A Voting Mechanism for Named Entity Translation in English–Chinese Question Answering

Ling-Xiang Tang[1], Shlomo Geva[1], Andrew Trotman[2], Yue Xu[1]

[1]Faculty of Science and Technology
Queensland University of Technology

`{l4.tang,s.geva,yue.xu}@qut.edu.au`

[2]Department of Computer Science
University of Otago

`andrew@cs.otago.ac.nz`

Abstract

In this paper, we describe a voting mechanism for accurate named entity (NE) translation in English–Chinese question answering (QA). This mechanism involves translations from three different sources: machine translation, online encyclopaedia, and web documents. The translation with the highest number of votes is selected. We evaluated this approach using test collection, topics and assessment results from the NTCIR-8 evaluation forum. This mechanism achieved 95% accuracy in NEs translation and 0.3756 MAP in English–Chinese cross-lingual information retrieval of QA.

1 Introduction

Nowadays, it is easy for people to access multi-lingual information on the Internet. Key term searching on an information retrieval (IR) system is common for information lookup. However, when people try to look for answers in a different language, it is more natural and comfortable for them to provide the IR system with questions in their own natural languages (e.g. looking for a Chinese answer with an English question: *"what is Taiji"*?). Cross-lingual question answering (CLQA) tries to satisfy such needs by directly finding the cor-

rect answer for the question in a different language.

In order to return a cross-lingual answer, a CLQA system needs to understand the question, choose proper query terms, and then extract correct answers. Cross-lingual information retrieval (CLIR) plays a very important role in this process because the relevancy of retrieved documents (or passages) affects the accuracy of the answers.

A simple approach to achieving CLIR is to translate the query into the language of the target documents and then to use a monolingual IR system to locate the relevant ones. However, it is essential but difficult to translate the question correctly. Currently, machine translation (MT) can achieve very high accuracy when translating general text. However, the complex phrases and possible ambiguities present in a question challenge general purpose MT approaches. Out-of-vocabulary (OOV) terms are particularly problematic. So the key for successful CLQA is being able to correctly translate all terms in the question, especially the OOV phrases.

In this paper, we discuss an approach for accurate question translation that targets the OOV phrases and uses a translation voting mechanism. This mechanism involves translations from three different sources: machine translation, online encyclopaedia, and web documents. The translation with the highest number of votes is selected. To demonstrate this mechanism, we use Google Translate

(GT)[1] as the MT source, Wikipedia as the encyclopaedia source, and Google web search engine to retrieve Wikipedia links and relevant Web document snippets.

English questions on the Chinese corpus for CLQA are used to illustrate of this approach. Finally, the approach is examined and evaluated in terms of translation accuracy and resulting CLIR performance using the test collection, topics and assessment results from NTCIR-8[2].

English Question Templates (QTs)
who [is | was | were | will], what is the definition of, what is the [relationship | interrelationship | inter-relationship] [of | between], what links are there, what link is there, what [is | was | are | were | does | happened], when [is | was | were | will | did | do], where [will | is | are | were], how [is | was | were | did], why [does | is | was | do | did | were | can | had], which [is | was | year], please list, describe [relationship | interrelationship | inter-relationship] [of | between], could you [please | EMPTY] give short description[s] to, who, where, what, which, how, describe, explain
Chinese QT Counterparts
之间有什么关系，的定义是什么，的关系是什么，发生了什么事，是什么关系，是什么时候，的关系如何，之间有什么，请简短简述，请简单简述，什么时候，什么关系，的关系是，有何关系，关系如何，有何相关，有何渊源，为什么会，为什么要，为什么能，是哪一年，什么时候，位于哪里，什么样的，你能不能，相互之间，代表什么，简短简述，简单简述，简短描述，简单描述，为什么，是什么，什么是，的关系，在哪里，怎么样，有哪些，什么事，是哪个，是哪家，有什么，请列出，请列举，请描述，哪一年，请简述，能不能，的定义，何时，谁是，是谁，如何，哪个，列举，请问，何谓，何以，为何，描述，有何，简述，哪些，什么，之间，有关,定义，解释

Table 1. Question templates

2   CLIR Issue and Related Work

In CLIR, retrieving documents with a cross-lingual query with out-of-vocabulary phrases has always been difficult. To resolve this problem, an external resource such as Web or Wikipedia is often used to discover the possible translation for the OOV term. Wikipedia and other Web documents are thought of as treasure troves for OOV problem solving because they potentially cover the most recent OOV terms.

