Using Wikipedia and Semantic Resources to Find Answer Types and Appropriate Answer Candidate Sets in Question Answering

Po-Chun Chen, Meng-Jie Zhuang, and Chuan-Jie Lin
Department of Computer Science and Engineering
National Taiwan Ocean University
No 2, Pei-Ning Road, Keelung, Taiwan ROC
{pcchen.cse, mjzhunag.cse, cjlin}@ntou.edu.tw

Abstract

This paper proposes a new idea that uses Wikipedia categories as answer types and defines candidate sets inside Wikipedia. The focus of a given question is searched in the hierarchy of Wikipedia main pages. Our searching strategy combines head-noun matching and synonym matching provided in semantic resources. The set of answer candidates is determined by the entry hierarchy in Wikipedia and the hyponymy hierarchy in WordNet. The experimental results show that the approach can find candidate sets in a smaller size but achieve better performance especially for ARTIFACT and ORGANIZATION types, where the performance is better than state-of-the-art Chinese factoid QA systems.

1 Introduction

1.1 Motivation

Answer type is the semantic category of an expected answer to a given question. Typical QA systems use different strategies to deal with different answer types (Allam and Haggag, 2012). If an answer type is a named entity type such as PERSON or LOCATION, a named entity recognition system (NER) is usually used to identify person names or location names as answer candidates.

NER has been a success for PERSON and LOCATION types (Nadeau and Sekine, 2007), but not for other NE types, especially ARTIFACT such as movies or songs. There are too many ARTIFACT types and most of them are difficult to be automatically recognized.

This paper proposed an alternative way to decide the answer type and the set of answer candidates at the same time. An answer type can be a Wikipedia category or a term in WorNet. Answer candidates are Wikipedia entry titles. By doing so, question answering can be no longer restricted by the ability of NER systems. The set of answer candidates can also be up-to-date since Wikipedia is frequently maintained. Although this study was done on Chinese datasets, our methods are mostly automatic and it is not hard to find comparable semantic resources in different languages. Adapting our methods to another language is possible.

1.2 Related Work

Question answering (QA) has been studied since 1990s. Large-scale benchmarks developed by international evaluation projects improved the performance of QA techniques in a great deal. Since 1999, TREC (**T**ext **RE**treival **C**onference) has held QA tracks for several times dealing with English monolingual question answering (Dang *et al*., 2007). NTCIR (**N**II **T**estbeds and **C**ommunity for **I**nformation access **R**esearch) dealt with multilingual QA in Japanese and Chinese (Sasaki *et al*., 2007), while CLEF supported multilingual QA in European languages (Peñas *et al*., 2014).

Proceedings of the Open Knowledge Base and Question Answering (OKBQA) Workshop, pages 1–10, Osaka, Japan, December 11 2016.

Two benchmarks on Chinese QA have been developed in NTCIR-5 CLQA1 (Sasaki *et al*., 2005) and NTCIR-6 CLQA2 tracks (Sasaki *et al*., 2007). Totally 350 Chinese questions with answers have been created. They are all factoid questions. Complex questions were studied in NTCIR-7 and NTCIR-8 (Sakai *et al*., 2010).

Most QA systems predefined several answer types and used different approaches to identify candidates of answers. Some used semantic resources (Harabagiu *et al*., 2006; Moldovan *et al*. 2007) and others used named entity recognition (NER) systems (Lee *et al*. 2007; Kwok *et al*. 2007; Lee *et al*., 2008; Sacaleanu *et al*., 2008). The ability of NER systems will affect the performance of QA systems.

Wikipedia-based QA is also a hot topic. Most research groups treated Wikipedia as a knowledge base (Furbach *et al*., 2008; Waltinger *et al*., 2008). They analyzed sentences in Wikipedia articles to find answers. Buscaldi and Rosso (2006) mapped common answer types to top-level Wikipedia categories in order to verify answers. Their method uses coarse-grained answer types, while ours focuses on fine-grained answer types. The closest research to our work was done by Adafre and van Genabith (2008), but they treated the substring matching between Wikipedia categories and answer types in WordNet as a scoring feature. They did not use the whole hierarchy of WordNet nor Wikipedia, either.

