Liu, C.-C. et al. (Eds.) (2014). Proceedings of the 22nd International Conference on Computers in Education. Indonesia: Asia-Pacific Society for Computers in Education

Question Type Analysis for Question-Answering Applications in Education

Yang-Yin Lee[a], Yong-Siang Shih[a], Chih-Chieh Shao[b], Yen-Pin Chiu[b], Hsin-Hsi Chen[a,b*], Chieh-Jen Wang[c] & Sen-Chia Chang[c]

[a]*Department of Computer Science and Information Engineering,*
[b]*Graduate Institute of Networking and Multimedia,*
National Taiwan University, Taipei, Taiwan
[c]*Intelligent Analytic Technology Division, Computational Intelligence Technology Center, Industrial Technology Research Institute, Hsinchu, Taiwan*
*hhchen@ntu.edu.tw

Abstract: In this paper, we present a question-answering (QA) system as a virtual tutor for students in the 5th and 6th grades. Students ask questions and the QA system gives answers to their questions based on a knowledge base. Teaching materials for history and geography are considered as a knowledge source. Because question log is not available in developing QA systems, multiple choice questions (MCQs) in the learning and testing materials are regarded as a training corpus to learn question types, answer types and keywords for retrieval, where an MCQ consists of a stem and a set of options. Options from the same MCQ are grouped into a cluster. Clusters with common elements are merged into a larger cluster. A cluster is labelled with a nominal element selected from the corresponding stems. We also mine question patterns from the stems for question type analysis in the QA system. Because the questions created by instructors in MCQs and the questions asked by students may be different, we develop a procedure to collect possible questions from students in the 6th grade. In the experiments, we first evaluate the question type classification systems using the MCQ corpus and the student corpus with 5-fold cross validation, respectively. Then we train question type classifiers with the complete MCQ corpus, and test them on the student corpus. The student's and the instructor's questions are compared and analyzed.

Keywords: Computer-assisted learning, question-answering systems, student intent analysis, text mining

1. Introduction

Question-answering (QA) systems, which extract answers from a background knowledge base to answer users' questions, attracts much attention in this decade. Famous information retrieval evaluation forums such as TREC, CLEF and NTCIR hold various question-answering evaluation tasks (Voorhees, 2002; Giampiccolo et al., 2007; Sasaki et al., 2005). Web platforms such as Yahoo Answers encourage knowledge partners to post questions and give answers collaboratively. The resulting huge QA database promotes the development of community QA systems (Zhou et al., 2013). Recently, DeepQA in Jeopardy! Challenge proves the feasibility of QA applications in real world (Ferrucci, 2012). Voice QAs such as Apple Siri and Google Now have been introduced into mobile devices.

Although commercial QA products are available currently, there are still many places to improve the performance of open domain question answering systems. In addition, specific domain question answering has its own demands for many real world applications. QA system can be introduced to the education domain for collaborative learning, online education and distance learning (Wang et al., 2006; Wen et al., 2012). In this paper, we focus on developing QA applications in education as virtual tutors. Students can post questions anytime anywhere after class, and computer-assisted learning systems response answers to their questions.

A typical question answering system is composed of question type analysis, query formulation, passage retrieval, and answer extraction (Ravichandran and Hovy, 2002). Question type analysis, which aims to understand users' intents, provides important clues for the latter tasks. The clues include the question part referencing to the answer, terms indicating the type of entities being asked for, and a classification of the question into some broad types (Lally et al., 2012). Query log which keeps users' information needs provide some prior knowledge for question type analysis, but it is not always available for all the application domains.

This paper utilizes multiple choice questions (MCQs) in learning and testing materials to mine question patterns to support QA systems in education domain. It is organized as follows. Section 2 describes a QA system for computer-assisted learning. Section 3 proposes methods to mine the important clues from MCQs. Section 4 presents a methodology to collect students' questions. The differences between students' questions in the student corpus and instructor's stems in the MCQ corpus are addressed. Section 5 evaluates the question type classifiers with these two corpora. Section 6 concludes the remarks.

2. A Question-Answering System for Computer-Assisted Learning

Figure 1 shows a question-answer system for computer-assisted learning, including the online and the offline parts. In the offline part, we transform history and geography textbooks in the 5th and 6th grades into a knowledge base. The online part accepts voice question inputs from students, and responses answers based on the knowledge base. It consists of the following three modules: question processing, information retrieval, and answer generation.

