A Japanese-Chinese Cross-Language Entity Linking Method with Entity Disambiguation Based on Document Similarity

Xiang Song, Jialiang Zhou, Fuminori Kimura, and Akira Maeda

***Abstract*—In this paper, we propose a method to automatically discover links between valuable keyphrases in a Japanese document and corresponding Chinese encyclopedia pages. The proposed method has three stages. First, we translate Japanese keyphrases into Chinese using a combination of three translation methods. Second, we extract all Chinese encyclopedia articles of the translated keyphrases. Third, we translate the original Japanese document into Chinese and make a vector of noun frequencies. We calculate the cosine similarities of original articles and all candidate Chinese encyclopedia ones. To find the appropriateness of term description pages for disambiguation, we make a rank with cosine similarity by comparing a Japanese document with Chinese encyclopedia articles. Finally, we add a link from a Japanese keyphrase to top-ranking Chinese encyclopedia article. In this paper, we use Wikipedia and Baidu Baike (an online encyclopedia published by Baidu, a Chinese search engine) articles to conduct our experiment. Although we achieved an accuracy rate of 81% by using Wikipedia, we achieved an accuracy rate of 97% by using Baidu Baike.**

***Index Terms*—Encyclopedia, cross-language link discovery, Wikification, Baidu Baike.**

I. Introduction

The Internet stores a lot of valuable knowledge and information in various languages. In addition, hyperlinks in Web pages link from keyphrases to related information as well as definitions. However, related information is not always provided in users' native languages, which causes difficulties for non-native speakers to understand document contents written in foreign languages.

Cross-Language Entity Linking (CLEL) links to the same concept in different language versions. Fig. 1 shows an example of Cross-Language Entity Linking (CLEL) between English and Chinese. The upper half of the figure is a part of an English-language news article about Amazon. The bottom of the figure provides a Chinese Wikipedia article about Amazon. If we have a link from the English to Chinese, the English news article can be better understood by Chinese readers.

Hyperlinks between keyphrases in a document and definition of such keyphrases in the native language will help foreign-language students to find out the meanings of the words they cannot understand. Therefore, these kinds of links may be helpful for foreign students to learn English. However, such cross-language links are rarely available in documents.

To solve the problem above, it is desirable to add automatic links to documents written in readers' native languages. Therefore, we propose a method to obtain Chinese encyclopedia articles that explain the meaning of keyphrases in a Japanese document. Such a mechanism is useful for Chinese students studying Japanese and could further enhance the utility of online encyclopedias such as Wikipedia and Baidu Baike.

In this paper, we use not only Wikipedia but also Baidu Baike to conduct our experiment. The main advantage of using Baidu Baike is that it has a larger number of articles, i.e. 15 times more than Chinese Wikipedia. Therefore, even if Wikipedia has no explanation for a concept, it can be searched for in Baidu Baike. For example, English Wikipedia has an article for "Massively multiplayer online game", but Chinese Wikipedia does not. In contrast, Baidu Baike has a corresponding article entitled "大型多人在线游戏".

Because the aim of CLEL is to link from Japanese to Chinese, we recruited Chinese students studying in Japan to help with our research. In our previous work, the system automatically selected keyphrases. In this study, we asked the Chinese students studying to manually select the words and phrases they did not understand or wanted to understand in more depth.

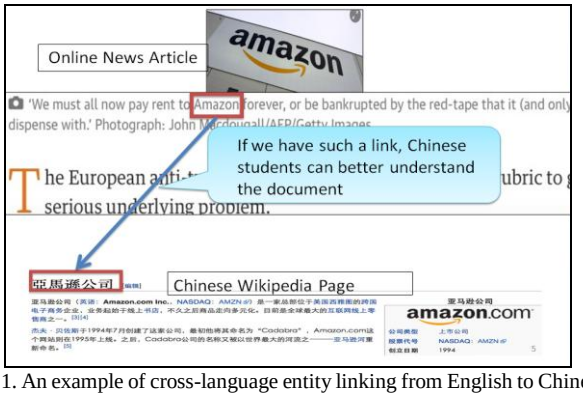

Fig. 1. An example of cross-language entity linking from English to Chinese.