This paper is organized as follows. Section 2 describes the proposed approach to determine answer types by Wikipedia and semantic resources. Section 3 explains how to determine answer candidate sets. Section 4 discusses the experimental results and Section 5 concludes this paper.

2 Answer Type Determination

Answer type is the semantic category of the information that a question is asking for. It is usually the semantic category of the sense described in a question focus.

Question focus of a question is the longest noun phrase (NP) which describes the expected answer. It can be the interrogative noun phrase (WHNP) without the interrogative word, such as “日本城市” (*Japanese city*) in the question “二次世界大戰時[哪個日本城市]遭投原子彈” ([*Which Japanese city*] *was atomic-bombed during World War II*, where the WHNP is bracketed and the question focus is underlined). It can also be the complement NP of a copula in a question, such as “一九九九年時國際足協主席” (*the president of FIFA in 1999*) in the question “誰是一九九九年時國際足協主席” (*Who was the president of FIFA in 1999*).

Wikipedia is a collaborative encyclopedia contributed by real users around the world. Each Wikipedia entry is often classified into several categories by its authors. These categories are also user-created, so are the hierarchical relationships between the categories. Here is an example of the semantic hierarchy where the Chinese Wikipedia entry “微軟” (Microsoft) belongs to:

Entry: 微軟 (Microsoft)
→ Category: 微軟 (Microsoft)
→ Category: 美國軟體公司 (Software companies of the United States)
→ Category: 各國軟體公司 (Software companies by country)
→ Category: 軟體公司 (Software companies)
→ Category: 科技公司 (Technology companies)
→ Category: 各行業公司 (Companies by industry)
→ Category: 各類公司 (Companies by type)
→ Category: 各類組織 (Organizations by activity)
→ Category: 組織 (Organizations)
→ Category: 社會 (Society)
→ Category: 頁面分類 (Fundamental categories)

As we can see, if we know the answer type of a given question is “軟體公司” (software company), all Wikipedia entries under that category, such as “微軟” (Microsoft), are appropriate answer candidates. We will discuss different methods to extract the longest Wikipedia category title from a given question focus in the following subsections.

2.1 Maximum Matching Strategy

The first straightforward method to extract an answer type from a question focus is to identify a Wikipedia category title by maximum matching algorithm. But because all these strings are noun phrases, the matched substring must also be a meaningful head of the question focus. This can be ensured by syntactic structure (such as removing of prepositional phrases) or trailing-matching strategy (i.e. matching the longest trailing substring). This kind of expected answer type will be referred to as ***Wikipedia-category answer type*** (WKtype) throughout this paper. Two examples are given as follows.

Q1: 一九九九年時聯合國秘書長是誰？
(Who was Secretary-General of the United Nations in 1999?)
QFocus: 一九九九年時聯合國秘書長 (United Nations Secretary-General in 1999)
WKtype: 聯合國秘書長 (United Nations Secretary-General)

Q2: 微軟公司推出的辦公室套裝軟體叫什麼?
(What is the name of the office software suite produced by Microsoft?)
QFocus: 微軟公司推出的辦公室套裝軟體 (the office software suite produced by Microsoft)
WKtype: 軟體 (software)

In both examples, the matched Wikipedia category titles ("聯合國秘書長" and "軟體") are trailing substrings of the question foci (denoted by QFocus). Sometimes the question focus itself is a Wikipedia category title.

As a backing method, we also define the maximum matching of a WordNet term in a question focus to be its ***WordNet-term answer type*** (WNtype). We use an extension version to develop our QA system, which was the Traditional Chinese version WordNet[1] extended by adding synonyms collected in the Extended Version of Tongyici Cilin[2] (同義詞詞林擴展版), a thesaurus collecting large sets of Chinese synonyms. In the two examples above, their WordNet-term answer types and their Wikipedia-category answer types happen to be the same.

2.2 Synonym Substitution and Maximum Matching

An important issue of maximum matching is the paraphrase problem. The maximum matching might fail to catch the longest one if a question focus is written in an expression different from a synonymous Wikipedia category title.