In question processing, basic language understanding such as word segmentation, part of speech tagging and dependency parsing is performed. The linguistic analysis supports information for question type classification, answer type identification, and keyword selection. If the information is not enough for retrieving relevant passage, then dialog manager initiates an interaction with students. In information retrieval, we formulate a query in terms of keywords selected from questions. To deal with the paraphrase problem, query expansion/reformulation may be done. The retrieval model selects passages from the knowledge base, and ranks their order by relevancy degree. In answer generation, keyword and relation matching confirms the relevancy of the retrieved passages, and extracts the answer based on the answer type. Finally, the answer is reported to students.

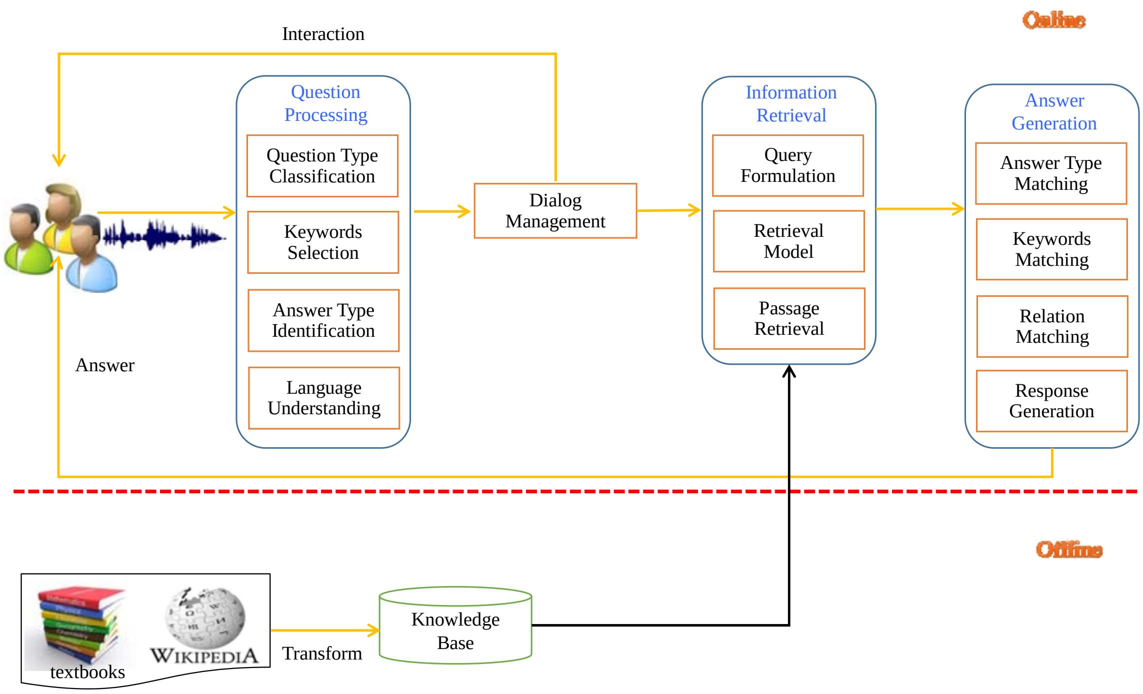

Figure 1. A Question-Answering System for Computer-Assisted Learning.

3. Question Pattern Mining

3.1 Multiple Choice Questions (MCQ) Corpus

Question type, answer type, and the keywords to retrieve potential passages containing answers are three major elements in a question. Query log recording users' questions is useful for mining the important information. Unfortunately, it is often not available when QA systems are developed. Here we use an MCQ corpus instead. The MCQ corpus is built based on the learning and testing materials. An MCQ consist of a stem and 4 options. The stem describes the questions to be asked and the options list the possible answers. (E1)-(E3) shows three Chinese MCQs and their English translation.

(E1) 臺灣與韓國之間隔著哪一個海域？ (1) 巴士海峽 (2) 東海 (3) 南海 (4) 臺灣海峽。
What sea area separates Taiwan and South Korea? (1) Bashi Channel (2) East China Sea (3) South China Sea (4) Taiwan Strait

(E2) 何者是臺灣與菲律賓群島之間相隔的海域？ (1) 臺灣海峽 (2) 太平洋 (3) 南海 (4) 巴士海峽。
What sea area separates Taiwan and the Philippines? (1) Taiwan Strait (2) Pacific Ocean (3) South China Sea (4) Bashi Channel

(E3) 琉球群島位在臺灣的哪一個方位？ (1) 西北方 (2) 西南方 (3) 東北方 (4) 東南方。
What position is Ryukyu Islands located from Taiwan? (1) North West (2) South West (3) North East (4) South East

Table 1 summarizes the three major elements in examples (E1)-(E3). The triggers for the question types of these three examples, i.e., what-questions, are 哪一個 (what) and 何者 (what). The triggers for the answer types are 海域 (sea area) and 方位 (position). The keywords for retrieval contain predicate such as 隔著 (separate), 相隔 (separate) and 位在 (is located), and the related arguments. These predicates tend to collocate with some triggers for the answer types. They are useful when the words to determine the answer type is not specified explicitly. (E4) is an example. Here a general question word 什麼 (what) is used without 海域 (sea area). The word 相隔 (separate) gives some clue for the answer type.

