Utilizing Entity Relation to Bridge the Language Gap in Cross-Lingual Question Answering System

Min Wu, Tomek Strzalkowski
Institute of Informatics, Logics and Security Studies
University at Albany, SS 261, 1400 Washington Avenue, NY 12222, USA
minwu@cs.albany.edu, tomek@csc.albany.edu

Abstract

We describe University at Albany's CLQA system and its performance in English-Chinese subtask evaluation in NTCIR-6 CLQA. Firstly we illustrate our submitted system, which was built in two weeks. (We had to finish our CLQA system in this time limit because we were late registered.) Then we would like to introduce the improved system which utilizes our ACE (Automatic Content Extraction) relation detection and recognition (RDR) system to help bridge the language gap when answering some types of questions. The experimental results show that our proposed method helps to improve the system performance in answering questions about some specific relation between entities.

Keywords: *English-Chinese Cross-Lingual Question Answering; entity relation; relation detection and extraction; ACE*

1. Introduction

Cross-Lingual Question Answering (CLQA) has been proposed as evaluation task since the startup of Cross Language Evaluation Forum (CLEF, http://www.clef-campaign.org) and NII-NACSIS Test Collection for IR Systems (NTCIR, http://research.nii.ac.jp/ntcir/index-en.html). [5] [6] Given a question in one language (source language) and data corpus in another language (target language), an automatic CLQA system is able to generate answers in source language. According to the evaluation report from CLEF and NTCIR in recent years, the performance of CLQA system still lags behind the monolingual QA system. The language gap between the source language and target language is one of the barriers to improving the performance of CLQA system up to the level of monolingual QA system.

Our experience in building English-Chinese CLQA system suggests that the language gap occur in question translation as a result of the following: translation ambiguity and inability to translate named entities. These two negative effects lead to poor document retrieval results and thus lower the performance of following QA components. So, we believe that solving these two aspects of question translation is the key to improving the accuracy of CLQA performance.

In the following sections, we firstly illustrate the architecture of our baseline system submitted to NTCIR-6 CLQA E-C subtask and the methods we used in the development. Then we introduce the background and examples that lead us to utilizing the ACE entity relation extraction methods. We then compare the performance of the baseline system with that of the improved system and conclude with discussion of how to employ more types of ACE entity relation and ACE event to the development of CLQA system.

2. Baseline English-Chinese QA system

Our baseline system consists of the following components: question translation, question analysis, document retrieval, passages filtering and answer extraction. [7] [8] The architecture is illustrated in Figure 1.

2.1 Question Translation

Question translation is the first step of the whole CLQA system. It is also the key component because the accuracy of question translation affects the performance of succeeding system components. We employed AltaVista's online machine translation tool Babel Fish (http://babelfish.altavista.com/) after comparison of sample question translation results with other machine translation tools such as Google. From our comparison, Babel Fish translation generate better