To solve such a problem, we proposed two different methods to substitute synonyms in a question focus and perform maximum matching as usual. The two methods used different semantic resources explained as follows. Sales *et al*. (2016) dealt with this problem by decomposing a category name into core+modifiers and measuring the similarity with word2vector (Mikolov *et al*., 2013). It is possible to adopt their methods in the future.

Tongyici Cilin synonym substitution

First, all Tongyici Cilin terms in the question focus are identified. By substituting these Cilin terms with their synonyms, a lot of new QFocus strings can be enumerated. The longest Wikipedia category title that can be matched in these new QFocus strings is the final decision, which we will refer to as the ***Cilin-rephrased Wikipedia-category answer type*** (CKtype) throughout this paper. For example,

Q3: 哪家是一九九八年最大的行動電話製造商？
(What was the biggest mobile phone manufacturer in 1998?)
QFocus: 行動電話製造商 (mobile phone manufacturer)
↓ 行動電話 = 手提電話 *in Tongyici Cilin*
CKtype: 手提電話製造商 (mobile phone manufacturers)
WKtype: N/A

[1] http://cwn.ling.sinica.edu.tw/ *and* http://lope.linguistics.ntu.edu.tw/cwn/
[2] http://ir.hit.edu.cn/ *and* http://www.ltp-cloud.com/

In this example, its WKtype cannot be determined because no matching of Wikipedia categories can be found. But by substituting "行動電話" (mobile phone) with its synonym "手提電話" (mobile phone) in Tongyici Cilin, the new QFocus string "手提電話製造商" (mobile phone manufacturers) itself is a Wikipedia category title and becomes the CKtype of this question.

The reason of using Tongyici Cilin instead of Chinese WordNet is that Cilin contains larger sets of synonyms in a sufficient number.

Wikipedia synonym substitution

It is okay to apply the method introduced in the previous subsubsection with a different resource of synonyms if available. In this paper, we try to recognize synonyms in Wikipedia so that we can handle named entities in a greater extent. The detected answer type will be referred to as the ***Wikipedia-rephrased Wikipedia-category answer type*** (KKtype) throughout this paper

Wikipedia does not have features denoting synonyms. The closest one is "重定向至" (redirect) page. A redirect page states that the information of an expression e_1 is contained in another Wikipedia entry e_2, mostly because e_1 is an alternative expression of e_2. For example, both "太空船" (spaceship) and "太空飛行器" (spaceplane) are redirected to the Wikipedia entry "太空載具" (spacecraft). We treat these terms connected by the redirect relationship as one type of ***Wikipedia synonyms***. (More Wikipedia synonym types will be introduced in Section 3.1.) The following example shows how to find an answer type by substituting Wikipedia synonyms.

Q4: 一九九九年時國際足協主席是誰？
(Who was the president of FIFA in 1999?)
QFocus: 國際足協主席 (president of FIFA)
↓ 國際足協 = 國際足球聯合會 *in Wikipedia*
CKtype: 國際足球聯合會主席 (presidents of FIFA)
WKtype: 主席 (president)

In this example, its WKtype is "主席" (president). But after substituting "國際足協" (FIFA) with its Wikipedia synonym "國際足球聯合會" (FIFA), a more specific Wikipedia category title "國際足球聯合會主席" (presidents of FIFA) can be matched and becomes the KKtype of this question.

WordNet maximum matching after synonym substitution

Again as a backing, we can perform maximum matching of WordNet terms in CKtype and KKtype if available. The matched term will be referred to as the ***Cilin-rephrased WordNet-term answer type*** (CNtype) and the ***Wikipedia-rephrased WordNet-term answer type*** (KNtype) throughout this paper. Note that CNtype and KNtype may be different from WNtype, if the synonym substitution happens at the head of the question focus. The following example demonstrates how KNtype is determined.

Q5: 請問涉嫌對台軍售弊案的前法國外長為誰？
(Which former French Minister of Foreign Affairs was involved in the Taiwan's armament purchase scandal?)
QFocus: 前法國外長 (former French Minister of Foreign Affairs)
↓ 外長 = 外交部長 *in Wikipedia*
KKtype: 法國外交部長 (French Foreign Ministers)
KNtype: 部長 (Ministers)
WNtype: 外長 (Foreign Ministers)

In this example, its WNtype is "外長" (foreign minister) matched in the original question focus. But after substituting "外長" with its Wikipedia synonym "外交部長" (foreign minister) and extracting the KKtype "法國外交部長" (French foreign ministers), its head "部長" (minister) becomes its KNtype. The term "外長" is an infrequent abbreviation of "外交部長".