(E4) 臺灣與韓國之間隔著什麼？ (What separates Taiwan and South Korea?)

Table 1: Major elements for examples (E1)-(E3).

	Triggers for Question Type	Triggers for Answer Type	Keywords for Retrieval
1	哪一個 (what)	海域 (sea area)	隔著 (separate), 臺灣 (Taiwan), 韓國 (South Korea)
2	何者 (what)	海域 (sea area)	相隔 (separate), 臺灣 (Taiwan), 菲律賓 (the Philippines)
3	哪一個 (what)	方位 (position)	位在 (is located), 琉球群島 (Ryukyu Islands), 臺灣 (Taiwan)

Because MCQs in the learning and test materials are designed for assessment, not all MCQs are suitable for training. Those MCQs whose items contain "Which is correct" or "Which is not correct" (refer to E5) and whose options contain "All of the above", "None of the above", or combination answers (refer to E6), will not be selected.

(E5) 有關臺灣位置的敘述，下列何者正確？
Which specification about the position of Taiwan is correct?

(E6) 臺灣的海洋產業包含下列哪些？A. 水產養殖, B. 海洋漁業, C. 海洋觀光, D. 冰河觀光。(1) ABCD (2) ABC (3) BCD (4) BD。
What does Taiwan's marine industry include in the following list? A. aquaculture, B. marine fishery, C. sea sightseeing, D. glacier sightseeing. (1) ABCD (2) ABC (3) BCD (4) BD

3.2 A Pattern Mining Algorithm

A typical information-gathering question concerns people, locations, time, things, and matter. It is often triggered by the 5w1h keywords such as who, what, where, when, why, and how. We can classify questions into these six broad types. The finer answer types that users are interested in are often specified explicitly or implicitly in the questions with some other nominal words. Table 2 shows some 5w1h keywords, the nominal words, the related answer types, and the examples. The nominal words specify finer answer types. Thus, the answer type identification will tell out the finer type of each question. The related issues are how many types will be formulated and which type this questions belong to.

Table 2: Some 5w1h keywords and the answer types.

5w1h	nominal	Answer Type	Examples
誰 (who)		person	端午節是為了紀念誰？(Who is commemorated in Dragon Boat Festival?)
什麼 (what)	朝代 (dynasty)	dynasty→time	開放安平跟基隆是什麼朝代？(In what dynasty did Anping and Keelung open?)
什麼 (what)	島 (island)	island→location	形狀看起來像蕃薯的是什麼島？(What island whose shape looks like a sweet potato?)
什麼 (what)		event/thing	阿美族的過年稱為什麼？(What Amis calls the New Year?)
哪 (which)	老師 (teacher)	teacher→person	哪一個老師的誕辰訂為教師節？(Which teacher's birthday is set as Teachers' Day?)
哪 (which)	時代 (era)	era→time	哪個時代是鄭成功創辦的？(Which era was founded by Zheng Chenggong?)
哪 (which)	山 (mountain)	mountain→location	台灣最高的山是哪一座？(Which is the highest mountain in Taiwan?)

We aim to mine entities, entity type, question type, and the relation among entities from MCQs in the teaching materials. We postulate options in MCQs denote the same type except how and why questions. A mining algorithm is shown below.

(1) Collect initial clusters from the options in the same MCQ.
(2) Merge clusters based on the transitivity. In other words, clusters with common elements are merged into a larger cluster.
(3) Partition stems in MCQs based on the clusters derived from (2).
(4) Assign a label to each cluster with nominal keywords in the stems.
(5) Find the question patterns for each cluster, where a pattern is in terms of 5h1w+nominal keyword.
(6) Extract the relations among entities.

The following takes (E1)-(E3) as examples to describe the above algorithm.

