Automatic Mining of Internet Translation Reference Knowledge Based on Multiple Search Engines

Baosheng Yin, Wei Wang, Ruixue Lu, Yang Yang

***Abstract*—With the increasing amount of bilingual resources on the Internet, translators increasingly solve various problems in the translation process by using Internet as a knowledge source. In order to improve translators' efficiency and effect in retrieving Internet translation reference knowledge, this paper provides an algorithm to automatically obtain the translation reference knowledge from Internet. By analyzing the retrieval logs which are generated by translators in translation process, we find that "Unit to be translated + heuristic word" is the most frequent and effective retrieval strategy used by translators. The proposed algorithm contains: generation and sorting of heuristic words, combination retrieval strategy, merging and ranking of the retrieval results from multiple search engines. Experimental results show that the recommendation precisions of TOP1, TOP3 and TOP5 achieve 53.5%, 69% and 77.5% respectively. The proposed method can significantly improve the efficiency of computer aided translation.**

***Index Terms*—knowledge-mining, search-engine, internet, computer-aided-translation.**

I. Introduction

THE translation competence of a translator not only depends on the ability of language, but also depends on the ability of utilizing translation tools and translation resources. With the development of Internet technology, network resource data is becoming an enormous free translation corpus that is updated in real time. The translators tend to utilize Internet to aid translation when they encounters difficult vocabulary. How to efficiently and accurately obtain the translation reference information implied in Internet and closely related to current translation requirement will be so important.

Now, there are three main assistant translation means based on Internet: online dictionary translation, online automatic machine translation and assistant translation based on Internet retrieval. Relevant terminology knowledge can be retrieved more rapidly by busing online dictionary than using paper dictionary. Thereby, the translator may save a lot of consultation time, but the on-line translation dictionary has not yet provided sufficient context translation information. When faced with the selection for numerous translation items, the translators would often be unable to make a decision, so its assistant effect is limited relatively. Online automatic translation system, such as Google translation etc., is subjected to the development level of machine translation technology, and the quality of translation is often unsatisfactory, which is still far from practical application. A great deal of bilingual information implied in multilingual official website, translation forum and translation community website on the Internet can be retrieved and applied by using network search engines. These information have dynamic performance and includes a large number of bilingual context information.

Relevant research on the retrieval, collecting, mining and utilization of translation resources of Internet has already been carried out successively. In the current research, some researches focus on locating and collecting of the Internet bilingual resources. For example, the method based on adaptive pattern learning was proposed by Microsoft effectively solves the problem of extracting the parallel sentence pair from mixed bilingual webpage [1]. Lin attempted to apply downloading strategy to find the single bilingual webpage, and used the mutual translation information to confirm webpage, and the parallel sentence pairs were be extracted according to length, dictionary, punctuation and number, abbreviation, etc. [2]. Zhu utilized webpage text extraction, keyword extraction and cross-language text similarity to design a mining system of parallel corpus and comparable corpus, which enhanced the utilization rate of Internet resource and the quality of Internet bilingual corpus mining [3], [4]. Liu put forward the identification and purification algorithm of webpage bilingual text [5], and also put forward matching of the webpage bilingual sentence pair and corpus mining algorithm to build a massive corpus [6]. Zhang used multiple features to identify parallel texts via a KNN classifier [7]. Dong, et al. provided the extraction method of terminology translation pair based on the glossary webpage [8], however, this method excessively depends on the glossary webpage, and cannot solve the problem that translation extraction pattern is matched faintly. Luo proposed a method to mining web bilingual resources based on frequent sequence pattern [9]. Guo, et al. put forward a method for automatic mining of the bilingual translation pair present in website by utilizing the features of bilingual translation pair in webpage, using the statistical discriminative model and multiple recognition features [10]. Zhang proposed the verification model of end

Manuscript received July 10, 2014, revised August 7, 2014. This work was supported in part by National Key Technology R&D Program of China(2012BAH14F05)

Baosheng Yin is with Human-Computer Intelligence Research Center, Shenyang Aerospace University, Shenyang, 110136, China, e-mail: (ybs@ge-soft.com)

Baosheng Yin, Wei Wang, Ruixue Lu, Yang Yang are with the Knowledge Engineering Research Center, Shenyang Aerospace University, Shenyang 110136, China

ISBN: 978-988-19252-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

analogy alignment verification, the bilingual alignment verification and the morphology verification, which could effectively verify the Web page terminology translation [11]. Cao et al. and Lin et al. proposed two different methods to extract term translations based on the observation that authors of many bilingual web pages, especially those whose primary language is Chinese, Japanese or Korean, sometimes annotate terms with their English translations inside a pair of parentheses [12], [13].