After that, we translate keyphrases by using our proposed method. The contents of Baidu Baike are updated more

Manuscript received December 5, 2015; revised March 21, 2016. This work was supported in part by the JSPS Grant-in-Aid forScientific Research (C) 24500300 "Research on Integrated InformationRetrieval from Multilingual Digital Archives" from Japan Society for the Promotion of Science (JSPS).

Xiang Song and Jialiang Zhou are with the Graduate School of Information Science and Engineering, Ritsumeikan University, Shiga, Japan (e-mail: gr0187xx@ed.ritsumei.ac.jp, is0095hx@ed.ritsumei.ac.jp).

Fuminori Kimura is with the Faculty of Economics Management and Information Science, Onomichi City University, Hiroshima, Japan (e-mail: f-kimura@onomichi-u.ac.jp).

Akira Maeda is with the College of Information Science and Engineering, Ritsumeikan University, Shiga, Japan (e-mail: amaeda@is.ritsumei.ac.jp).

doi: 10.18178/ijke.2016.2.3.065

frequently than those of Wikipedia. Moreover, our proposed method links Japanese keyphrases to articles in Baidu Baike by finding appropriate articles through the search function, rather than use whole content downloaded beforehand, as in our previous work using Wikipedia. Therefore, when a new entry appears, our method using Baidu Baike can always link to the latest Chinese article. Another advantage of using the search function is that we can filter out articles that are unrelated to the keyphrases and thus narrow down the candidate articles for disambiguation significantly, which greatly reduces the cost of calculation while improving the accuracy of the results.

The most obvious feature of Baidu Baike compared with Wikipedia is it offers not just an encyclopedia but also a dictionary function. If we search for a keyphrase in Baidu Baike, we can find that it shows not only the exact meaning of the word but also its origin, as well as some culture-related meaning. For example, "art" is translated as "艺术" in Chinese. Baidu Baike gives has both the exact meaning, origin of the word, as well as an explanation of it as a cultural term in the "art" field. Therefore, we believe that such a feature is very practical for language learners.

II. Related Work

Wikipedia is a multilingual online encyclopedia that anyone can use and edit on a Web browser. Articles for the same topic in different languages are usually linked via inter language links. However, some articles in some languages do not have appropriate versions in different languages. To solve this problem, a variety of research has been done so far.

In the Text Analysis Conference (TAC) [1], the task of Cross-language Entity Linking (CLEL) is being performed. The purpose of this task is to extract PER (person), ORG (organization), and GPE (geopolitical entity) from Chinese or Spanish documents. Then they link them to appropriate English documents. In this paper, the target entities are not limited to places or names, which is the main difference between the proposed method and CLEL.

Cross-lingual link discovery (CLLD) is concerned with automatically finding potential links between documents in different languages. In contrast to conventional information retrieval tasks where queries are not attached to an explicit context, or only loosely attached to context, CLLD algorithms actively recommend a set of meaningful anchors in the context of a source document and establish links to documents in an alternative language [2]. Moreover, other researchers have studied OKSAT [3], KECIR [4], UKP [5] and RDLL [6]. However, they aimed to link valuable keyphrases in a language other than English to related English Wikipedia pages.

Chen *et al*. [7] proposed an approach in which the first step is extracting n-grams from the query source documents as potential anchors. The next step is the anchor expansion and ranking. The final step of the anchor selection process is to re-rank anchors by computing the similarity between the title of the current query Wikipedia page and each element in the vector of expanded potential anchors using Wikipedia Miner (http://wikipedia-miner.cms.waikato.ac.nz/). Different from our work, they only use Google Translate to translate the potential anchors. In this paper, we translate the keyphrase using three methods and extract all the Chinese articles of the translated keyphrase. Finally, we make a ranking using cosine similarity comparison of the Japanese document and Chinese Wikipedia articles.