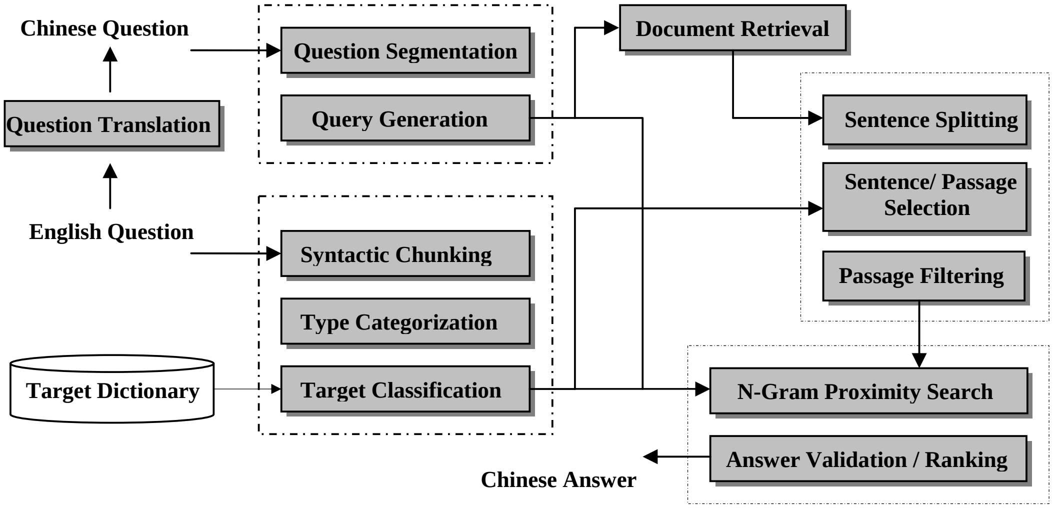


Figure 1. Architecture of baseline CLQA system

translation of common English words, however Google is better at translation of named entities. We didn’t employ both translation tools because sometimes it is hard to choose which one is better among the two translation results. The translated question is segmented. The token words are filtered and organized as query.

2.2 Question Analysis

The rule-based question analysis component is imported from our monolingual QA system ILQUA which has participated in TREC QA track for three years. Question analysis classifies English questions by syntactic structure and answer target. We used the parser developed by Stanford University NLP group. (http://nlp.stanford.edu/downloads/lex-parser.shtml) Syntactic chunking splits question into a list of question terms with syntactic tags. For example, the question “What company is South Korea’s No.1 carmaker?” will be chunked with the syntactic structure of “What_NP_Be_NP_POS_NP”. For questions beginning with the words “When”, “Where” and “Who”, the answer target assigned is “Date”, “Location” and “Person” respectively. For questions beginning with pattern “How+Adj.”, there are hand-crafted rules to assign answer target according to the adjectives. For questions beginning with “What_Be”, “What_NP”, “Which_Be”, “Which_NP”, the key term of noun phrase “NP” is mapped to appropriate answer target type. We set up a noun-target map of 7917 entries to correlate nouns with named entity types which can be processed by the system. The assigned entity type is set as the major answer target type and the noun is set as the minor answer target type. For example, if the major answer target is “Organization”, the minor answer target could be “company”, “university”, “party”, “team” etc. This two-level answer target categorization is helpful to answer validation.

2.3 MSRSeg and Name Entity Tagging

Since the answer extraction of our CLQA system is deployed on NE-tagged text with N-gram proximity searching, the plain-text data corpus must be pre-processed by NE-tagging tools. The NE-tagging tool used in our E-C CLQA system is Microsoft Chinese word segmenter and NE tagger (MSRSeg) developed by Microsoft Research Asia. [2] This pragmatic system consists of a generic segmenter that is based on the statistical framework of word segmentation and unknown word detection, and a set of output adaptors for adapting the output of the segmenter to different application-specific standards such as named entities and factoids.

2.4 Document Retrieval & Passage Filtering

The IR engine is built on Lucene augmented with Chinese segmenter. The top 50 document names retrieved by Lucene are used to locate corresponding NE-tagged documents. These documents are split into sentences and filtered by question target and question terms.

2.5 Answer Extraction, Validation & Ranking

Tokenized question terms are matched as n-grams around every named entity in the filtered candidate sentences. We match the longest possible sequence of tokenized word within the 100 word sliding window around each named entity. Once a sequence is matched, the corresponding word tokens are removed from the token list and the same search and matching process is repeated until the token list is empty or no sequence can be matched. The candidate named entity is scored by the average weighted distance score of matched question terms.