The researches above mainly focused on the identification, collecting of Internet bilingual resources and establishment of knowledge base, and have not yet met the requirements of the translators for the real-time application of Internet dynamic information. In interactive retrieval assistant translation research based on Internet, Wang explained by way of examples how professional translators make use of the Internet resources such as online dictionaries, translation websites, etc. and verified that internet resources could improve translation effectively [14]. Zhu summarized 10 tips forward word searching to assist translation based on years of translation practice using search engine forward word search on the Internet [15]. Zhou provided the basic steps that the translators utilize Internet search engine to carry out manual retrieval for assistant translation [16], but the automation achievement of the assistant translation based on Internet is not reflected, it requires constant manual attempt to obtain more accurate translation, and is time consuming and effortful.

To enhance the efficiency and effect of translators in real-time retrieval for translation knowledge from the Internet in the translation process, and based on the analysis of manual network search strategy of translators, this paper proposes an automatic mining algorithm of translation reference knowledge which is built on multiple search engines.

II. Analysis of Internet Retrieval Logs of Translators

A. Analysis of Internet Retrieval Content of Translators

This paper collected and analyzed Internet retrieval content of nearly 30 translators of GE-SOFT Translation Service Center (GE-SOFT.com) in the practical translation process. More than 14,000 Internet retrieval logs sampled randomly there from were manually analyzed, including retrieval contents, search engine address and retrieval time. All difficult terms and phrases input by translators in retrieval process could be collectively called the unit to be translated. The translation difficulty during translation can be summarized as follows:

1) Proper Noun.

Generally, translators may encounter some proper nouns in translation process, such as person name, place name, institution name, but the current dictionary often could not provide accurate and available translation. For instance, “中国商用飞机有限责任公司(Commercial Aircraft Corporation of China Ltd, COMAC)”, the correct translation cannot be obtained in the network automatic translation system and the online dictionary, and translators require to retrieve the accurate name in official website.

2)Professional Terminology.

Most of frequently-used vocabularies applied to the professional field may have their conventional translation, thus, the context is required to determine whether it is the usual translation or specific territoriality translation. For instance, “Acknowledgement” can be translated into “领知(Acknowledgement)” in aviation field, but it may be translated into “认知(Cognition)” if the translator gets the meaning from on-line dictionary while he does not understand the knowledge of aviation field, whereas, the accurate translation cannot be obtained.

3) Abbreviation.

Meanings of the same abbreviation in different professional fields are often different enormously. The difficulty of translation is increasing if the translators are unfamiliar with the specialized vocabulary. For example, the correct translation of GA in aviation field literature is “通用航空(General Aviation)”, but other translations can be found in on-line dictionary, such as “总代理人(general agent)”, “遗传算法(Genetic algorithm)”, “赤霉素(Gibberellin acid)” and so on. All these translations are independent of aviation field. This shows that dictionary are unable to completely include all the abbreviations in different professional fields.

4) Newly-Emerging Vocabulary.

With the development of social networking website, newly-emerging vocabulary also rapidly appears and is used widely. On-line dictionary also gradually collects new vocabulary, but it is still far from their update speed. For example, ROM brain means that a person is “死脑筋(One-Track Mind)”, but it may be translated for "Luo's brain" in on-line dictionary, so the effect is not ideal, which may influence the quality and efficiency of translators.

B. Analysis of Internet Retrieval Strategy of Translators

In addition to the analysis of retrieval contents of translators, this paper also particularly analyzed the manual Internet retrieval strategy of translators. To accurately retrieve the reference knowledge required for translation, translators no longer only use the unit to be translated to retrieve the translation knowledge in retrieval process, but also take different combination strategies to attempt retrieval repeatedly. The results of statistical analysis of retrieval strategy of translators are shown in Table I, and the strategy that the unit to be translated is directly used for retrieval is not included in this statistical results.