(1) Three initial clusters, i.e., {Bashi Channel, East China Sea, South China Sea, Taiwan Strait}, {Taiwan Strait, Pacific Ocean, South China Sea, Bashi Channel}, and {North West, South West, North East, South East}, are collected.
(2) The first two clusters have three common elements, thus a new cluster is formed, i.e., {Bashi Channel, East China Sea, South China Sea, Taiwan Strait, Pacific Ocean}.
(3) MCQs is partitioned into {E1, E2} and {E3}.
(4) {Bashi Channel, East China Sea, South China Sea, Taiwan Strait, Pacific Ocean} and {North West, South West, North East, South East}are tagged with labels “sea area” and “position”.
(5) Two question patterns, i.e., “what+sea area” and “what position”, are found.
(6) Two relation keywords, i.e., “separate” and “is located”, are extracted.

4. Collect Students’ Questions

4.1 Procedure to Collect Questions

The users of the proposed question-answering system are students in the 5th and 6th grades. We all know there is an information gap between an information need and a real query even for adult searchers. In this study, we would like to know how children formulate their questions, and the effects of fuzzy questions and incomplete questions on students' question type analysis and question-answering.

Total 28 students in the 6th grade of an elementary school in New Taipei City, Taiwan, took part in the study. We divided these students into 7 group. After a short introduction, each group was asked to provide questions for some predefined topics. Three different methods are proposed to guide students to formulate questions. At most 5 questions are provided for each method. The following shows a sample topic about a history of Taiwan and the three methods to collect questions.

Sample topic: 鄭成功率軍渡海來臺，驅逐荷蘭人，使臺灣成為反清復明的基地。 Zheng Chenggong led his troops across the sea to Taiwan, expelled the Dutch, and made Taiwan to be a base for rebelling Qing dynasty and rebuilding Ming dynasty.

Method 1: Please change the above declarative sentence into an interrogative sentence whose answer is 鄭成功 (Zheng Chenggong).

Method 2: Please list questions whose answer is 鄭成功 (Zheng Chenggong).

Method 3: Please list questions related to 鄭成功 (Zheng Chenggong) and their answers.

Method 1, which rewrites the given sentence, is the simplest way. Method 2, which keeps the same answer, but presents different views, is more complex. Method 3, whose answer is different from the original topic, is the most complex. (E7)-(E9) show examples for Methods 1-3, respectively.

(E7) 誰驅逐了佔領臺灣的荷蘭人？ (Who expelled the Dutch who occupied Taiwan?)

(E8) 被封為延平郡王的是哪一位？ (Who was called Koxinga?)

(E9) 鄭成功是在哪裡登陸臺灣的？ (Where did Zheng Chenggong land Taiwan?)

4.2 Analysis of Students' Questions

Because students have different backgrounds and linguistic capabilities, the question they formulate may be incomplete, vague, ambiguous, or erroneous. Total 683 questions were collected. Of them, 155 questions contain at least such a phenomenon. Table 3 lists several types of phenomena found in the students' questions. These types are defined as follows. The top three types are ambiguous sentences, incorrect uses of words, and uses of wrong characters. They occupy 56.78% in total. Near 83% of ambiguous questions are the-most questions, e.g., 最有名 (the most famous), 最聰明 (the smartest), and so on.

(1) Wrong characters: Students do not write down the correct characters in their questions.

(2) Missing characters: Some characters are missing, e.g., a character "至" is omitted from the question "誰被稱為聖先師？" by students.

(3) Additional characters: Some extra characters are inserted in the questions.

(4) Incorrect uses of words: Unsuitable words are used, e.g., "臺灣是幾月幾日誕生？" (When is the birthday of Taiwan?) and "水稻住在什麼地？" (Where does rice live?)

(5) Mixing English words: English words are placed in Chinese questions, e.g., "台灣中部最高的山在where?" (Where is the highest mountain in the middle Taiwan?)

(6) Incomplete sentences: Sentences are incomplete, e.g., "什麼山被稱呼？" (What mountain is called?"

(7) Ambiguous/subjective sentences: Sentences have more than one interpretation, e.g., "去台中一定要帶回來的食物是什麼？" (What food must be brought back from Taichung?)

(8) Uninterpretable sentences: Sentences cannot be understandable, e.g., "貼在門上的東西是什麼顏色？" (What colors of the things on the door?)

(9) Illogical sentences: Sentences are not logical, e.g., "愛國詩人是在哪個節日自殺的？" (At which festival did the Patriotic Poet suicide?)

(10) Without answers: There may be answers, but they cannot be found in our knowledge base, e.g., "臺灣最早的口足畫家是誰？" (Who is the first painter using foot and mouth?)

(11) Complex questions: Questions cannot be answered even by human, e.g., "如果無孔子那怎辦？" (How do we do without Confucius?)

(12) Others: Some questions contain more than one type of phenomena.