Liu *et al*. [8] divided their cross-language link discovery task into three sub-problems. Their approach has three steps: anchor mining, cross-lingual linking to related articles, and disambiguation. Similar with Chen *et al*.'s work, they choose Google Translate as the anchor translation tool. In this paper, we use three translation methods and make a ranking using a cosine similarity comparison of the Japanese document and Chinese Wikipedia articles find the appropriateness of term description pages.

Wang *et al*. [9] proposed a cross-lingual knowledge linking approach for building cross-lingual links across Wiki knowledge bases. Their approach uses only language-independent features of articles and employs a graph model to predict new cross-lingual links. They also used Baidu Baike for their experiment. Different from our research, they aim to find the corresponding document in Baidu Baike to explain the full article in English Wikipedia, but we propose a method to find the appropriate document that can explain the keyphrase extracted.

Blanco *et al*. [10] proposed a new probabilistic model and algorithm for entity linking in web search queries that extracts a large number of candidate aliases for entities from click-through information and anchor texts. Their method leverages information from query logs and anchor texts to automatically obtain a large number of aliases for the entities in a knowledge base and uses a probabilistic model to rank entities in query segments. Different from their search, we have students studying in Japan who actually read the documents extract keyphrases, and we make a ranking using cosine similarity comparison of Japanese documents and Chinese articles. According to the calculation results, we can solve the ambiguity of Chinese candidate articles.

Chisholm and Hachey [11] use their method to demonstrate the potential for web links to both complement and completely replace Wikipedia derived data in entity linking. Additionally, Mihalcea and Cosmai [12] introduced the use of an encyclopedia as a resource to support accurate algorithms for keyphrase extraction and word sense disambiguation.

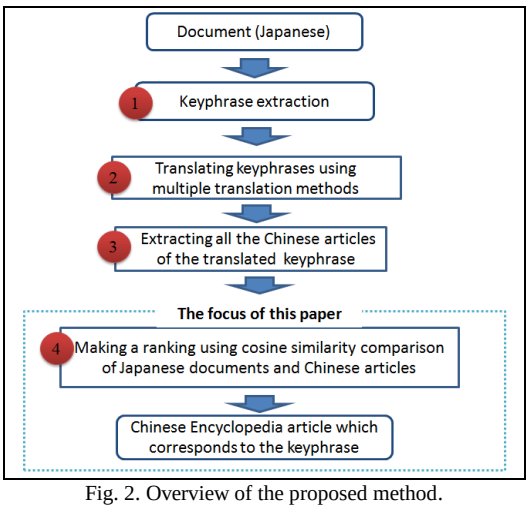

Fig. 2. Overview of the proposed method.

III. Proposed Method

In this section, we describe the method to detect an appropriate Chinese encyclopedia article for a keyphrase in a Japanese document. Fig. 2 shows the overview of the proposed method. Fig. 3 shows an example of the proposed method. The proposed method consists of four processes as follows:

1) Extraction of keyphrases

We asked Chinese students who are studying in Japan to manually select the words and phrases they did not understand or wanted to understand in more depth. Then, we select two keyphrases from each news article.

The Top Consecutive Nouns Cohesion (TCNC) method [13] that we used in our previous work is for extracting candidate keyphrases. TCNC connects continuous nouns and treats them as one compound word. Since keyphrase extraction is not the object of this paper, we manually extracted keyphrases.

2) Translation of the keyphrases using multiple translation methods

Since our aim is to find corresponding Chinese articles for Japanese keyphrases, we have to translate Japanese keyphrases into Chinese. We use two kinds of machine translations and a shift method of literal code to fulfill our task.