The Web-based translation method was shown to be an effective way to solve the OOV phrase problem (Chen et al., 2000; Lu et al., 2007; Zhang & Vines, 2004; Zhang et al., 2005). The idea behind this method is that a term/phrase and its corresponding translation normally co-exist in the same document because authors often provide the new terms' translation for easy reading.

In Wikipedia the language links provided for each entry cover most popular written languages, therefore, it was used to solve a low coverage issue on named entities in EuroWordNet (Ferrández et al., 2007); a number of research groups (Chan et al., 2007; Shi et al., 2008; Su et al., 2007; Tatsunori Mori, 2007) employed Wikipedia to tackle OOV problems in the NTCIR evaluation forum.

3   CLQA Question Analysis

Questions for CLQA can be very complex. For example, "*What is the relationship between the movie "Riding Alone for Thousands of Miles" and ZHANG Yimou?*". In this example, it is important to recognise two named entities (*"Riding Alone for Thousands of Miles"* and *"ZHANG Yimou"*) and to translate them precisely.

In order to recognise the NEs in the question, first, English question template phrases in Table 1 are removed from question; next, we use the Stanford NLP POS tagger (The Stanford Natural Language Processing Group, 2010) to identify the named entities; then translate them accordingly. Chinese question template phrases are also pruned from the translated question at the end to reduce the noise words in the final query.

There are three scenarios in which a term or phrase is considered a named entity. First, it is consecutively labelled NNP or NNPS (University of Pennsylvania, 2010). Second, term(s) are grouped by quotation marks. For example, to extract a named entity from the example question above, three steps are needed:

1. Remove the question template phrase "*What is the relationship between*" from the question.
2. Process the remaining using the POS tagger, giving "*the_DT movie_NN `` _ `` Riding_NNP Alone_NNP for_IN Thou-*

*sands_NNS   of_IN   Miles_NNP
``_``and_CC ZHANG_NNP Yimou_NNP
?_."*

3. *"Riding Alone for Thousands of Miles"* is between two tags (``) and so is an entity, and the phrase *"ZHANG Yimou"*, as indicated by two consecutive NNP tags is also a named entity.

Third, if a named entity recognised in the two scenarios above is followed in the question by a phrase enclosed in bracket pairs, this phrase will be used as a *tip term* providing additional information about this named entity. For instance, in the question *"Who is David Ho (Da-i Ho)?"*, *"Da-i Ho"* is the tip term of the named entity *"David Ho"*.

4   A Voting Mechanism for Named Entity Translation (VMNET)

Observations have been made:

- Wikipedia has over 100,000 Chinese entries describing various up-to-date events, people, organizations, locations, and facts. Most importantly, there are links between English articles and their Chinese counterparts.

- When people post information on the Internet, they often provide a translation (where necessary) in the same document. These pages contain bilingual phrase pairs. For example, if an English term/phrase is used in a Chinese article, it is often followed by its Chinese translation enclosed in parentheses.

- A web search engine such as Google can identify Wikipedia entries, and return popular bi-lingual web document snippets that are closely related to the query.

- Statistical machine translation relying on parallel corpus such as Google Translate can achieve very high translation accuracy.

Given these observations, there could be up to three different sources from which we can obtain translations for a named entity; the task is to find the best one.

4.1   VMNET Algorithm

A Google search on the extracted named entity is performed to return related Wikipedia links and bilingual web document snippets. Then from the results of Web search and MT, three different translations could be acquired.

Wikipedia Translation

The Chinese equivalent Wikipedia pages could be found by following the language links in English pages. The title of the discovered Chinese Wikipedia page is then used as the *Wikipedia translation*.

Bilingual Clue Text Translation

The Chinese text contained in the snippets returned by the search engine is processed for *bilingual clue text translation*. The phrase in a different language enclosed in parentheses which come directly after the named entity is used as a candidate translation. For example, from a web document snippet, *"YouTube - Sean Chen（陳信安）dunks on Yao Ming..."*, " 陳信安" can be extracted and used as a candidate translation of *"Sean Chen"*, who is a basket ball player from Taiwan.

Machine Translation

In the meantime, translations for the named entity and its tip term (if there is one) are also retrieved using Google Translate.