Let $Num(t_i...t_j)$ denote the number of all matched n-grams, $d(E, t_i...t_j)$ denote the word distance between the named entity and the matched n-gram, $W(t_i...t_j)$ denote the weight of the matched n-gram. The value assigned to weight W is determined by λ, the ratio of matched n-gram length to question term length as follows:

$$
\begin{array}{ll}
W(t_i...t_j)=0.4 & \text{if } \lambda<0.4 \\
W(t_i...t_j)=0.6 & \text{if } 0.4 \leq \lambda<0.6 \\
W(t_i...t_j)=0.8 & \text{if } \lambda>0.6 \\
W(t_i...t_j)=0.9 & \text{if } \lambda<0.75
\end{array}
$$

The weighted distance score $D(E,QTerm)$ of the question term and the final score $S(E)$ of the named entity are calculated as follows:

$$D(E, QTerm) = \frac{\sum_{t_i...t_j} \frac{d(E, t_i...t_j)}{W(t_i...t_j)}}{Num(t_i...t_j)}$$

$$S(E) = \frac{\sum_{i}^{N} D(E, QTerm_i)}{N}$$

After the n-gram proximity search, the matched named entities are sorted by their scores and organized in an answer candidate list. Then, the succeeding answer validation process filters out ill-formatted or not-in-category answer candidates. The remaining candidates are re-scored with base proximity score S(E) and frequency score.

3. Issues in Question Translation

Let's now take a closer look back at the "language gap" problem mentioned before. To illustrate the problem better, we consider an example question from NTCIR6 E-C CLQA subtask "Which Taiwanese party does Shui-bian Chen belong to?" We submit this question to two up-to-date online translation tools and in each case we got inaccurate results.

Question in English:
Which Taiwanese party does Shui-bian Chen belong to?

Translated Question in Chinese:
a. (By Alta Vista Babel Fish)
哪个台湾党水 bian 陈属于?
b. (By Google)
其中台湾一方当事人陈水扁同属?

Babel Fish (translation a) can not translate named entity "Shui-bian Chen" into corresponding Chinese name. Google Translation Tool translated the person name correctly, however stumbled into translation ambiguity. It translated term "party" as "当事人" (meaning: a person or group taking one side of a question, dispute, or contest), the correct translation of "party" in this question context should be "党"(meaning: a group of persons organized for the purpose of directing the policies of a government).

Although a lot of research effort [3] [4] has been devoted to improve the accuracy of machine translation, such problems as translation ambiguity, out-of-dictionary still plague even the most up-to-date machine translation tools.

In the example question, although it is difficult to translate "Shui-bian Chen" from English to correct Chinese, it is much easier to translate corresponding Chinese unit "陈水扁" into English. Can we utilize this "unbalanced translation difficulty" to help bridge the language gap during question translation? If some organization affiliation facts about "陈水扁" have already been extracted from Chinese data corpus and some simple translation of the named entity "陈水扁" has been deployed prior to the question translation, then this information will be very helpful to answer the question. In this way, the system performs translation in a combined way: source-to-target direction on question and target-to-source direction on candidate answers in data corpus.

The above method is viable as long as the extraction and translation of facts attain some reasonable performance. Thus we propose a method based on entity relation extraction, entity translation and bi-directional relation searching and matching.

4. Utilizing Entity Relation in CLQA

Many NTCIR CLQA questions seek relations between entities. For example, the answer to the question "Which Taiwanese party does Shui-bian Chen belong to?" is an affiliation between an organization and a person which is defined as Org-Affiliation relation in ACE.

4.1 ACE Entity & Relation

The ACE (Automatic Content Extraction, http://projects.ldc.upenn.edu/ace/) program conducted by NIST aims to develop automatic content extraction technology to support the automatic processing of language data. ACE tasks focus on extracting entity, time, value, relation and event. Currently ACE defines 7 types of entities (Person, Organization, Location, Geo-Political Entity, Facility, VEH and Weapon). An ACE relation is a relation between two ACE entities, which are called the relation arguments. ACE relation types are Org-affiliation, Gen-Affiliation, Part-whole, Physical, Person-Social, Artifact etc.

4.2 Entity Relation in NTCIR Question

To investigate how entity relations can be utilized in CLQA, we mainly focus on three types of relations: ORG-AFF, GEN-AFF and PART-WHOLE. These relations frequently occur in both ACE training data and in NTCIR English-Chinese test questions. Among the 150 questions in NTCIR6 CLQA E-C subtask, 41 questions concern ORG-AFF relation and 11 questions concern GEN-AFF relation and 2 questions concern PART-WHOLE relation.