3 Answer Candidate Set Determination

3.1 Entries under a Specific Wikipedia Category

Among all the answer types introduced in Section 2, WKtype, CKtype, and KKtype are Wikipedia category titles. All the Wikipedia entries in these categories and their sub-categories are answer candidates. We will refer to such kind of answer candidate sets as ***Wikipedia-entry candidates*** (WKcand).

Note that an answer candidate from Wikipedia will be further extended with its Wikipedia synonyms in order to increase the probability of matching in the knowledge base of a QA system. Besides redirect relationships, we also derive synonymous terms by removing specific punctuations or phrases. All the Wikipedia synonym cases are listed in Table 1 with examples.

Synonym Case	**Origin Term**	**Synonym**
Redirect pages	"太空船" (spaceship)	"太空載具" (spacecraft)
Disambiguation pages	"豐田汽車" (Toyota Motor Corporation)	"豐田" (Toyota) *which has a disambiguation saying that "豐田汽車" is one of its possible meanings*
Disambiguation tags	"Trainspotting (film)"	"Trainspotting" *where the disambiguation tag "(film)" is removed; the tag denotes that the entry is about a film*
Comma-separated clauses	"Bothell, Washington"	"Bothell" *where the complement phrase is removed*
Interpuncts	"哈利·波特" (Harry Potter)	"哈利波特" (Harry Potter) *where "·", an interpunct inserted between first name and last name is removed*

Table 1. Cases of Wikipedia Synonyms

3.2 WordNet-Connected Wikipedia Entries

The answer types WNtype, CNtype, and KNtype are WordNet terms. We proposed two methods to bridge between Wikipedia and WordNet in order to obtain an up-to-date answer candidate set which are modern proper nouns in the following subsubsections.

There are two reasons that we need to bridge these two resources. (1) We do not use the set of hyponyms in WordNet directly, because WordNet terms are often common words rather than proper nouns. (2) The hierarchy of Wikipedia categories does not always stick to hypernym relationship. For example, one of the categories of the entry "台北市市長" (Mayor of Taipei) is "台北市政府" (Government of Taipei), which is not hypernymy but rather ontological relationship. Ponzetto and Strube (2007) have made a study on the hierarchy of Wikipedia. We would try to distinguish IS-A relationships from ontological relationships in the future.

Selecting entries under Wikipedia categories having heads of WordNet answer types

During the development of our QA system, each Wikipedia category was assigned a "WordNet head" which was the longest trailing substring of its title being a WordNet term. After a WordNet answer type is determined, its answer candidates are those Wikipedia entry titles which belong to any category having a WordNet head as a synonym or hyponym of the WordNet answer type. We will refer to such kind of answer candidate sets as ***WordNet-connected Wikipedia-category candidates*** (NCcand). For example,

Q6: 請問美國史上最大宗的企業破產事件為哪一家企業?
(What is largest company bankruptcy case in the US history?)
WNtype: 企業 (enterprise)
Answer: 安隆公司 (Enron)
→ Category: 美國已結業公司 (Defunct companies of the United States)
↪ Head: 公司 (company) *in the WordNet synset* {企業, 公司, 事業} (enterprise)

The question's WNtype is “企業” (enterprise). Its correct answer “安隆公司” (Enron) belongs to a Wikipedia category “美國已結業公司” (Defunct companies of the United States). The category's WordNet head is “公司” (company), which is a synonym of “企業” (enterprise) in WordNet. So the correct answer is successfully included in the answer candidate set by this method.

Selecting entries whose titles have heads of WordNet answer types

The second method to bridge between Wikipedia and WordNet is to match the longest WordNet term in a Wikipedia entry title itself. We call it the “WordNet head” of a Wikipedia entry. A Wikipedia entry is an answer candidate if its WordNet head is a synonym or hyponym of WNtype. We will refer to such kind of answer candidate sets as ***WordNet-connected Wikipedia-entry candidates*** (NEcand).