Regarding the translation using Wikipedia, the number of results could be more than one because of ambiguity. So for a given named entity, we could have at least one, but possibly more than three candidate translations.

With all possible candidate translations, the best one then can be selected. Translations from all three sources are equally weighted. Each translation contributes one vote, and the votes for identical translation are cumulated. The best translation is the one with the highest number of votes. In the case of a tie, the first choice of the best translation is the Wikipedia translation if only one Wiki-entry is found; otherwise, the priority for choosing the best is bilingual clue text translation, then machine translation.

4.2   Query Generation with VMNET

Because terms can have multiple meanings, ambiguity often occurs if only a single term is given in machine translation. A state-of-the-art MT toolkit/service could perform better if more contextual information is provided. So a better translation is possible if the whole sentence is given (e.g. the question). For this rea-

son, the machine translation of the question is the whole query and not with the templates removed.

However, issues arise: 1) how do we know if all the named entities in question are translated correctly? 2) if there is an error in named entity translation, how can it be fixed? Particularly for case 2, the translation for the whole question is considered acceptable, except for the named entity translation part. We intend to keep most of the translation and replace the bad named entity translation with the good one. But finding the incorrect named entity translation is difficult because the translation for a named entity can be different in different contexts. The missing boundaries in Chinese sentences make the problem harder. To solve this, when a translation error is detected, the question is reformatted by replacing all the named entities with some nonsense strings containing special characters as place holders. These place holders remain unchanged during the translation process. The good NE translations then can be put back for the nearly translated question.

Given an English question Q, the detailed steps for the Chinese query generation are as following:

1. Retrieve machine translation T_{mt} for the whole question from Google Translate.
2. Remove question template phrase from question.
3. Process the remaining using the POS tagger.
4. Extract the named entities from the tagged words using the method discussed in Section 3.
5. Replace each named entity in question Q with a special string S_i,($i =0,1,2,..$) which makes nonsense in translation and is formed by a few non-alphabet characters. In our experiments, S_i is created by joining a double quote character with a ^ character and the named entity *id* (a number, starting from 0, then increasing by 1 in order of occurrence of the named entity) followed by another double quote character. The final S_i, becomes **"^*id*"**. The resulting question is used as Q_s.
6. Retrieve machine translation T_{qs} for Q_s from Google Translate. Since S_i consists

of special characters, it remains unchanged in T_{qs}.

7. Start the VMNET loop for each named entity.
8. With an option set to return both English and Chinese results, Google the named entity and its tip term (if there is one).
9. If there are any English Wikipedia links in the top 10 search results, then retrieve them all. Else, jump to step 12.
10. Retrieve all the corresponding Chinese Wikipedia articles by following the languages links in the English pages. If none, then jump to step 12.
11. Save the title $NET_{wiki}(i)$ of each Chinese Wikipedia article *Wiki(i)*.
12. Process the search results again to locate a bilingual clue text translation candidate - NET_{ct}, as discussed in Section 4.1.
13. Retrieve machine translation NET_{mt}, and NET_{tip} for this named entity and its tip term (if there is one).
14. Gather all candidate translations: $NET_{wiki}(*)$, NET_{ct}, NET_{tip}, and NET_{mt} for voting. The translation with the highest number of votes is considered the best (NET_{best}). If there is a tie, NET_{best} is then assigned the translation with the highest priority. The priority order of candidate translation is $NET_{wiki}(0)$ (if $sizeof(NET_{wiki}(*))=1$) > NET_{ct} > NET_{mt}. It means when a tie occurs and if there are more than one Wikipedia translation, all the Wikipedia translations are skipped.
15. If T_{mt} does not contain NET_{best}, it is then considered a faulty translation.
16. Replace S_i in T_{qs} with NET_{best}.
17. If NET_{best} is different from any $NET_{wiki}(i)$ but can be found in the content of a Wikipedia article (*Wiki(i)*), then the corresponding $NET_{wiki}(i)$ is used as an additional query term, and appended to the final Chinese query.
18. Continue the VMNET loop and jump back to step 8 until no more named entities remain in the question.
19. If T_{mt} was considered a faulty translation, use T_{qs} as the final translation of Q. Otherwise, just use T_{mt}. The Chinese question template phrases are pruned from the translation for the final query generation.