3) Obtaining of the corresponding Chinese articles

In this process, we obtain the corresponding Chinese articles for Japanese keyphrases that have been translated into Chinese in the previous process. When using Wikipedia as the target encyclopedia, we can obtain the whole Wikipedia dump data beforehand. However, when using Baidu Baike, we cannot obtain the whole data. Consequently, we have to search for text information for a keyphrase using the Web interface of Baidu Baike and extract all the text content as the corresponding Chinese articles by using Web scraping.

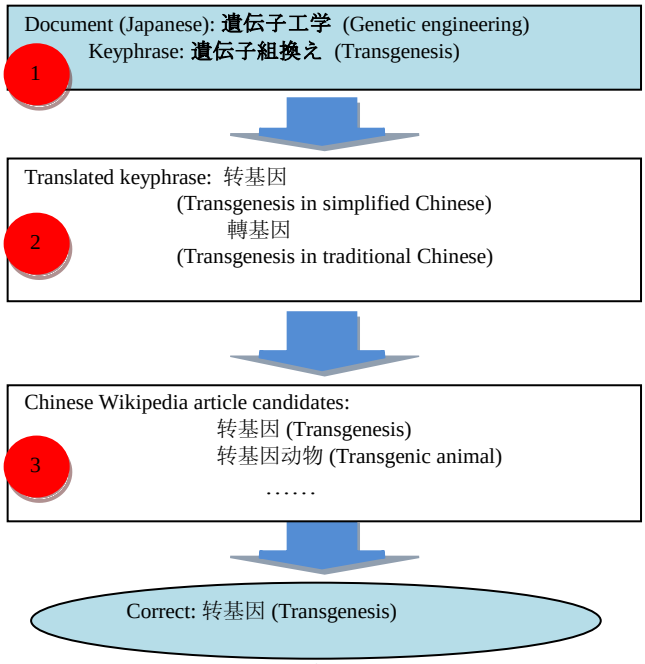

Fig. 3. An example of the proposed method.

4) Ranking of Chinese candidate articles

Lastly, we make a rank using cosine similarity by comparing Japanese news articles with corresponding Chinese articles. According to the calculation results, we can solve the ambiguity of Chinese candidate articles.

A. Translation using Multiple Methods

In the proposed method, we translate the Japanese keyphrases into Chinese using a combination of three translation methods: two kinds of machine translations (Google Translate (https://translate.google.com/) and Bing Translator (https://www.bing.com/translator/)) and a character code conversion method. We consider all the obtained Chinese translations from all three translation methods (including incorrect translations) as translation candidates for a Japanese keyphrase in order to prevent the appropriate translation from being missed. Incorrect translations will be eliminated in the ranking process, which will be explained in the next section.

B. Ranking of Chinese Candidate Articles

In this procedure, we rank obtained Chinese encyclopedia articles in order to decide the most appropriate Chinese article for each Japanese keyphrase. Fig. 4 shows the flow of this ranking procedure.

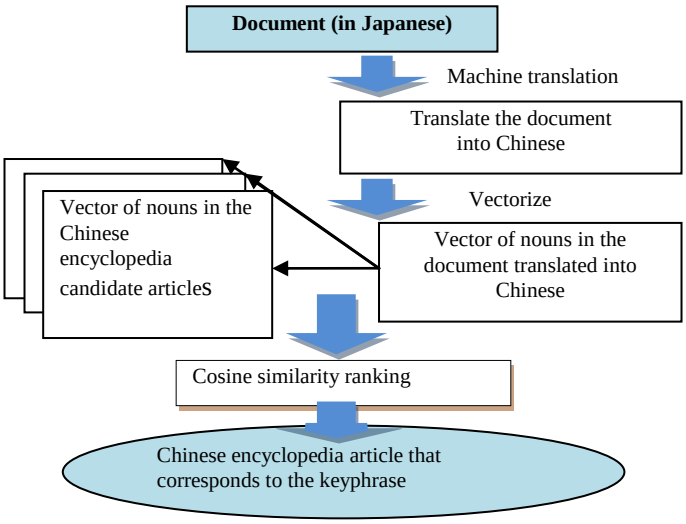

Fig. 4. Processing flow of the ranking of Chinese candidate articles.