In practical translation process, when the unit to be translated is directly retrieved by means of search engine, the translation reference information obtained is often little, but introducing "Heuristic Word" for combined retrieval enables translators to obtain more. For example, when “连续增量调制 (Continuous Delta Modulation, CDM)” is directly retrieved on Baidu, there are one pair of unit to be translated and translation, but when retrieved by “连续增量调制 +Continuous”, six translation pairs can be obtained. What kind of heuristic word can help translators discover the translation information more rapidly requires constant attempts. The search results of "安装孔(Mounting Hole)" obtained from Baidu search engine with different retrieval strategies are shown in Table II:

ISBN: 978-988-19252-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

TABLE I
ANALYSIS OF MANUAL RETRIEVAL STRATEGY

Serial Number	User Retrieval Pattern	Example[1]	Number	Proportion
1	Unit to be translated + heuristic word	中文信息学报 +information	4,789	80.8%
2	Unit to be translated “什么意思”	中文信息学报 什么意思	758	12.8%
3	Unit to be translated “英文”	中文信息学报 英文	243	4.1%
4	Unit to be translated “英语怎么说”	中文信息学报 英语怎么说	140	2.3%

Note 1 : “中文信息学报” means “Journal of Chinese Information Processing”.
“什么意思” means “what's meaning”. "英文” means “english”.
“英语怎么说” means “how to say in english”.

TABLE II
ALL KINDS OF RETRIEVAL STRATEGIES

Retrieval Strategy	Translation Retrieval Record	Number of Translation of the Unit to be Translated in the First 30 Returned Webpage Abstracts
Unit to be translated	安装孔	2
Unit to be translated "英语"	安装孔 英语	3
Unit to be translated "英语怎么说"	安装孔 英语怎么说	3
Unit to be translated "什么意思"	安装孔 什么意思	1
Unit to be translated + heuristic word	安装孔 +hole	13
	安装孔 +install	4
	安装孔 +mout	2

A conclusion is drawn from analysis of Table I and II that "Unit to be translated + heuristic word" is the retrieval strategy that is used more frequently and effectively by translators. Simultaneously, user’s selection of translation heuristic word often requires constant attempt, and when no satisfied translation reference is obtained by certain search engine, the user would change another search engine to attempt once again so that a large number of time is consumed. An automatic mining algorithm of related translation reference knowledge based on multiple search engines and heuristic words is put forward in view of above analysis.

III. AUTOMATIC MINING ALGORITHM OF TRANSLATION REFERENCE KNOWLEDGE

The main concept of this algorithm is: extracting the key word from "Unit to be translated", using online dictionary (Kookge.com online dictionary is used in this paper) to obtain the translation items of the key word of the unit to be translated, forming a translation candidate heuristic word set, sorting the translation candidate heuristic word set, retrieving in multiple search engines respectively, and finally optimizing by sorting of the results obtained. Overall process is shown in Figure 1:

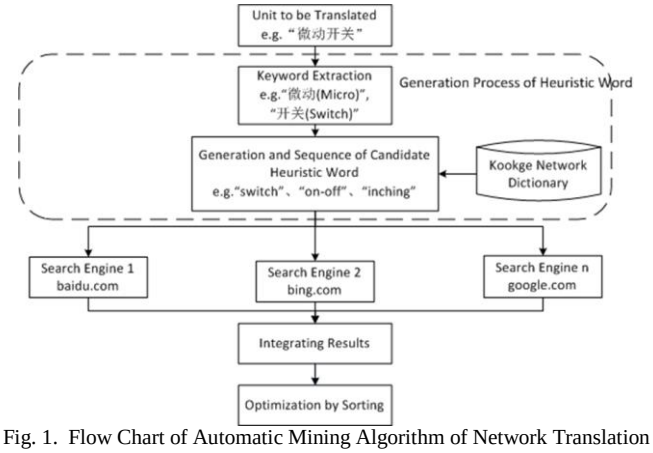

Fig. 1. Flow Chart of Automatic Mining Algorithm of Network Translation Reference Knowledge

A. Key Word Extraction

To obtain the heuristic word, the Chinese word segmentation technology or the Chinese keyword extraction technology can be used to extract key word from units to be translated, then obtain the translations of key words from online dictionary. But the syncopation effect of Chinese word segmentation technology on professional terminology and proper noun generally cannot reach a satisfactory level. To preferably syncopate the unit to be translated and obtain the high-accuracy heuristic word, this paper utilized keyword extraction method based on TextRank to select key word[17].