Some example questions from NTCIR6 questions are in the table below. The English entity relation in question can be easily detected and extracted with some predefined rules because question is usually expressed as a short sentence and has well-organized syntactic format. The relation is translated into Chinese with the aid of bilingual English-Chinese dictionary. When the translation ambiguity occurs, we refer to the translated question from Babel Fish and Google translation tool. If some English phrase can't be translated by either method, we just ignore it.

English Question: *Which Taiwanese party does Shui-bian Chen belong to?*
Entity Relation in English and Chinese: **Shui-bian Chen – (ORG-AFF) –Taiwanese party** **Shui-bian Chen – (ORG-AFF) --** 台湾 党
English Question: *What is Japan's largest car maker?*
Entity Relation in English and Chinese: **largest car maker – (GEN-AFF) – Japan** 最大的 汽车 制造者 **– (GEN-AFF) --** 日本
English Question: *What is the second largest city in Japan?*
Entity Relation in English and Chinese: **second largest city – (PART-WHOLE) – Japan** 第二 最大的 城市 **– (PART-WHOLE) --** 日本

4.3 Relation Extraction in Chinese Document

We employed a machine-learning based method to perform entity relation extraction. [1] [9] [10] Figure 2 shows the framework of the relation extraction sub-system.

Document Preprocessing: The raw SGML texts are parsed and processed with MSRSeg. In the processed texts, Chinese word boundary is marked, NEs and factoids are tagged. In addition, the metadata such as the document ids and paragraph ids are saved for later reference.

Entity Detection and Recognition: To simplify the problem, we reduce EDR to a three-step classification process. Since ACE entities have a wider coverage than the NEs tagged by MSRSeg, the first step is entity boundary detection, that is, the system tries to identify all possible ACE entity candidates in a Chinese text segment. This can be done with the aid of NEs tagged by MSRSeg and entity trigger word list. Each named entity tagged by MSRSeg is directly placed into the entity candidate list. At the same time, the system also examines if there is an ACE entity trigger word within a sliding window of 7 words around the tagged name entity. The phrase activated by the trigger word is also added into the entity candidate list. The trigger word list is built ahead of time from the ACE training data. All the ACE entities in the training data are extracted, examined manually and trimmed manually if necessary. We choose the coverage of sliding window as 7 words based on the observation of the training data. (This is somewhat arbitrarily and we would like to try better methods later.)

The system collects features for each entity candidate, the features we used include bag of words feature of entity head word, entity extent, the words preceding the entity and the words succeeding the entity, POS of each word, NE types (optional). To collect the bag of words feature, we build a dictionary of most common Chinese words to keep a reasonable length of SVM feature vector. The current performance of Chinese parser is still not satisfactory when applied to long and complex sentences and parsing every sentence in the document is too time-consuming. Therefore we use a Chinese syntactic dictionary to assign the POS of each word. This syntactic dictionary is built from different downloaded web data source and ACE training data.

The last subtask in EDR is to classify and label the entity candidates. Instead of doing one-versus-all classification, we did the two-level SVM classification. The first-level is multivariate classification process which determines the entity type and the second-level is a multiple classification which decides entity-subtype and mention type.

Relation Detection and Recognition: The RDR component implemented in similar way as the EDR component. Firstly, the system searches all possible combinations of ACE entity pairs that could be linked as an ACE relation candidate within a sentence. (In our system, we just consider entity pairs within one sentence to simplify the implementation.) Sometimes it is very obvious that some entity pairs will not be formed as an ACE relation by the definition of ACE requirements. We added some pragmatic rules to filter out these entity pairs to simplify the task of SVM classification.