In the previous example, the correct answer “安隆公司” (Enron) has a WordNet head “公司” (company), which is a synonym of the WNtype “企業” (enterprise). So the correct answer is also successfully included in the answer candidate set by this method.

4 Experiments

4.1 Experiment Setup

Our main interest in this study is to detect a precise answer type and determine its answer candidate set when NER has its limitation, especially for the classes of artifacts and organizations. Unfortunately there are not many QA benchmarks providing answer type information, nor providing evaluation results according to individual answer types. Hence we chose NTCIR QA datasets even if the number of questions were not large enough.

Two benchmarks on Chinese QA have been developed in NTCIR (Sasaki et al., 2005; Sasaki et al., 2007). NTCIR-5 CLQA1 constructed 200 questions and NTCIR-6 CLQA2 tracks constructed 150 questions classified in nine coarse-grained answer types. We only focused on 4 types including PERSON, LOCATION, and especially ARTIFACT and ORGANIZATION, because they were harder to be answered correctly in the previous evaluation.

Top 1000 relevant documents for each question were retrieved by a typical tf.idf VSM IR module from the official NTCIR CLQA corpus. Answer candidates were searched inside these relevant documents and ranked by several scoring functions in our previous QA system (Lin and Liu, 2008) which included frequencies of candidates and keywords, and their distances in a document.

The usefulness of answer type determination methods is measured in terms of the size of the answer candidate set and its coverage of correct answers. The performance of a QA system is evaluated by MRR (mean reciprocal rank, the average of the inverse of the highest rank where a correct answer is proposed) and Top-1 accuracy (the percentage of questions whose top-1 answers are correct).

4.2 Performance Upper Bound

Table 2 depicts upper bound of our system. There are totally 247 questions in ARTIFACT (ART), ORGANIZATION (ORG), LOCATION (LOC), and PERSON (PRS) types. Among them, only 221 questions have explicit question foci. The other questions are expressed only by interrogative words.

Among these 221 questions, the correct answers of 196 questions are Wikipedia entry titles. But for only 177 of them, the correct answers appear in their top 1000 relevant documents, so the upper bound performance of the baseline QA system is 0.792.

# \ Atype	ART	ORG	LOC	PRS	All
Q with Focus	20	31	57	113	221
QFocus with Wiki Ans	15	29	56	96	196
QFocus with Wiki Ans in 1000doc	15	27	53	80	175

Table 2. Number of Questions in Four Answer Types

4.3 Coverage of Correct Answers in Answer Candidate Sets

Several answer candidate sets were generated by using 4 answer-type determination methods and 3 candidate-set extraction methods. Their definitions are:

- WKtype: the maximum matched Wikipedia category title in a question focus
- CKtype: the maximum matched Wikipedia category title in a Cilin-rephrased question focus
- KKtype: the maximum matched Wikipedia category in a Wikipedia-rephrased question focus
- KNtype: the maximum matched WordNet term in a Wikipedia-rephrased question focus
- WKcand: all entries under a Wikipedia category which is the answer type
- NCcand: all entries under Wikipedia categories whose heads are WordNet-connected to the answer type
- NEcand: all entries whose head is WordNet-connected to the answer type
- Union: union of all the answer candidate sets listed above
- WikiAll: using all the Wikipedia entries in different types (upper bound of the coverage)

Model		Q with Focus and Wiki Ans					Q with Focus and Wiki Ans in 1000doc				
Atype	CandSet	ART	ORG	LOC	PRS	All	ART	ORG	LOC	PRS	All
WKtype	WKcand	12	27	49	91	179	12	25	45	76	158
CKtype	WKcand	12	28	49	91	180	12	26	45	76	159
KKtype	WKcand	12	28	49	91	180	12	26	45	76	159
KNtype	NCcand	13	28	51	84	176	13	27	49	71	160
KNtype	NEcand	15	29	54	85	183	15	27	52	71	165
Union		15	29	56	91	191	15	27	52	76	170
WikiAll		15	29	56	96	196	15	27	53	80	175

Table 3. Number of Questions Having Correct Answer Candidates with Different Methods