The key words extracted by TextRank method can better indicate the core of the unit to be translated, thus decreasing noise heuristic word and enhancing the accuracy of candidate heuristic word sets. This paper utilized Fudan Natural Language Processing Toolkit FudanNLP[2] to extract key words. This toolkit outputs the key words extracted in accordance with importance degree (in descending order). For example, “关于建立汽车行业退出机制的通知(Notice on establishing withdrawal mechanism of automobile industry)”, the result after the processing of Chinese keyword extraction technology of FudanNLP toolkit is as follows: “建立(establish)”, “通知(notice)”, “机制(mechanism)”, “汽车(automobile)”, “行业(industry)”, “退出(withdrawal)”.

B. Generation and Sorting of Heuristic Word

This paper adopted the online dictionary method (kookge.com) to obtain the translation items and to form the candidate heuristic word set of the key words extracted in part A of III.

Due to the large number of candidate heuristic words, the heuristic words are sorted to enhance the performance of algorithm, To determine the sorting strategy of heuristic word, the following steps are performed: selecting 100 units to be translated, extracting keywords according to part A of III,

Notes2:Fudan Natural Language Processing Toolkit FudanNLP
download address :https://code.google.com/p/fudannlp/downloads/list

ISBN: 978-988-19252-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

TABLE III
ANALYSIS RETURNED RESULTS OF SEARCH ENGINES

Unit to be Translated	Key Word	Combined Retrieval Mode	Baidu	Google	Bing	Multiple Search Engine	
			Number of Returned Results	Number of Returned Results	Number of Returned Results	Total Number of Returned Results	Average Number of Returned Results
计算机世界	计算机	“计算机世界”+computer	204,000	606,000	54,000	864,000	249,158
		“计算机世界”+calculator	0	23,800	4,390	28,190	
		“计算机世界”+calculating machine	0	7,580	36	7,616	
		“计算机世界”+counting machine	0	96,800	26	96,826	
	世界	“计算机世界”+world	258,000	1,080,000	131,000	1,469,000	1,469,000
…							

obtaining translation items of each key word based on Kookge, forming the corresponding candidate heuristic word set, and combining unit to be translated with heuristic word to retrieve the translation on search engines, such as Baidu, Google and Bing. The results are shown in Table III.

The experimental analysis shows that the less the translation items of key word in dictionary is, the higher accuracy the translation of this key word can be, thus more accuracy retrieval results is available by way of the combined retrieval of translation heuristic word. In addition, the heuristic word with more returned results is better. Accordingly, the heuristic words are required to be sorted, that is, key words are sorted in accordance with the number of translation items in the dictionary, that the number of translation items is less is preferred, then the set of heuristic words is sorted in descending order with the number of returned results in search engines, forming the ordered list. The key words with the same number of translation items are sorted in accordance with the importance by keyword extraction.

C. Information Acquisition of Multiple Search Engines

Considering that the results returned from different search engines are different, this paper collected and merged the retrieval results of multiple search engines. The retrieval process is the combined retrieval of the unit to be translated with each heuristic word, to obtain the retrieval result returned from search engine each time, including title, URL address, abstract and source website, etc. For instance, the retrieval results of “中文信息学报” on Baidu are shown in the Table IV.

TABLE IV
STRUCTURE FOR INFORMATION

Field	Example
Title	中文信息学报 - Journal of Chinese Information Processing
Url	http://www.baidu.com/link?url=D3_1__lpG3Xb9eS1o02UmjYLElcxRhiy7ldfIftjsRFGPFSftZNYiFwLgf3xfmX
Summary	Journal of Chinese Information Processing核心期刊 核心期刊| 期刊首页引文评价期刊简介:《中文信息学报》中文版于1986年创刊,是中国中文信息学会会刊。国内外...
Date	2013-05-01
Source	www.cqvip.com

D. Merging and Sorting of Retrieval Results

The retrieval results returned from each search engine require to be merged for removing noise data, such as commercial advertisement, and sorting the retrieval results in accordance with the validity.