A short question translation example is given below:

- For the question "*What is the relationship between the movie "Riding Alone for Thousands of Miles" and ZHANG Yimou?*", retrieving its Chinese translation from a MT service, we get the following: 之间有什么电影"利民为千里单独的关系"和张艺谋.

- The translation for the movie name *"Riding Alone for Thousands of Miles"* of "*ZHANG Yimou*" is however incorrect.

- Since the question is also reformatted into "*What is the relationship between the movie "^0" and "^1"?*", machine translation returns a second translation: 什么是电影之间的关系"＾0"和"＾1"?

- VMNET obtains the correct translations: 千里走单骑 and 张艺谋, for two named entities *"Riding Alone for Thousands of Miles"* and "*ZHANG Yimou*" respectively.

- Replace the place holders with the correct translations in the second translation and give the final Chinese translation: 什么是电影之间的关系"千里走单骑"和"张艺谋"?

5 Information Retrieval

5.1 Chinese Document Processing

Approaches to Chinese text indexing vary: Unigrams, bigrams and whole words are all commonly used as tokens. The performance of various IR systems using different segmentation algorithms or techniques varies as well (Chen et al., 1997; Robert & Kwok, 2002). It was seen in prior experiments that using an indexing technique requiring no dictionary can have similar performance to word-based indexing (Chen, et al., 1997). Using bigrams that exhibit high mutual information and unigrams as index terms can achieve good results. Motivated by indexing efficiency and without the need for Chinese text segmentation, we use both bigrams and unigrams as indexing units for our Chinese IR experiments.

5.2 Weighting Model

A slightly modified BM25 ranking function was used for document ordering.

When calculating the inverse document frequency, we use:

$$IDF(q_i) = log \frac{N}{n} \qquad (1)$$

where N is the number of documents in the corpus, and n is the document frequency of query term q_i. The retrieval status value of a document d with respect to query $q(q_1, ..., q_m)$ is given as:

$$rsv(q, d) =$$

$$\sum_{i=0}^{m} \frac{tf(q_i, d) * (k_1 + 1)}{tf(q_i, d) + k_1 * \left(1 - b + b * \frac{len(d)}{avgdl}\right)} * IDF(q_i) \qquad (2)$$

where $tf(q_i, d)$ is the term frequency of term q_i in document d; $len(d)$ is the length of document d in words and *avgdl* is the mean document length. The number of bigrams is included in the document length. The values of the tuneable parameters k_1 and b used in our experiments are 0.7 and 0.3 respectively.

6 CLIR Experiment

6.1 Test Collection and Topics

Table 2 gives the statistics of the test collection and the topics used in our experiments. The collection contains 308,845 documents in simplified Chinese from *Xinhua News*. There are in total 100 topics consisting of both English and Chinese questions. This is a NTCIR-8 collection for ACLIA task.

Corpus	#docs	#topics
Xinhua Chinese (simplified)	308,845	100

Table 2. Statistics of test corpus and topics

6.2 Evaluation Measures

The evaluation of VMNET performance covers two main aspects: translation accuracy and CLIR performance.

As we focus on named entity translation, the translation accuracy is measured using the precision of translated named entities at the topic level. So the translation precision -*P* is defined as:

$$P = \frac{c}{N} \qquad (3)$$

where c is the number of topics in which all the named entities are correctly translated; N is the number of topics evaluated.

The effectiveness of different translation methods can be further measured by the resulting CLIR performance. In NTCIR-8, CLIR performance is measured using the mean average precision. The MAP values are obtained by running the ir4qa_eval2 toolkit with the assessment results [3] on experimental runs(NTCIR Project, 2010). MAP is computed using only 73 topics due to an insufficient number of relevant document found for the other 27 topics (Sakai et al., 2010). This is the case for all NTCIR-8 ACLIA submissions and not our decision.

It also must be noted that there are five topics that have misspelled terms in their English questions. The misspelled terms in those 5 topics are given in Table 3. It is interesting to see how different translations cope with misspelled terms and how this affects the CLIR result.