The system collects features of the relation candidates. The features collected are: bag of words of each attribute, attribute's ACE entity type, attribute

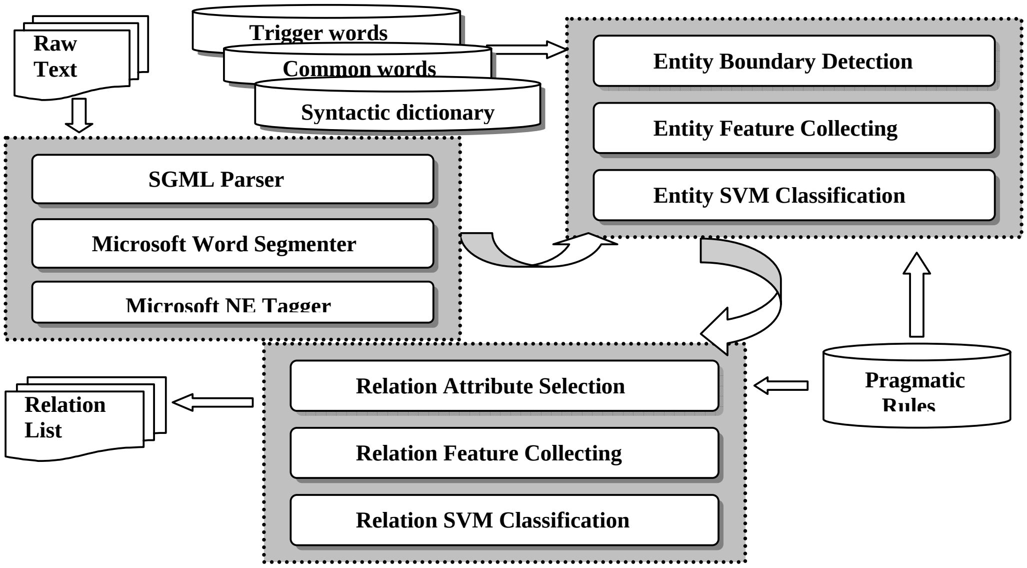

Raw Text
Trigger words
Common words
Syntactic dictionary
Entity Boundary Detection
Entity Feature Collecting
Entity SVM Classification
SGML Parser
Microsoft Word Segmenter
Microsoft NE Tagger
Relation List
Relation Attribute Selection
Relation Feature Collecting
Relation SVM Classification
Pragmatic Rules

Figure 2. Entity relation extraction based on machine learning

position order etc. Similar to EDR, the succeeding process is also a two-level classification to determine the relation type and subtype separately.

4.4 Relation translation

The extracted relations from Chinese data corpus need to be translated before they are postulated into a database. Since we have not developed an automatic machine translation tool for entity translation, the two entities in each relation can not be fully translated. Currently as a tentative resolution, our system only translates Chinese name, organization name and location name contained in the entity. Chinese names can be translated into English by their pronunciation with pre-defined rules. These rules here refer to the different pronunciation and spelling between China mainland, HongKong and TaiWan. Organization names and location names are translated with the aid of Chinese-English bilingual organization dictionary and location dictionary. These bilingual dictionaries are manually collected and well organized in a machine readable format. However, this method only works out for well-known organizations, locations and countries that included in the dictionaries.

4.5 Database population

The database management system used is MySql. Each relation type gets a database table created and the data is loaded from the output file of the relation extraction subsystem. Each database entry corresponds to the description information of an entity relation. The entry consists of document ID, paragraph