Atype	CandSet	ART	ORG	LOC	PRS	All
KKtype	WKcand	1551.2	1893.3	4461.7	2475.5	2520.3
KNtype	NCcand	1035.8	548.6	2774.6	676.8	970.5
KNtype	NEcand	531.4	400.9	699.1	514.0	512.4
WikiAll		11822.0	5707.1	21093.5	11190.5	11222.5

Table 4. Average Number of the Distinct Answer Candidates Found in Top 1000 Documents

WikiAll is our baseline model. We collected several Wikipedia infobox templates of and mapped them into the four question types. For example, when an entry has an infobox written in the format of "infobox:組織" (infobox:organization), it is an answer candidate to an ORGANIZATION question.

The left part of Table 3 gives the coverage of different candidate sets which contain correct answers. The right part of Table 3 gives the number of questions whose correct answers appear in the top 1000 relevant documents. All the methods have very similar coverage rates. But they proposed different candidate sets, because the union sets have the greatest coverage of correct answers.

Table 4 shows the average number of distinct answer candidates found in the 1000 relevant documents. We argue that more candidates will cause more noise. Apparently WikiAll has the most candidates. Averagely every question has 11,222.5 candidates to be scored thus is quite noisy.

We can see from Table 3 and 4 that KNtype+NEcand can successfully narrow down the size of candidates to be 512.4 in average but still has the best correct-answer coverage except the union method.

Note that we did not list the results of WNtype and CNtype, because they had worse experimental results than KNtype. Although WNtype and CNtype can capture more accurate answer types, unfortunately the correct answers are neither Wikipedia entries nor instances of the detected type.

4.4 Question Answering Performance

Table 5 and Table 6 show the performance of our QA system in MRR score and top-1 accuracy, where results in Table 6 were measured on all questions and Table 5 only on questions with explicit foci. The answer candidates for questions without foci were the entire WikiAll sets.

These two tables give the same conclusions. The union of the candidate sets achieves better performance than other models. It greatly outperformed WikiAll, which provided too much candidates.

	MRR					Top-1 accuracy				
Model	**ART**	**ORG**	**LOC**	**PRS**	**All**	**ART**	**ORG**	**LOC**	**PRS**	**All**
KKtype+WKcand	0.438	0.443	0.343	0.351	0.370	0.350	0.355	0.259	0.283	0.293
KNtype+NCcand	0.450	0.519	0.428	0.336	0.396	**0.400**	0.454	0.345	**0.295**	0.340
KNtype+NEcand	0.442	**0.582**	**0.452**	0.321	0.403	0.386	**0.499**	**0.411**	0.283	**0.356**
Union	**0.492**	0.490	0.449	**0.370**	**0.418**	**0.400**	0.387	0.345	0.292	0.329
WikiAll	0.229	0.319	0.299	0.272	0.282	0.150	0.194	0.207	0.177	0.185

Table 5. Performance of Answering Questions with QFocus

	MRR					Top-1 accuracy				
Model	**ART**	**ORG**	**LOC**	**PRS**	**All**	**ART**	**ORG**	**LOC**	**PRS**	**All**
KKtype+WKcand	0.438	0.433	0.340	0.361	0.371	0.350	0.333	0.246	0.288	0.288
KNtype+NCcand	0.450	0.528	0.452	0.358	0.415	**0.400**	0.455	0.348	**0.297**	0.341
KNtype+NEcand	0.442	**0.587**	0.458	0.340	0.414	0.350	**0.486**	**0.377**	0.296	**0.349**
Union	**0.492**	0.479	**0.470**	**0.378**	**0.426**	**0.400**	0.365	0.348	**0.297**	0.329
WikiAll	0.229	0.316	0.286	0.271	0.278	0.150	0.180	0.188	0.175	0.178

Table 6. Performance of Answering All Questions

In order to compare our work with previous NTCIR QA systems, we adapted our QA system (Lin and Liu, 2007) to use the union of answer candidates by our proposed models. The choice of using a typical QA system was based on the reason that our main interest was to observe the improvement when introducing new sets of answer candidates.