After analysis of a great number of retrieval results, a sorting method based on the combined evaluation of the position, distance, information source and default sorting of search engine present in results of unit to be translated and heuristic words is put forward. The particular description is as follows:

1) Correlation degree score function based on position information. The correlation degree that the unit to be translated or the heuristic word present in title is higher than that in abstract, when both present in title or abstract simultaneously, the correlation degree of this result is particularly outstanding.

Define T_1 denotes whether the unit to be translated presents in title, if it presents , T_1=1, or else T_1=0, T_2 denotes whether the heuristic word appears in title, S_1 denotes whether the unit to be translated presents in abstract, S_2 denotes whether the heuristic word presents in abstract, the valuing method of T_2, S_1, S_2 is the same as that of T_1. Supposing that the weight of unit to be translated and heuristic word present in title is α (0< α<1), then the weight of the unit to be translated and heuristic word present in abstract is (1- α), and the formula of position information score function R_1 is denoted as Formula 1:

$$R_1 = \alpha(T_1 + T_2)^2 + (1-\alpha)(S_1 + S_2)^2 \qquad (1)$$

2) Correlation degree score function based on distance information. On the basis of position information when both the unit to be translated and the heuristic word present in title or abstract, their relative distance information is introduced. The closer the distance is, the larger the correlation degree is.

Define the specific location that the unit to be translated

ISBN: 978-988-19252-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

presents in title is TL_1 ($TL_1 > 0$), in abstract is SL_1($SL_1 > 0$),and the specific location that the heuristic word appears in title is TL_2($TL_2 > 0$), in abstract is SL_2($SL_2 > 0$). α denotes the weight of unit to be translated and heuristic word present in title, which is the same as α in Formula 1, the distance between unit to be translated and heuristic word in title is $|TL_1 - TL_2|$ ($|TL_1 - TL_2| \geq 1$), in abstract is $|SL_1 - SL_2|$ ($|SL_1 - SL_2| \geq 1$),then the formula of distance information score function R_2 is denoted as Formula 2:

$$R_2 = \frac{\alpha}{\log(|TL_1 - TL_2| + 1)} + \frac{1 - \alpha}{\log(|SL_1 - SL_2| + 1)} \quad (2)$$

3) Correlation degree score function base on ranking information in each search engine of returned results. The former the ranking of results is, the higher the correlation degree is.

Define the number of the results selected from each search engine is N (general 10<N<100), the ranking of result in the ith search engine is n_i, i denotes the serial number of search engine (i=1,2,3), λ_i denotes the weight of the ith search engine, then the formula of ranking information score function R_3 is denoted as Formula 3:

$$R_3 = \lambda_i \log\frac{N}{n} \quad (3)$$

4) Correlation degree score function based on result source information, which uses the website type file to determine the type of result source, such as educational website(.edu), government website(.gov or .gov.cn) and authority website, etc. The quality of the results returned from these kinds of websites is higher.

The formula of the correlation degree of result source information score function R_4 is denoted as Formula 4, μ ($0 < \mu < 1$) denotes the reliability of result source information.

$$R_4 = \begin{cases} \mu & authority\ website, etc. \\ 0 & other\ website \end{cases} \quad (4)$$

This paper ultimately utilized linear combination to combine the above sorting conditions to form the final sorting score function. The weight of each sorting condition in the experiment is set to 1, and the specific form is denoted as Formula 5:

$$R = \beta_1 R_1 + \beta_2 R_2 + \beta_3 R_3 + \beta_4 R_4 \quad (5)$$

Finally, all returned result are sorted in accordance with the R value of correlation degree in descending order.