First, we translate the original Japanese document from which we extracted the keyphrases into Chinese. Second, we extract nouns from the translated Japanese document and all the obtained Chinese encyclopedia articles. Third, the frequencies of these nouns are regarded as a vector, and we calculate the cosine similarities between the original document and all the Chinese encyclopedia articles. Fourth, we adopt the highest ranked article as the corresponding Chinese article for the Japanese keyphrase. We can find the corresponding Chinese article for each Japanese keyphrase by this procedure even if we obtain many Chinese articles as candidates of the corresponding keyphrase.

In our previous work, we used Chinese Wikipedia as the target encyclopedia for the proposed method. In this paper, we use Baidu Baike as the target encyclopedia. Since the focus of this paper is the disambiguation process in entity linking, we did not automatically extract keyphrases from the article but asked Chinese students studying in Japan to manually select the words and phrases they did not understand

or wanted to understand in more depth. Note that the students differed in terms of Japanese language ability.

Because Baidu Baike does not provide the dump of whole data as Wikipedia does, we cannot directly download all data. With Wikipedia, we can download all article data directly, and each article has its corresponding ID, through which we can easily find the corresponding article. However, Baidu Baike does not provide such data. Therefore, we first use the translated keyphrase to find the corresponding article in Baidu Baike. Then, through the corresponding article located on the Baidu Baike Web site by using the Web scraping, we extract the article contents. Finally, we obtain all the corresponding articles.

Baidu Baike is different from Wikipedia since it has only Chinese articles. However, it has many more articles (almost 12 million) than Chinese Wikipedia (approx. 800 thousand). Thanks to the large number of articles, we have more chances to find the corresponding Chinese article in Baidu Baike than in Chinese Wikipedia.

IV. Experiment

A. Overview of the Experiments

To evaluate the proposed method, we conducted two experiments to obtain Chinese encyclopedia articles corresponding to the keyphrases in Japanese documents. The details of the experiment are described in our previous paper [13].

In the first experiment, we selected 20 Japanese Wikipedia articles in which the Chinese students were interested, and for each article, 55 keyphrases were manually extracted [14].

Although our proposed method includes the procedure to extract keyphrases, the purpose of this experiment is to confirm the accuracy of the ranking procedure, and thus we extracted keyphrases for each article manually in this experiment.

All the extracted keyphrases were important to their respective articles and difficult for the Chinese students to understand. The proposed system might acquire a lot of corresponding article candidates if the keyphrase is short. In this experiment, for 55 keyphrases in 20 articles, the proposed system acquired 1,350 articles in total from the candidate keyphrases. The average value of the acquired articles is 26, and the keyphrase "battle" acquired the maximum number of corresponding articles: 187. The relevance of the Chinese corresponding articles to the keyphrase in Japanese was decided by one of the authors by reading the acquired articles.

In the second experiment, we selected 30 Japanese news articles from the Asahi Shimbum Web site, and two keyphrases were extracted from each article by Chinese students who read them all. They selected two keyphrases that they could not understand, and we extracted the two keyphrases that were selected the most. Finally, we extracted 58 keyphrases that had corresponding articles in Baidu Baike. In this experiment, for 58 keyphrases in 30 articles, we acquired 97 articles in total from the candidate keyphrases by using the proposed system.

The details of the translation procedure and its experiments are described in our previous paper [15].

B. Ranking of Chinese Candidate Articles

To evaluate the accuracy of the ranking procedure, we conducted an experiment of the ranking of Chinese candidate encyclopedia articles by calculating the cosine similarities between the source document in Japanese and the corresponding article candidates in Chinese.

