic relationship *r*. If the candidate English translation has a POS sequence ($POS_1|POS_2$) and matches a POS selection template, then we use the probability in the template as the translation probability of the word pair. We add all the *N* probabilities which match a POS selection template as the POS selection probability of the phrase *c*. Formula 8 shows the calculation. If *c* has only one word, then we set the probability to a constant α_3 .

$$p_{POS\_sl}(c,e)=\begin{cases}\alpha_3\\ \frac{\sum_{i=1}^{|N|}p(w_1,w_2,r)}{|N|}\end{cases} \tag{8}$$

This paper used 452,781 Chinese-English terms from the patent titles given by State Patent Office of China to extract the templates. Our trained SVM model and CRF model are applied to perform semantic analysis on these terms. The Stanford postagg[2] is applied for POS tagging and GIZA++ is applied for word alignment.

5 Experimental Results and Analysis

In this section, we describe the experiments which we carried out to test the performance of the improvements presented in the previous sections.

5.1 Data Setup

A. Data Setup for Term Semantic Analysis

Our experiments are carried out on 642,908 terms extracted from patent documents. The term dependency analysis model and the semantic analysis model are trained with 3000 manually labelled terms and we choose 238 terms as a testing corpus. Each term has an average length of 5.07 words.

B. Data Setup for Term Translation

Our term translation experiments are carried out on the 451,500 terms mentioned in the above section. Table 3 shows some statistical characteristics of the corpus for term translation.

[2] http://nlp.stanford.edu/software/

We use the IRLAS[3] tool to perform Chinese word segmentation. The English terms are tokenized, lowercased and POS tagged by the Stanford tool. The GIZA++ tool is used to perform bilingual word alignment, and NiuTrans is taken as the baseline system.

Table 3. The characteristics of the corpus

		Ch	En
Train	Term pairs	450000	
	Ave length	4.262	4.829
Dev	Term pairs	500	
	Ave length	4.250	4.896
	Perplexity	465.06	439.22
Test	Term pairs	1000	
	Ave length	4.318	4.931
	Perplexity	392.52	378.05

5.2 Results

A. The result of semantic analysis.

We build two baseline systems for comparison. In the first system, all words depend on the head word in the term. In the second system, all words depend on the right nearest neighbour. Table 4 gives the performances of the baseline systems and our systems with "basic feature"(system1), "basic feature + mutual information" (system2)and "basic feature + mutual information + first sememe in HowNet" (system3). Word pair accuracy (p_{wpa}) and term accuracy (p_{ta}) are used to evaluate the systems' performances. Word pair accuracy is calculated by formula 9, and term accuracy by formula 10. The results are shown in table 4.

$$p_{wpa} = \frac{c_{wr}}{c_w} \times 100\% \qquad (9)$$

$$p_{ta} = \frac{c_{tr}}{c_t} \times 100\% \qquad (10)$$

[3] http://www.ir.hit.edu.cn

where c_{wr} is the number of correct arcs, c_w is the total number of arcs, c_{tr} is the number of correct terms and c_t is the total number of terms.

Table 4. The result of dependency and semantic analysis

	Word pair accuracy		Term accuracy
Dependency analysis	baseline1	33.67%	17.84%
	baseline2	71.78%	50.41%
	system1	81.36%	51.51%
	system2	82.55%	55.89%
	system3	***87.85%***	***65.65%***
Semantic analysis	***76.98%***		***59.20%***

It can be seen from the experimental results that the word accuracy of baseline2 is 38.11% higher than baseline1, which indicates that the probability of dependency between nearest words is better than that between a word and its head word. On the other hand, the word accuracy and term accuracy of system2 increase by 1.19% and 4.38% over system1 respectively. Therefore mutual information is a good measure to identify the strength of association between two words. However, due to the large number of unknown words, the effect is not obvious. After adding the feature of first sememe in HowNet, the word pair accuracy and term accuracy have increased 5.3% and 9.76% respectively, which shows that the interdependence between two words highly depends on their semantic. We can also see that the accuracy of word pairs is nearly 80%, and the lexical reordering templates, word selection templates and POS selection templates are extracted on word pair level, which can guarantee the accuracy of the templates.

B. The result of term translation.

Firstly, we will determine the values of α_1, α_2 and α_3 mentioned in section 4. Taking the process of deciding the value of α_1 in formula 4-5 as an example, we made experiments with different values of α_1 varying from 0 to 1 on the development set, and the result is shown in figure 7.

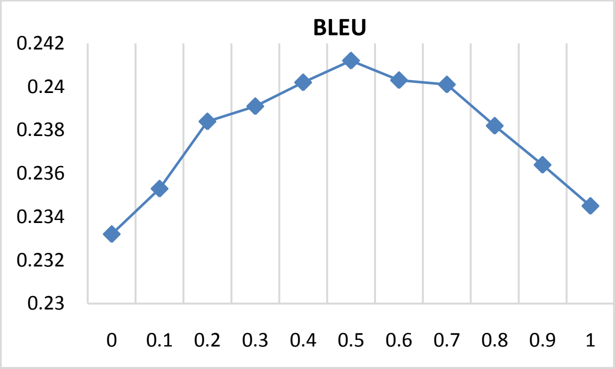

Fig.7. The BLEU scores with different α

So we set α_1 to 0.5 in the following experiments. The other parameters are decided by the same method. α_2 is set to 0.2, and α_3 is set to 0.5.