IV. Experiment and Analysis

Test corpus used in this experiment includes more than 10 thousand network retrieval records generated from Chinese-English translation project of GE-SOFT Translation Service Center (GE-SOFT.com) in aviation field. We randomly sampled 500 records and extracted units to be translated with its accurate translation, then retrieved in Baidu/Google/Bing respectively, the first 20 records are selected from each search engine. Kookge.com is selected as the online dictionary. If the results include the units to be translated and the reference translation, the recommended result is deemed to be effective. If the reference translation is accurate, the unit to be translated is the recommended successful unit. The recommendation precision is used to evaluate the experimental result, and the specific formula of recommendation precision is denoted as Formula 6, RP denotes recommendation precision, RSU denotes recommended successful unit, and UT denotes the unit to be translated.

$$RP = \frac{The\ Number\ of\ RSU}{Total\ Number\ of\ UT} \quad (6)$$

By combining the unit to be translated with heuristic word to retrieve on a single search engine and multiple search engines, and count the number of effective results of TOP5, TOP10 and TOP20 respectively. Statistic and comparison of the number of effective results are shown in Table V:

TABLE V
NUMBER OF EFFECTIVE RESULTS

Number of Heuristic Words	Single Search Engine(Bing)			Multiple Search Engine		
	TOP5	TOP10	TOP20	TOP5	TOP10	TOP20
1	3.01	4.61	6.01	4.18	7.8	13.51
2	3.6	6.07	9.03	4.25	8.36	15.28
3	3.85	6.57	11.03	4.42	8.67	16.07
4	3.88	6.61	11.14	4.46	8.71	16.17

The RP with multiple search engines and multiple heuristic words are shown in Figure 2, and HW denotes heuristic words.

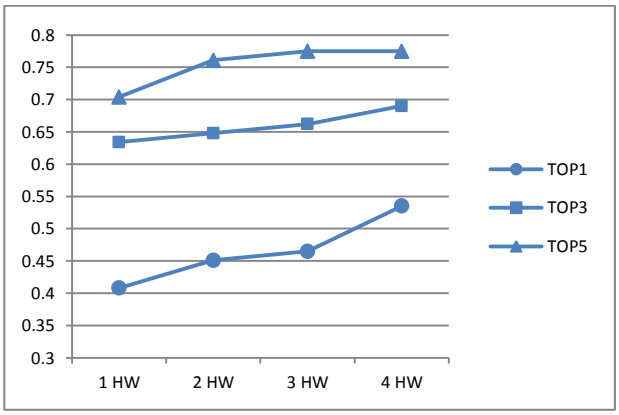

Fig. 2. Multiple Search Engine and Multiple Heuristic Words

In order to verify the effect on sorting of heuristic words, we performed the experiment with heuristic word from the first to the fourth and counted the RP1 of top 1 in multiple search engines. The specific form of RP1 is denoted as Formula 7:

$$RP1 = \frac{The\ Number\ of\ RSU}{Number\ of\ UT_k} \quad (7)$$

UT_k denotes the unit to be translated with at least k (k = 1,2,3,4,5) heuristic words. For example, the RP1 with the

ISBN: 978-988-19252-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

second heuristic word is equivalent to the number of RSU divided by the number of UT contains two or more heuristic words. Statistical results are shown in Figure 3.

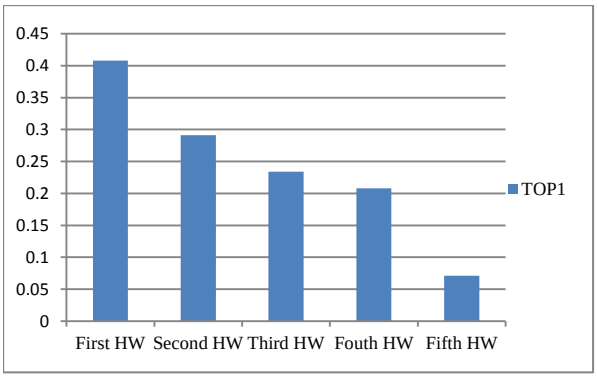

Fig. 3. Multiple Search Engine and Single Heuristic Words

Analysis of experimental results:

1) According to Table V, compared with single search engine, more translation reference knowledge could be acquired through multiple search engines. The sorting strategy in this paper could be more effective of obtaining available translation reference knowledge than the default sorting result of the search engine.
2) According to Figure 2, utilization combination retrieval strategy in multiple search engines can effectively improve the RP level. With 4 heuristic words, the RP of TOP1, TOP3 and TOP5 achieve 53.5%, 69% and 77.5% respectively.
3) According to Figure 3, the first heuristic word has the highest RP1. It can be seen that the strategy of sorting heuristic words is effective.