In the first experiment, we used the same 20 Wikipedia articles as the previous experiment. From the 55 keyphrases, we selected 26 keyphrases for the experiment. Due to the effort required to find the appropriate article that corresponds to Japanese keyphrases from a large number of candidate articles, we selected 26 keyphrases for which appropriate articles were relatively easy to find. Second, we translated the original Japanese document from which keyphrases were extracted into Chinese. Third, we calculated the vector of nouns in the document translated into Chinese and all the Chinese Wikipedia candidate articles. Then, we calculated cosine similarity between the translated original document and each candidate article one by one. Lastly, we ranked Chinese candidate articles by these cosine similarities. In this experiment, we used ANSJ_Seg as the Chinese morphological analyzer to extract nouns from Chinese texts.

Among the 26 keyphrases, 21 articles were correctly ranked at the top by the cosine similarity ranking. Therefore, the accuracy rate of selecting the relevant articles for keyphrases is approximately 81%. The results of ranking within the top 3 and top 5 are summarized in Table I.

TABLE I: Accuracy Rate of Ranking for Wikipedia

Ranking	Accuracy Rate
Top 1	21/26 (81%)
Within Top 3	25/26 (96%)
Within Top 5	26/26 (100%)

In the second experiment, we used the 30 Japanese news articles from the Asahi Shimbun and translated the articles from which keyphrases were extracted into Chinese. Second, we calculated the vector of nouns in the news articles translated into Chinese and all the Chinese Baidu Baike candidate articles. Then, we calculated cosine similarity between the translated news and each candidate article one by one. Lastly, we ranked Chinese candidate articles by these cosine similarities. For the Chinese candidate articles, we first used only the abstract to calculate the cosine similarity and then used the full articles to calculate it.

TABLE II: Accuracy Rate of Ranking for Abstract (Baidu Baike)

Ranking	Accuracy Rate (Abstract)
Top 1	50/58 (86%)
Within Top 3	54/58 (93%)
Within Top 5	57/58 (98%)

TABLE III: Accuracy Rate of Ranking for Full Article (Baidu Baike)

Ranking	Accuracy Rate (Full Article)
Top 1	51/58 (87%)
Within Top 3	53/58 (91%)
Within Top 5	55/58 (94%)

When we used the Chinese candidate articles (abstract) from Baidu Baike among the 58 keyphrases, 50 articles were correctly ranked at the top by the cosine similarity ranking. Therefore, the accuracy rate of selecting the relevant articles for keyphrases is approximately 86%. The results of ranking

within the top 3 and top 5 are summarized in Table II.

When we used the Chinese candidate articles (full article) from Baidu Baike among the 58 keyphrases, 51 articles were correctly ranked at the top by the cosine similarity ranking. Therefore, the accuracy rate of selecting the relevant articles for keyphrases is approximately 87%. The results of ranking within the top 3 and top 5 are summarized in Table III.

C. Weighting Linked Phrases in Similarity Ranking

In order to further improve the accuracy, we conducted an additional experiment, in which phrases with hyperlinks in the candidate articles are given higher weights in similarity calculation. In this experiment, we used another 30 Japanese news articles from the Asahi Shimbun and keyphrases were extracted by Chinese students. 29 keyphrases were manually extracted in total. Table IV shows all of the keyphrases extracted for this experiment.

TABLE IV: KEYPHRASES FOR THE ADDITIONAL EXPERIMENT

Keyphrase List	
富士山 (Mount Fuji)	立候補 (candidacy)
不祥事 (scandal)	アスリート (athlete)
賄賂 (bribe)	陸連 (Japan Association of Athletics Federations)
総裁 (president)	戦闘機 (fighter aircraft)
フィリピン (Philippines)	インフラ (infrastructure)
披露宴 (reception)	台頭 (rise)
肉声 (natural voice)	葬儀 (funeral)
俳優 (actor)	フォーラム (forum)
字幕 (caption)	湧水 (spring water)
頭蓋骨 (skull)	コンセプト (concept)
ガバナンス (governance)	クジラ (whale)
水族館 (aquarium)	遺伝子 (gene)
アユ (sweetfish)	暗渠 (ditch)
ペンギン (penguin)	海坊主 (sea monster)
坂本龍馬 (Sakamoto Ryoma)	