Table 7 shows the performance of our system comparing to the best teams in CLQA1 and CLQA2 (Lee *et al*., 2007; Kwok *et al*., 2007) according to 4 answer types. Our system outperforms CLQA best teams on ARTIFACT and ORGANIZATION types as we have expected. Although our methods were implemented on a baseline QA system, we believe that other QA systems can also be improved by our methods.

Note that our methods did not improve QA performance on PERSON and LOCATION types. We found that the CLQA questions were created from news articles and some of them were asking information about local events. It did not violate the design of ad hoc QA task (i.e. finding answers in a given corpus), but the answers were not world-wide famous so there were no Wikipedia entries introducing them. It reveals one weakness of our methods.

	CLQA1		CLQA2		
Atype	**Our Work**	**ASQA**	**Our Work**	**ASQA**	**Pircs**
ARTIFACT	**0.385**	0.159	**0.714**	0.286	0.429
ORGANIZATION	**0.556**	0.389	0.533	**0.563**	0.313
LOCATION	0.415	**0.457**	0.438	**0.875**	0.500
PERSON	0.375	**0.563**	0.422	**0.660**	0.575

Table 7. Comparison to the Best Teams in CLQA Tasks

5 Conclusion

This paper proposes a method to bridge Wikipedia and WordNet (together with other semantic resources) in order to find a proper-sized answer candidate sets inside Wikipedia. The experimental results showed that the union of the sets of answer candidates suggested by our methods could provide a suitable-sized set of answer candidates yet still improve a baseline QA system.

In our proposed QA system, an answer type is determined by finding a trailing substring of the question focus which is also a Wikipedia category. The question focus may be rephrased by synonyms (in WordNet or Wikipedia) before the answer type determination.

The answer candidate set is determined by collecting either all Wikipedia entries in the subtree under the answer type in the hierarchy of Wikipedia categories, or all entries under the categories which have heads related to the answer type in WordNet, or all entries having heads related to the answer type in WordNet. Our final system uses the union of these kinds of candidates and achieves the best performance among different models.

Although the experimental results seem promising, it is a pity that the dataset is too small and no other suitable benchmarks are available. We wish to find a different way to setup the experiments in the future in order to verify our conclusion with stronger evidence.

Adapting our methods to another language, such as English, is a good way to have larger experiment sets. English Wikipedia uses the same strategy to build hierarchies thus we can obtain answer candidates in the same way. WordNet itself is built in English thus synonym-rephrasing is also possible during answer type determination or candidate scoring. We would like to see if the proposed methods have similar conclusions in English in the future.

Reference

Sisay Fissaha Adafre and Josef van Genabith (2008) "Dublin City University at QA@CLEF 2008," *Proceedings of CLEF 2008 - 9th Workshop of the Cross-Language Evaluation Forum*, pp. 353-360.

Ali Mohamed Nabil Allam and Mohamed Hassan Haggag (2012) "The Question Answering Systems: A Survey," *International Journal of Research and Reviews in Information Sciences (IJRRIS)*, Vol. 2, No. 3, pp. 211-221.

David Buscaldi and Paolo Rosso (2006) "Mining Knowledge from Wikipedia from the Question Answering Task," *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*, pp. 727-730.

Hoa Trang Dang, Diane Kelly, and Jimmy Lin (2007) "Overview of the TREC 2007 Question Answering Track," *Proceedings of TREC 2007*.

Ulrich Furbach, Ingo Glöckner, Hermann Helbig, and Björn Pelzer (2008) "Loganswer - a Deduction-Based Question Answering System," *Proceedings of the 4th international joint conference on Automated Reasoning, (IJCAR '08)*, pp. 139-146.

Sanda Harabagiu and Andrew Hickl (2006) "Methods for Using Textual Entailment in Open-Domain Question Answering," *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the ACL*, pp. 905-912.

Dan Klein and Christopher D. Manning (2003) "Accurate Unlexicalized Parsing," *Proceedings of the 41st Meeting of the Association for Computational Linguistics*, pp. 423-430.

Kui-Lam Kwok, Peter Deng, and Norbert Dinstl (2007) "NTCIR-6 Monolingual Chinese and English-Chinese Cross-Lingual Question Answering Experiments using PIRCS," *Proceedings of NTCIR-6 Workshop Meeting*, pp. 190-197.