V. Conclusion

In order to enhance translator's retrieval efficiency and effect of Internet translation reference knowledge, this paper proposed an automatic mining algorithm of Internet translation reference knowledge based on multiple search engines and multiple translation heuristic words after the analysis of Internet retrieval logs generated in the practical translation process. The experimental results show the efficiency of the proposed algorithm, and a translator can retrieve translation reference knowledge on the Internet with high efficiency and accuracy. The future work will focus on involving context message of "Unit to be translation" and using multiple heuristic words instead of single word.

References

[1] L. Jiang, S. Q. Yang and M. Zhou, "Mining Bilingual Data from the Web with Adaptively Learn Patterns," in Proc. ACL Conf., 2009, pp. 870–878.
[2] Z. Lin, Y. J. LV, Q. Liu and X. R. Ma, "Mining Parallel Corpora from Web and Its Application in Machine translation," Journal of Chinese Information Processing, vol. 24, no. 05, pp. 85–91, Sep. 2010.
[3] Z. D. Zhu, M. Li, "Automatically Mining Parallel Corpora for Minority Languages from Web Pages," in Proc. Asian Language Processing Conf., Hanoi, 2012, pp. 97–100.
[4] Z. D. Zhu, "Research on Key Technology in Mining Web Bilingual Corpora," University of Science and Technology of China, 2014.
[5] D. F. Liu, B. Su, "Research in Identification and Purification of the Bilingual Web Page," in Proc. International Colloquium on Computing, Communication, Control and Management Conf., Guangzhou, 2008, pp. 576–579.
[6] D. F. Liu."Research on Methods of Intelligent Bilingual Search and Search Engine," Wuhan University of Technology, 2009.
[7] Y. Zhang and P. Vines, "Using the Web for Automated Translation Extraction in Cross-Language Information Retrieval," in Proc. SIGIR Conf., Sheffield, 2004, pp. 162–169.
[8] Y. J. Dong and D. F. Cai, "Research on Web-Based Chinese-English Term Translation," Journal of Shenyang Institute of Aeronautical Engineering, vol. 27, no. 2, Apr. 2012.
[9] Y. Luo, D. Ji, G. P. Zhang and Y. Y. Wang, "A Method of Mining Bilingual Resources from Single Bilingual Web Pages," Journal of Chinese Information Processing, vol.25, no. 1, pp.110–115, Jan. 2011.
[10] J. Guo, Y. J. Lv and Q. Liu, "An Effective Method to Extract Translation Pairs from Web Corpora," Journal of Chinese Information Processing, vol.22, no. 6, pp.110–115, Jul. 2008.
[11] J. Zhang, C. G. Cao and S. Wang, "Web-based Term Translation Extraction and Verification Method," Journal of Computer Science, vol. 39, no. 7, pp. 170–174, July. 2012.
[12] G. H. Cao, J. F. Gao and J. Y. Nie, "A system to mine large-scale bilingual dictionaries from monolingual web pages," MT summit. pp. 57–64. 2007.
[13] D. Lin, S. Zhao, B. Durme and M. Pasca, "Mining Parenthetical Translations from the Web by Word Alignment," in Proc. ACL Conf., New York, 2008, pp 994–1002.
[14] J. L. Wang, "Application of Internet Resources," Chinese Science &Technology Translations Joural,vol.20,no.2,pp.36–40,May.2007.
[15] X. D. Zhu and D. R. Hua, "Tips Forward Word Search Using Search Engine," Chinese Science &Technology Translations Joural,vol.20,no.3,pp.33–35,Aug.2007.
[16] J. Zhou, "Search Engine Assisted Translation," Journal of Computer-Assisted Foreign Language Education, vol. 22, no. 6, pp. 62–65, May. 2007.
[17] R. Mihalcea and P. Tarau. "Textrank: Bringing order into texts," in Proc. EMNLP Conf., Barcelona, 2004, pp. 404–411.

ISBN: 978-988-19252-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)