Among the 29 keyphrases, there are 30 Chinese candidate articles and 26 articles were correctly ranked at the top by the cosine similarity ranking without weighting linked phrases in the candidate articles. In this case, the accuracy rate of selecting the relevant articles for keyphrases is approximately 87%.

For weighting linked phrases, we make 10 times weights for the phrases that have hyperlinks in the Chinese candidate articles. We calculate the cosine similarities of translated Japanese news and Chinese candidate articles in the same way as the previous experiment. As a result, 3 correct keyphrases rise to the top compared to the case of not weighting linked phrases, and 1 keyphrase has no change. In this case, 29 articles out of 30 were correctly ranked at the top by the cosine similarity ranking. The accuracy rate of selecting the relevant articles for keyphrases is greatly improved by weighting linked phrases, and it achieved approximately 97%.

V. DISCUSSION

In the first experiment, when there is an article whose title fully matches the translated keyphrase, a high accuracy rate was achieved. However, the most appropriate articles are not ranked as the top even if the Japanese keyphrase was translated correctly. There are two reasons for this.

First, "Emperor Meiji and the Russo-Japanese War" is the name of a movie. In this article, most personal names that appear are consistent with the original Japanese document, so the similarity degree of the top ranked article is higher than that of the most appropriate article.

Second, the Chinese articles about "Eboshi" and "Eboshi parent" are relatively short and both only have about 50 nouns. Therefore, the similarity degree calculation did not work well.

In the second experiment, we used Baidu Baike instead of Wikipedia. Baidu Baike is an online encyclopedia established by Baidu in April 2006. It is a search engine and also a knowledge base that everyone can use for free. Different from Wikipedia, Baidu Baike has just Chinese articles, but it has about 15 times as many encyclopedia articles as Chinese Wikipedia. If we can translate the keyphrase correctly, the probability of finding the corresponding article is very high.

In the experiment of weighting linked phrases, we were able to achieve the accuracy rate of 97%. The only error was the keyphrase "不祥事 (scandal)". In fact, there are no hyperlinks in the correct article, so the weighting had no effect at all.

VI. CONCLUSION

In this paper, we proposed a method to link between keyphrases in a Japanese document with the appropriate articles in an online Chinese encyclopedia automatically.

However, the proposed method still has some problems that need be addressed.

First, machine translation is primarily designed to be used for translating sentences, rather than individual keyphrases as our current method is. We are planning to consider the context of a word for resolving the word ambiguities.

Second, in our proposed method, when we translate a keyphrase and find Chinese encyclopedia articles, we still obtain some ambiguous pages. In our future work, we are planning to consider a method to find the most appropriate article.

REFERENCES

[1] H. Ji, J. Nothman, and B. Hachey, "Overview of tac-kbp2014 entity discovery and linking tasks," in *Proc. the Text Analysis Conference*, 2014.

[2] L. X. Tang, I. Kang, F. Kimura, Y. S. Lee, A. Trotman, S. Geva, and Y. Xu, "Overview of NTCIR-10 cross-lingual link discovery task," in *Proc. the 10th NTCIR Conference*, 2013, pp. 8-38.

[3] S. Takashi, "Osaka Kyoiku University at NTCIR-10 crosslink-2 link," in *Proc. the 10th NTCIR Conference*, 2013, pp. 47-50.