Topic ID	Misspelling	Correction
ACLIA2-CS-0024	*Qingling*	*Qinling*
ACLIA2-CS-0035	*Initials D*	*Initial D*
ACLIA2-CS-0066	*Kasianov*	*Kasyanov*
ACLIA2-CS-0074	*Northern Territories*	*northern territories*
ACLIA2-CS-0075	*Kashimir*	*Kashmir*

Table 3. The misspelled terms in topics

6.3 CLIR Experiment runs

A few experimental runs were created for VMNET and CLIR system performance evaluation. Their details are listed in Table 7. Those with name *CS-CS* are the Chinese monolingual IR runs; and those with the name *EN-CS* are the English-to-Chinese CLIR runs. Mono-lingual IR runs are used for benchmarking our CLIR system performance.

7 Results and Discussion

7.1 Translation Evaluation

The translations in our experiments using Google Translate reflect only the results retrieved at the time of the experiments because Google Translate is believed to be improved over time.

The result of the final translation evaluation on the 100 topics is given in Table 4. Google Translate had difficulties in 13 topics. If all

thirteen named entities in those topics where Google Translate failed are considered OOV terms, the portion of topics with OOV phrases is relatively small. Regardless, there is an 8% improvement achieved by VMNET reaching 95% precision.

Method	c	N	P
Google Translate	87	100	87%
VMNET	95	100	95%

Table 4. Translation Evaluation Results

There are in total 14 topics in which Google Translate or VMNET failed to correctly translate all named entities. These topics are listed in Table 8. Interestingly, for topic (ACLIA2-CS-0066) with the misspelled term "*Kasianov*", VMNET still managed to find a correct translation (米哈伊尔·米哈伊洛维奇·卡西亚诺夫). This has to be attributed to the search engine's capability in handling misspellings. On the other hand, Google Translate was correct in its translation of "Northern Territories" of Japan, but VMNET incorrectly chose "Northern Territory" (of Australia). For the rest of the misspelled phrases (*Qingling, Initials D, Kashimir*), neither Google Translate nor VMNET could pick the correct translation.

7.2 IR Evaluation

The MAP values of all experimental runs corresponding to each query processing technique and Chinese indexing strategy are given in Table 5. The results of mono-lingual runs give benchmarking scores for CLIR runs.

As expected, the highest MAP 0.4681 is achieved by the monolingual run VMNET-CS-CS-01-T, in which the questions were manually segmented and all the noise words were removed.

It is encouraging to see that the automatic run VMNET-CS-CS-02-T with only question template phrase removal has a slightly lower MAP 0.4419 than that (0.4488) of the best performance CS-CS run in the NTCIR-8 evaluation forum (Sakai, et al., 2010).

If unigrams were used as the only indexing units, the MAP of VMNET-CS-CS-04-T dropped from 0.4681 to 0.3406. On the other hand, all runs using bigrams as indexing units either exclusively or jointly performed very well. The MAP of run VMNET-CS-CS-05-T using bigrams only is 0.4653, which is slightly

lower than that of the top performer run VMNET-CS-CS-01-T, which used two forms of indexing units. However, retrieval performance could be maximised by using both unigrams and bigrams as indexing units.

The highest MAP (0.3756) of a CLIR run is achieved by run VMNET-EN-CS-03-T, which used VMNET for translation. Comparing it to our manual run VMNET-CS-CS-01-T, there is around 9% performance degradation as a result of the influence of noise words in the questions, and the possible information loss or added noise due to English-to-Chinese translation, even though the named entities translation precision is relatively high.

The best EN-CS CLIR run (MAP 0.4209) in all submissions to the NTCIR-8 ACLIA task used the same indexing technique (bigrams and unigrams) and ranking function (BM25) as run VMNET-EN-CS-03-T but with "query expansion based on RSV" (Sakai, et al., 2010). The MAP difference 4.5% between the forum best run and our CLIR best run could suggest that using query expansion is an effective way to improve the CLIR system performance.

Runs VMNET-EN-CS-01-T and VMNET-EN-CS-04-T, that both used Google Translate provide direct comparisons with runs VMNET-EN-CS-02-T and VMNET-EN-CS-03-T, respectively, which employed VMNET for translation. All runs using VMNET performed better than the runs using Google Translate.