<table>
<tr><th>Doc Id</th><th>Passage Id</th><th>Entity1</th><th>Entity1*</th><th>Entity2</th><th>Entity2*</th><th>Relation Head</th><th>Relation Head*</th></tr>
<tr><td colspan="3">udn xxx 19980927 0164</td><td>3</td><td colspan="3">台北市</td><td>Taipei</td></tr>
<tr><td colspan="3"></td><td></td><td colspan="3">陳水扁</td><td>Shui-bian Chen</td></tr>
<tr><td colspan="3"></td><td></td><td colspan="3">市長</td><td></td></tr>
<tr><td colspan="3">udn xxx 19981122 0291</td><td>2</td><td colspan="3">民進黨</td><td>Democratic Progressive Party</td></tr>
<tr><td colspan="3"></td><td></td><td colspan="3">陳水扁</td><td>Shui-bian Chen</td></tr>
<tr><td colspan="3"></td><td></td><td colspan="3">台北市 長 候選人</td><td>Taipei</td></tr>
<tr><td colspan="3">udn xxx 19990803 0288</td><td>3</td><td colspan="3">海峽交流基金會</td><td>Straits Exchange Fundation</td></tr>
<tr><td colspan="3"></td><td></td><td colspan="3">辜振甫</td><td>Gu Zhenfu / Koo Chen-fu</td></tr>
<tr><td colspan="3"></td><td></td><td colspan="3">董事長</td><td>President / Chairman</td></tr>
<tr><td colspan="3">udn xxx 19990808 0210</td><td>1</td><td colspan="3">美國 國防部</td><td>USA Department of Defense</td></tr>
<tr><td colspan="3"></td><td></td><td colspan="3">坎柏</td><td></td></tr>
<tr><td colspan="3"></td><td></td><td colspan="3">助理 副 部長</td><td></td></tr>
</table>

Figure 3. Database entry samples

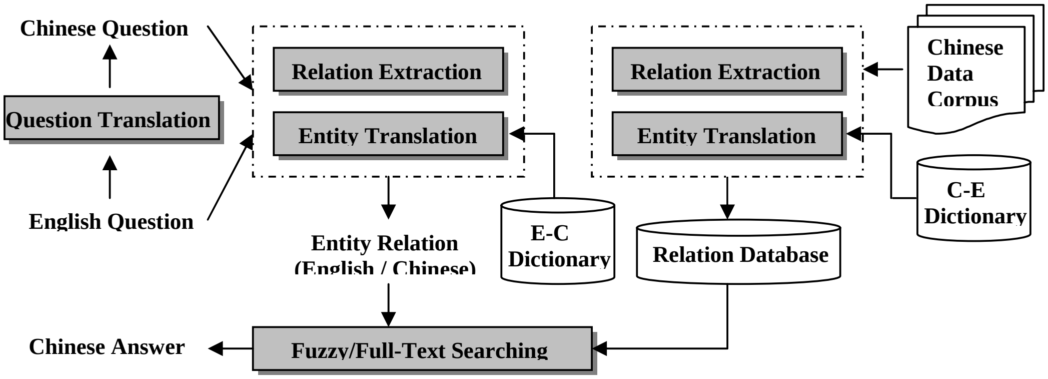


Figure 4. CLQA system utilizing entity relation

ID, entity, head word of the relation, translation of entity and translation of the relation head word. Since the entities in relation are not fully translated (see section 4.4), some attributes may be empty. Figure3 shows some example entries from the relation database.

4.6 Relation Searching

Given a NTCIR question, the entity relation in this question is extracted and represented in both English and Chinese. The key step is to search and match the question relation in the relation database built from Chinese data corpus.

In our experiment, we did two different searches: i). One-way search is the process of matching Chinese question relation (Q-Rel-Ch) against Chinese attributes of database entry (D-Rel-Ch). ii). Two-way search is the process of matching both Chinese and English question relations (Q-Rel-Ch and Q-Rel-En) against the corresponding Chinese and English attributes of database entry (D-Rel-Ch and D-Rel-En). The database search operation must be fuzzy and full-text searching because the exact search will miss a lot of matching entries. MySql 5.0 provides both fuzzy and full-text search functions.