[4] J. X. Zheng, Y. Bai, C. Guo, and D. F. Cai, "KECIR at NTCIR-10 cross-lingual link discovery task," in *Proc. the 10th NTCIR Conference*, 2013, pp. 51-56.
[5] J. Kim and I. Gurevych, "UKP at crosslink2: CJK-to-English subtasks," in *Proc. the 10th NTCIR Conference*, 2013, pp. 57-61.
[6] F. Kimura, K. Horita, Y. Konishi, H. Harada, and A. Maeda, "RDLL at crosslink anchor extraction considering ambiguity in CLLD," in *Proc. the 10th NTCIR Conference*, 2013, pp. 82-86.
[7] S. Chen, G. J. F. Jones, and N. E O'Conner, "DCU at NTCIR-10 cross-lingual link discovery (crosslink-2) task," in *Proc. the 10th NTCIR Conference*, 2013, pp. 74-78.
[8] Y. L. Liu, J. Boisson, and J. S. Chang, "NTHU at NTCIR-10 crosslink-2: An approach toward semantic features," in *Proc. the 10th NTCIR Conference*, 2013, pp. 62-68.
[9] Z. C. Wang, J. Z. Li, Z. G. Wang, and J. Tang, "Cross-lingual knowledge linking across wiki knowledge bases," in *Proc. the 21st International Conference on World Wide Web*, 2012, pp. 459-468.
[10] R. Blanco, G. Ottaviano, and E. Meij, "Fast and space-efficient entity linking for queries," in *Proc. the Eighth ACM International Conference on Web Search and Data Mining*, 2015, pp. 179-188.
[11] A. Chisholm and B. Hachey, "Entity disambiguation with web links," *Transactions of the Association for Computational Linguistics*, vol. 3, pp. 145-156, 2015.
[12] R. Mihalcea and A. Csomai, "Wikify!: Linking documents to encyclopedic knowledge," in *Proc. CIKM' 07*, 2007, pp. 233-242.
[13] K. Horita, F. Kimura, and A. Maeda, "Automatic keyword extraction for Wikification of east Asian language documents," *International Journal of Computer Theory and Engineering*, vol. 8, no. 1, pp. 32-35, 2016.
[14] J. L. Zhou, X. Song, F. Kimura, and A. Maeda, "A cross-language entity linking method using combination of multiple translation methods," in *Proc. the 2015 4th ICT International Student Project Conference*, 2015.
[15] X. Song, J. L. Zhou, F. Kimura, and A. Maeda, "A Japanese-Chinese cross-language entity linking method based on appropriateness of term description pages," in *Proc. the 4th IIAI International Congress on Advanced Applied Informatics*, 2015.

Xiang Song is a master's student at the Graduate School of information science and engineering, Ritsumeikan University, Shiga, Japan. He was born in Liaoning province of China. He graduated from the Dalian University of Foreign Languages. His research interests include text mining and multilingual information processing.

Jialiang Zhou is a master's student at the Graduate School of information science and engineering, Ritsumeikan University, Shiga, Japan. He was born in Shandong province of China. He graduated from the Ritsumeikan University.

Fuminori Kimura is a lecturer in the Faculty of Economics, Management and information Science, Onomichi City University. He obtained a Ph.D. degree in engineering from Nara Institute of Science and Technology in 2007. His research interests include information retrieval, text mining, and multilingual information processing.

Akira Maeda is a professor at the Department of Media Technology, College of Information Science and Engineering, Ritsumeikan University. He received B.A. and M.A. degrees in library and information science from the University of Library and Information Science (ULIS) in 1995 and 1997 and received the Ph.D. degree in engineering from Nara Institute of Science and Technology (NAIST) in 2000. He visited Virginia Polytechnic Institute and State University (Virginia Tech) from October 2000 through March 2001 as a postdoctoral visiting scholar. He worked as a postdoctoral researcher of the CREST program, Japan Science and Technology Corporation (JST) from April 2001 through March 2002. He was a visiting professor at King's College London from September 2011 through September 2012. His research interests include digital libraries, digital humanities, information retrieval, and multilingual information processing.