Run Name	MAP
NTCIR-8 CS-CS BEST	**0.4488**
VMNET-CS-CS-01-T	0.4681
VMNET-CS-CS-02-T	**0.4419**
VMNET-CS-CS-03-T	0.4189
VMNET-CS-CS-04-T	0.3406
VMNET-CS-CS-05-T	0.4653
NTCIR-8 EN-CS BEST	**0.4209**
VMNET-EN-CS-01-T	0.3161
VMNET-EN-CS-02-T	0.3408
VMNET-EN-CS-03-T	**0.3756**
VMNET-EN-CS-04-T	0.3449

Table 5. Results of all experimental runs

The different performances between CLIR runs using Google Translate and VMENT is the joint result of the translation improvement and other translation differences. As shown in Table 8, VMNET found the correct translations for 8 more topics than Google Translate.

It should be noted that there are two topics (ACLIA2-CS-0008 and ACLIA2-CS-0088) not included in the final CLIR evaluation (Sakai, et al., 2010). Also, there is one phrase, "Kenneth Yen (K. T. Yen) (严凯泰)", which VMNET couldn't find the correct translation for, but it detected a highly associated term "Yulon - 裕隆汽车", an automaker company in Taiwan; Kenneth Yen is the CEO of *Yulon*. Although *Yulon* is not a correct translation, it is still a good query term because it is then possible to find the correct answer for the question: "*Who is Kenneth Yen?*". However, this topic was not included in the NTCIR-8 IR4QA evaluation.

Moreover, it is possible to have multiple explanations for a term. In order to discover as many question-related documents as possible, alternative translations found by VMNET are also used as additional query terms. They are shown in Table 6. For example, 丁克 is the Chinese term for DINK in Mainland China, but 顶客族 is used in Taiwan. Furthermore, because VMNET gives the Wikipedia translation the highest priority if only one entry is found, a person's full name is used in person name translation rather than the short commonly used name. For example, *Cheney* (former vice president of U.S.) is translated into 迪克·切尼 rather than just 切尼.

NE	VMNET	Wiki Title
Princess Nori	纪宫公主	黑田清子
DINK	丁克	顶客族
BSE	疯牛病	牛海绵状脑病
Three Gorges Dam	三峡大坝	三峡工程

Table 6. Alternative translations

The biggest difference, 3.07%, between runs that used different translation is from runs VMNET-EN-CS-03-T and VMNET-EN-CS-04-T, which both pruned the question template phrase for simple query processing. Although the performance improvement is not obvious, the correct translations and the additional query terms found by VMNET are still very valuable.

8 Conclusions

General machine translation can already achieve very good translation results, but with our proposed approach we can further improve the translation accuracy. With a proper adjust-

ment of this approach, it could be used in a situation where there is a need for higher precision of complex phrase translation.

The results from our CLIR experiments indicate that VMNET is also capable of providing high quality query terms. A CLIR system can achieve good results for answer finding by using the VMNET for translation, simple indexing technique (bigrams and unigrams), and plain question template phrase pruning.

Run Name	Indexing Units	Query Processing
VMNET-CS-CS-01-T	U + B	Manually segment the question and remove all the noise words
VMNET-CS-CS-02-T	U + B	Prune the question template phrase
VMNET-CS-CS-03-T	U + B	Use the whole question without doing any extra processing work
VMNET-CS-CS-04-T	U	As VMNET-CS-CS-01-T
VMNET-CS-CS-05-T	B	As VMNET-CS-CS-01-T
VMNET-EN-CS-01-T	U + B	Use Google Translate on the whole question and use the entire translation as query
VMNET-EN-CS-02-T	U + B	Use VMNET translation result without doing any further processing
VMNET-EN-CS-03-T	U + B	As above, but prune the Chinese question template from translation
VMNET-EN-CS-04-T	U + B	Use Google Translate on the whole question and prune the Chinese question template phrase from the translation

Table 7. The experimental runs. For indexing units, U means unigrams; B means bigrams.