Cheng-Wei Lee, Min-Yuh Day, Cheng-Lung Sung, Yi-Hsun Lee, Mike Tian-Jian Jiang, Chia-Wei Wu, Cheng-Wei Shih, Yu-Ren Chen, and Wen-Lian Hsu (2007) "Chinese-Chinese and English-Chinese Question Answering with ASQA at NTCIR-6 CLQA," *Proceedings of NTCIR-6 Workshop Meeting*, pp. 175-181.

Cheng-Wei Lee, Min-Yuh Day, Cheng-Lung Sung, Yi-Hsun Lee, Tian-Jian Jiang, Chia-Wei Wu, Cheng-Wei Shih, Yu-Ren Chen, and Wen-Lian Hsu (2008) "Boosting Chinese Question Answering with Two Lightweight Methods: ABSPs and SCO-QAT," *ACM Transactions on Asian Language Information Processing (TALIP)*, Vol. 7, Issue 4, pp. 12:1-12:29.

Chuan-Jie Lin and Ren-Rui Liu (2008) "An Analysis of Multi-Focus Questions," *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2008), Workshop on Focused Retrieval*, pp. 30-36.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean (2013) "Efficient Estimation of Word Representations in Vector Space," *Proceedings of the Workshop in the International Conference on Learning Representations* (*ICLR 2013*).

Dan Moldovan, Christine Clark, Sanda Harabagiu, and Daniel Hodges (2007) "Cogex: A Semantically and Contextually Enriched Logic Prover for Question Answering," *Journal of Applied Logic*, Vol. 5, Issue 1, pp. 49-69.

David Nadeau and Satoshi Sekine (2007) "A survey of named entity recognition and classification," *Linguisticae Investigationes*, Vol. 30, No. 1, pp. 3-26.

Anselmo Peñas, Christina Unger, and Axel-Cyrille Ngonga (2014) "Overview of CLEF Question Answering Track 2014," *Information Access Evaluation. Multilinguality, Multimodality, and Interaction, Lecture Notes in Computer Science*, Vol. 8685, pp. 300-306.

Simone Paolo Ponzetto and Michael Strube (2007) "Knowledge Derived from Wikipedia for Computing Semantic Relatedness," *Journal of Artificial Intelligence Research* (*JAIR*), Vol. 30, pp. 181-212.

Bogdan Sacaleanu, Constantin Orasan, Christian Spurk, Shiyan Ou, Oscar Ferrandez, Milen Kouylekov, and Matteo Negri (2008) "Entailment-Based Question Answering for Structured Data," *Proceedings of 22nd International Conference on Computational Linguistics (COLING 2008)*, pp. 173-176.

Tetsuya Sakai, Hideki Shima, Noriko Kando, Ruihua Song, Chuan-Jie Lin, Teruko Mitamura, Miho Sugimito, and Cheng-Wei Lee (2010) "Overview of NTCIR-8 ACLIA IR4QA," *Proceedings of the 8th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering, and Cross-Lingual Information Access (NTCIR-8)*, pp. 63-93.

Juliano Efson Sales, André Freitas, Brian Davis, and Siegfried Handschuh (2016) "A Compositional-Distributional Semantic Model for Searching Complex Entity Categories," *Proceedings of the 5th Joint Conference on Lexical and Computational Semantics* (**SEM*), pp. 199-208.

Yutaka Sasaki, Hsin-Hsi Chen, Kuang-hua Chen, and Chuan-Jie Lin (2007) "Overview of the NTCIR-5 Cross-Lingual Question Answering Task (CLQA1)," *Proceedings of NTCIR-5 Workshop Meeting*, pp. 175-185.

Yutaka Sasaki, Chuan-Jie Lin, Kuang-hua Chen, and Hsin-Hsi Chen (2007) "Overview of the NTCIR-6 Cross-Lingual Question Answering (CLQA) Task," *Proceedings of NTCIR-6*, pp. 153-163.

Ulli Waltinger, Alexa Breuing, and Ipke Wachsmuth (2011) "Interfacing Virtual Agents with Collaborative Knowledge: Open Domain Question Answering Using Wikipedia-Based Topic Models," *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI-11)*, pp. 1896-1902.