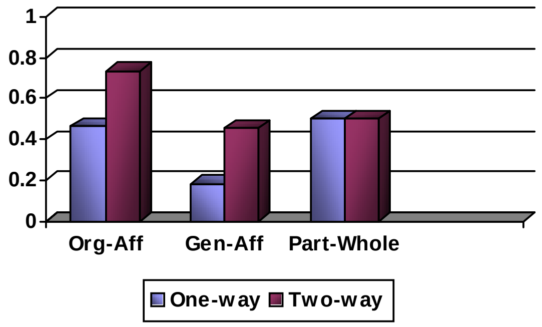
5. Experiment and evaluation results

5.1 One-way search vs. two-way search

The answers to some questions can be extracted via one-way search. For example, test question “T3016: Who was the president of Korea in 1999?” is parsed and the entity relation in this question is an Org-Affiliation relation “Korea – president – person”. This English question relation is translated into Chinese “韓國 – 總統 --人”. It is not difficult to form a SQL search statement from this Chinese relation and match it against the relation database. The SQL statement is “SELECT * FROM orgaff WHERE entity1zh LIKE '%韓國%' AND head1zh LIKE '%總統%'. In some cases, however, the Chinese question relation doesn’t perfectly match the database entry. For example, from test question “T3143: Who is the secretary-general of the Environmental Quality Protection Foundation?” system extracts Org-Affiliation relation “the Environment Quality Protection Foundation – secretary-general – person”. It is translated into “環境質量保護基礎 -- 秘書長 -- 人”. If system uses SQL statement “SELECT * FROM orgaff WHERE entity1zh LIKE '%環境質量保護基礎%' AND head1zh LIKE '%秘書長%'” to search the database, we can’t get any result returned. In such situation, N-grams of entity1 is used to form another new SQL statement “SELECT * FROM orgaff WHERE entity1zh LIKE '%環境%' AND phrase1zh LIKE '%秘書長%'” and the passages containing the correct answer are returned.

One-way searching can’t handle questions with incorrect entity translation. For example, test question “T3081: Which Taiwanese party does Shui-bian Chen belong to?" contains an Org-Affiliation relation “Taiwanese party – null – Shui-bian Chen”. In this relation, named entity “Shui-bian Chen” can’t be translated correctly with either the bilingual dictionary or machine translation tool. In this case, our system just ignore it. The partially translated relation is “台湾党 – null -- Shui-bian Chen”. To match this question relation against the database, the system did one-way search with SQL statement “SELECT * FROM orgaff WHERE entity1zh LIKE '%台湾党%'”, followed by search with SQL statement “SELECT * FROM orgaff WHERE entity1zh LIKE '%台湾%'”, and finally with “SELECT * FROM orgaff WHERE entity1zh LIKE '%党%'”. Unfortunately none of the above operations works; they either return no results or too many results to continue the process. In order to perform two-way search, the system uses the English named entity “Shui-bian Chen” and forms SQL statement “SELECT * FROM orgaff WHERE entity1zh LIKE '%党%' AND entity2en LIKE '%Shui-bian Chen%'”. This query returns a small number of related database entries for later processing.

The following chart shows the comparison of the top-1 answer (the first answer returned by the system) accuracy of the two search methods work on questions containing different relation types.



5.2 Evaluation results of different systems

We compare the top-1 answer accuracy of three systems: the baseline system, the system utilizing relation database populated from Chinese data corpus and a combined system of both. The results are shown in Table2. The top-1 answer accuracy is improved from 8.67% to 24% when relation database is utilized and again improved to 28.67% with the third system.

6. Conclusion and discussion

In this paper, we proposed a method to utilize entity extraction to populate a relation database and two-way entity relation search to help bridge the language gap during the question translation. This method exploits the "un-balanced translation difficulty" between the source language (English) and target language (Chinese) by trying to sidestep the hard problem of "translating English named entity to Chinese named entity" by the replacing it with the easier translation of "Chinese named entity to English named entity". This approach works well on questions containing specific types of relations such as Org-Affiliation, Gen-Affiliation and Part-Whole. Currently our system only covers these three types of relations because our ACE relation extraction system performs relatively well on these relations.