Topic ID	Question with OOV Phrases	Correct	GT	VMNET
ACLIA2-CS-0002	What is the relationship between the movie "*Riding Alone for Thousands of Miles*" and ZHANG Yimou?	千里走单骑	利民为千里单独	千里走单骑
ACLIA2-CS-0008	Who is *LI Yuchun*?	李宇春	李玉春	李宇春
ACLIA2-CS-0024	Why does *Qingling* build "panda corridor zone"	秦岭	宋庆龄	宋庆龄
ACLIA2-CS-0035	Please list the events related to the movie "*Initials D*".	头文字 D	缩写 D 的事件	缩写 D 的事件
ACLIA2-CS-0036	Please list the movies in which *Zhao Wei* participated.	赵薇	照委	赵薇
ACLIA2-CS-0038	What is the relationship between Xia Yu and *Yuan Quan*.	袁泉	袁区广	袁泉
ACLIA2-CS-0048	Who is *Sean Chen(Chen Shin-An)*?	陈信安	肖恩陈（陈新的）	陳信安
ACLIA2-CS-0049	Who is *Lung Yingtai*?	龙应台	龙瀛台	龙应台
ACLIA2-CS-0057	What is the disputes between China and Japan for the undersea natural gas field in the *East China Sea*?	东海	东中国海域	东海
ACLIA2-CS-0066	What is the relationship between two Russian politicians, *Kasianov* and Putin?	卡西亚诺夫	Kasianov	米哈伊尔·米哈伊洛维奇·卡西亚诺夫
ACLIA2-CS-0074	Where are Japan's *Northern Territories* located?	北方领土	北方领土	北领地
ACLIA2-CS-0075	Which countries have borders in the *Kashimir* region?	克什米尔	Kashimir	Kashimir
ACLIA2-CS-0088	What is the relationship between the Golden Globe Awards and *Broken-back Mountain*?	断臂山	残破的背山	断臂山
ACLIA2-CS-0089	What is the relationship between *Kenneth Yen(K. T. Yen)* and China?	严凯泰	肯尼思日元（观塘日元）	裕隆汽车

Table 8. The differences between Google Translate and VMNET translation of OOV phrases in which GT or VMNET was wrong.

References

Chan, Y.-C., Chen, K.-H., & Lu, W.-H. (2007). *Extracting and Ranking Question-Focused Terms Using the Titles of Wikipedia Articles.* Paper presented at the NTCIR-6.

Chen, A., He, J., Xu, L., Gey, F. C., & Meggs, J. (1997, 1997). *Chinese text retrieval without using a dictionary.* Paper presented at the SIGIR '97: Proceedings of the 20th annual international ACM SIGIR conference on Research and development in information retrieval.

Chen, A., Jiang, H., & Gey, F. (2000). Combining multiple sources for short query translation in Chinese-English cross-language information retrieval. 17-23.

Ferrández, S., Toral, A., Ferrández, Ó., Ferrández, A., & Muñoz, R. (2007). Applying Wikipedia's Multilingual Knowledge to Cross–Lingual Question Answering *Natural Language Processing and Information Systems* (pp. 352-363).

Lu, C., Xu, Y., & Geva, S. (2007). Translation disambiguation in web-based translation extraction for English–Chinese CLIR. 819-823.

NTCIR Project. (2010). Tools. from http://research.nii.ac.jp/ntcir/tools/tools-en.html

Robert, W. P. L., & Kwok, K. L. (2002). A comparison of Chinese document indexing strategies and retrieval models. *ACM Transactions on Asian Language Information Processing (TALIP), 1*(3), 225-268.

Sakai, T., Shima, H., Kando, N., Song, R., Lin, C.-J., Mitamura, T., et al. (2010). *Overview of NTCIR-8 ACLIA IR4QA.* Paper presented at the Proceedings of NTCIR-8, to appear.

Shi, L., Nie, J.-Y., & Cao, G. (2008). *RALI Experiments in IR4QA at NTCIR-7.* Paper presented at the NTCIR-7.

Su, C.-Y., Lin, T.-C., & Wu, S.-H. (2007). *Using Wikipedia to Translate OOV Terms on MLIR.* Paper presented at the NTCIR-6.

Tatsunori Mori, K. T. (2007). *A method of Cross-Lingual Question-Answering Based on Machine Translation and Noun Phrase Translation using Web documents.* Paper presented at the NTCIR-6.

The Stanford Natural Language Processing Group. (2010). Stanford Log-linear Part-Of-Speech Tagger. from http://nlp.stanford.edu/software/tagger.shtml

University of Pennsylvania. (2010). POS tags. from http://bioie.ldc.upenn.edu/wiki/index.php/POS_tags

Zhang, Y., & Vines, P. (2004). *Using the web for automated translation extraction in cross-language information retrieval.* Paper presented at the Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval.

Zhang, Y., Vines, P., & Zobel, J. (2005). Chinese OOV translation and post-translation query expansion in Chinese–English cross-lingual information retrieval. *ACM Transactions on Asian Language Information Processing (TALIP), 4*(2), 57-77.