To evaluate the performance of Chinese-English term translation, we build six systems: NiuTrans(baseline), "NiuTrans + lexical reordering feature "(system1), "NiuTrans + word selection feature"(system2), "NiuTrans + POS selection feature"(system3), and "NiuTrans + lexical reordering feature + POS selection feature"(system4), and "NiuTrans + lexical reordering feature + word selection feature + POS selection feature"(system5).We use BLEU and NIST for evaluation. Experimental results are shown in table 5. In order to describe the experiment more clearly, we use f_1, f_2 and f_3 to represent the lexical reordering feature, word selection feature and POS selection feature.

Table 5. The result of term translation

	BLEU	NIST
Baseline	0.2324	5.7427
system1(baseline+f_1)	0.2409	5.7823
system2(baseline+f_2)	0.2356	5.7448
system3(baseline+f_3)	0.2400	5.7811
system4(baseline+f_1+f_3)	0.2436	5.7846
system5(baseline+f_1+f_2+f_3)	***0.2482***	***5.7963***

It can be seen from the results that compared with the baseline system, using the lexical reordering feature, the BLEU score increased by 0.85 percent. However, adding the word selection feature makes the BLEU score improved by 0.32 percent. But using the POS selection feature, the BLEU score improves 0.76 percent, because the

POS templates are easier to be matched than the word templates. So we add the word selection feature and lexical reordering feature together, and the system performance is improved by 1.12 percent. Finally we add the three features all together, and the BLEU score increased by 1.58 percent. This proves that semantic information plays a positive role in term translation.

For example, "蛋糕音乐装饰卡"(music decorative card for cake) is translated into "cake music decorative card" by the baseline system. But according to the lexical reordering templates, when the semantic relationship between word pair (蛋糕|卡) is "category", the probability of inversed order is higher than that of straight order. Therefore after adding the semantic features it is translated into "music decorative card for cake". Another example, "新型装饰灯"(novel decorative light) is translated into "novel decorate light" by the baseline system. But according to the POS selection templates, when the semantic relationship between word pair (装饰|灯) is "category", the probability of POS sequence "JJ NN" is higher than that of "VB NN", therefore the system after adding semantic features translated it into "novel decorative light".

C. Error Analysis .

By analyzing the experimental results of semantic analysis, we find that the main reasons that lead to the errors in semantic analysis are: (1) word segmentation error; (2) many words in terms are quite difficult to understand even for human beings. Moreover, the boundaries between some semantic relationships are not clear.

We also find some reasons that effect the accuracy of term translation: (1) the semantic analysis errors which are produced by the above work; (2) the types of semantic relationships we defined in this paper cannot cover all the internal relationships of terms.

6 Conclusion

In this paper, we presented a Chinese-English term translation method based on semantic structure. We use a SVM model with features of mutual information and the first sememe in Hownet for dependency analysis, and then use a CRF model for semantic analysis. Then we extracted three features on the basis of term semantic structure and integrated them into the phrase-based statistical machine translation framework. Experimental showed the effectiveness of the semantic analysis method as well as the term translation method, which illustrates that the internal semantic structure of terms is important information for term translation.

Acknowledgments

This paper is funded by National Key Technology R&D Program of China (2012BAH14F05).

References.

1. Cao, Y., Li, H.: Base noun phrase translation using web data and the EM algorithm. In Proceedings of the 19th international conference on Computational linguistics-Volume 1. pp. 1-7. Association for Computational Linguistics (2002)
2. Fang, G., Yu, H., Nishino, F.: Chinese-English term translation mining based on semantic prediction. In Proceedings of the COLING/ACL on Main conference poster sessions. pp. 199-206. Association for Computational Linguistics (2006)
3. Wang, J., Zhang, G., Ye, N., Zhou, L.: Research on Japanese-Chinese Term Translation Technique Based on Multi-Features. In Pattern Recognition, 2009. CCPR 2009. Chinese Conference on IEEE. pp. 1-5. (2009)
4. Kang, B. K., Chen, Y. R., Chang, B. B., Yu, S. W.: Translating multi word terms into Korean from Chinese documents. In Natural Language Processing and Knowledge Engineering, 2005. IEEE NLP-KE'05. Proceedings of 2005 IEEE International Conference on IEEE. pp. 449-454. (2005)
5. Wu, X., Okazaki, N., Tsunakawa, T., Tsujii, J. I.: Improving English-to-Chinese translation for technical terms using morphological information. In AMTA-2008. MT at work: Proceedings of the Eighth Conference of the Association for Machine Translation in the Americas. pp. 202-211. (2008)
6. Tsuji, K.: Automatic extraction of translational Japanese-KATAKANA and English word pairs from bilingual corpora. International Journal of Computer Processing of Oriental Languages, 15(03), 261-279 (2002)
7. Xiao, T., Zhu, J., Zhang, H., & Li, Q.: NiuTrans: an open source toolkit for phrase-based and syntax-based machine translation. In Proceedings of the ACL 2012 System Demonstrations. pp. 19-24. Association for Computational Linguistics (2012)
8. Beale, S., Nirenburg, S., Mahesh, K.: Semantic analysis in the Mikrokosmos machine translation project. In Proceedings of the 2nd Symposium on Natural Language Processing . pp. 297-307. (1995)
9. Dong, Z., Dong, Q.: HowNet and the Computation of Meaning . pp. 1-316. Singapore: World Scientific (2006)
10. Liu, Q., Li, S.: Word similarity computing based on How-net. Computational Linguistics and Chinese Language Processing, 7(2), 59-76 (2002)
11. Och, F. J., Ney, H.: Discriminative training and maximum entropy models for statistical machine translation. In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. pp. 295-302. Association for Computational Linguistics (2002)