More work need to be done in the future development. From the evaluation results, the accuracy of question type of ARTIFACT, DATE, MONEY and NUMEX are still very low. We are considering utilizing more types of relations (or maybe events) to handle these types of questions. For example, some DATE questions concern political events, sport events and natural disaster etc. Some NUMEX questions and MONEY questions concern countries and organizations. Further improvement of the performance of our ACE relation extraction system and the availability of better bilingual dictionaries to help the entity translation are also necessary.

From the experience in developing our NTCIR CLQA system, we feel that the language gap occurring during the question translation can be bridged in multiple ways and sometimes the system can't choose the most appropriate one. Thus the difficulty lies in how to integrate these methods into one stable and effective strategy to boost the CLQA system performance to the same level as the monolingual QA system.

Acknowledgement

Thanks to Dr. Jianfeng Gao from Microsoft Research to provide the Chinese segmenter and NE tagger MSRSeg. Also we would like to thank ACE organizers to provide the ACE training data and evaluation tool.

Reference

[1] A. Culotta, J. Sorensen. "Dependency Tree Kernels for Relation Extraction". *In Proceedings of ACL-2004*, 2004

[2] J. F. Gao, A. D. Wu, C. N. Huang, H. Q. Li, X. S. Xia, H. W. Qin. "Adaptive Chinese Word Segmentation". *In Proceedings of ACL-2004*, 2004

[3] J. F. Gao, J. Y. Nie, E. D. Xun, J. Zhang, M. Zhou, C. N.

Table 2: Comparison of System Performance Based on Top1 Accuracy

		Baseline		Relation		Baseline + Relation	
QType	**#Q**	**#Top-1**	**Accuracy**	**#Top-1**	**Accuracy**	**#Top-1**	**Accuracy**
ARTIFACT	7	0	0.0000	0	0.0000	0	0.0000
DATE	39	4	0.1026	0	0.0000	4	0.1026
LOCATION	16	1	0.0625	4	0.25	5	0.3125
MONEY	8	1	0.1250	0	0.0000	1	0.1250
NUMEX	11	0	0.0000	0	0.0000	0	0.0000
ORGANIZATION	16	2	0.1250	6	0.375	7	0.4375
PERCENT	4	0	0.0000	0	0.0000	0	0.0000
PERSON	47	5	0.1064	26	0.5532	26	0.5532
TIME	2	0	0.0000	0	0.0000	0	0.0000
Total	150	13	0.0867	36	0.2400	43	0.2867

Huang. “Improving Query Translation for Cross-Language Information Retrieval using Statistical Models”. *In proceedings of SIGIR-2001*, 2001.

[4] J. F. Gao, J. Y. Nie, H. Z. He, W. J. Chen, M. Zhou. “Resolving Query Translation Ambiguity using a Decaying Co-occurrence Model and Syntactic Dependency Relations”. *In proceedings of SIGIR-2002*, 2002.

[5] K. L. Kwok, P. Deng, N. Dinstl, S. Choi. “NTCIR-5 English-Chinese Cross Language Question-Answering Experiments using PIRCS”. *In proceedings of NTCIR-5 Workshop Meeting”.* December 2005, Tokyo, Japan.

[6] Y. Sasaki. “Overview of the NTCIR-5 Cross-Lingual Question Answering Task (CLQA1)”. *In proceedings of NTCIR-5 Workshop Meeting”.* December 2005, Tokyo, Japan.

[7] M. Wu, T. Strzalkowski. “Utilizing Co-Occurrence of Answers in Question Answering“. *In proceedings of ACL-2006,* July 2006, Sydney, Australia.

[8] M. Wu, T. Strzalkowski. “ILQUA--An IE-Driven Question Answering System”. *In proceedings of TREC 2005.* November 2005, Washington DC.

[9] M. Zhang, J. Zhang, J. Su. “*Exploring Syntactic Features for Relation Extraction using a Convolution Tree Kernel*”. In Proceedings of HLT-2006, June 2006. New York City.

[10] S. B. Zhao, R. Grishman. “*Extracting Relations with Integrated Information Using Kernel Methods*”. In Proceedings of ACL-2005, 2005.