Table 3: Distribution of error types in the students' questions.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
19	4	4	33	9	10	36	14	4	8	3	11
12.26%	2.58%	2.58%	21.29%	5.81%	6.45%	23.23%	9.03%	2.58%	5.16%	1.94%	7.10%

5. Results and Discussions

5.1 Corpora Comparison

In the experiments, two corpora are available for evaluating the question type analysis. Questions in the MCQ corpus are written by instructors, and questions in the student corpus are provided by the students at the 6th grade. There are 11 common question types shown as follows in these corpora. We label a question type for each question.

(T1) people, e.g., 端午節時，我們會包粽子、划龍舟，這些活動是為了紀念誰？ (At the Dragon Boat Festival, we package dumplings and hold dragon boat racing. Who do these activities commemorate?)

(T2) things, e.g., 哪一個事件發生後，使得臺灣正式成為漢人統治的地區？ (Which event made Taiwan governed by the Hans?)

(T3) time, e.g., 第一枚原子彈被運用在戰爭上，始於哪一年？ (When was the first atomic bomb used in the war?)

(T4) location, e.g., 何處能看見火山地形？ (Where can I see the volcanic terrain?)

(T5) matter, e.g., 清朝時期，想要在臺灣開墾土地的人，必須先向政府申請什麼？ (In Qing Dynasty, what must be applied to the government first when people wanted to reclaim land in Taiwan?)

(T6) numerical value, e.g., 臺灣島的面積大約有多少平方公里？ (How many square kilometers is the area of Taiwan Island?)

(T7) application, e.g., 在地圖上會標示許多圖例，請問圖例的功用是什麼？ (Many legends are marked on the map. What is the function of the legend?)

(T8) phenomena, e.g., 哪一種天氣現象可以豐富家鄉的水資源？ (What kind of weather phenomena can enrich water resource of hometown?)

(T9) reason, e.g., 中亞地區的鹹海面積逐漸縮小，其主要原因是什麼？ (Why is area of the Aral Sea in Central Asia gradually reduced?)

(T10) title, e.g., 鄭成功二十一歲時，明代唐王對他賞識有加，賜姓「朱」，後人因此而稱他為什麼？ (When Zheng Chenggong was twenty-one years old, the Ming Tang bestowed the surname "Zhu" to him. Thus, what people call him later?)

(T11) language, e.g., 荷蘭人引進哪一種拼音方式，書寫原住民的語言？ (What spelling system was introduced by the Dutch for Aboriginal language writing?)

Table 4 lists the distribution of the 11 types in the MCQ corpus (M) and the student corpus (S). These two corpora contain 593 and 395 instances, respectively. The top 5 types in the MCQ corpus are location (45.36%), matter (12.48%), phenomena (11.80%), reason (9.28%) and people (7.92%). They occupy 86.84% of instances. Comparatively, the top 5 types in the student corpus are location (23.04%), time (21.77%), matter (17.97%), people (16.46%) and things (8.61%). They cover 87.85% of instances. Location is the most interesting type in the two corpora. Matter and people types are also interesting for instructors and students. Surprisingly, only 0.84% of the instances are related to things type in the MCQ corpus. In contrast, 8.61% of students' questions related to this type. Similarly, 11.80% of instructors' questions touch on phenomena, but only 1.01% of students' questions belong to this type. It may be because this type of questions is difficult to be formulated by students. In both corpora, numerical value, title and language are the three minority types.

Table 4: Distribution of question types in the two experimental corpora.

	(T1)	(T2)	(T3)	(T4)	(T5)	(T6)	(T7)	(T8)	(T9)	(T10)	(T11)
M	47	5	30	269	74	9	31	70	55	1	2
%	7.92	0.84	5.06	45.36	12.48	1.52	5.23	11.80	9.28	0.17	0.34
S	65	34	86	91	71	3	10	4	23	1	7
%	16.46	8.61	21.77	23.04	17.97	0.76	2.53	1.01	5.82	0.25	1.77

5.2 Experimental Setup

In the first set of experiments, we train and test the question type classifier by 5-fold cross validation with the MCQ corpus and the student corpus, respectively. In the second set of experiments, we train the question type classifier with the complete MCQ corpus, and test it with the student corpus. Those incomplete, vague, ambiguous, or erroneous questions are removed from the student corpus for testing.

We aim to classify each question into one of the 11 types. L2-regularized L2-loss support vector classification in LIBSVM (Chang and Lin, 2011) is adopted. Two sets of features are explored. The first contains triggers for question types, triggers for answer types and the relation keywords. They are extracted from the corpora. The second contains bigrams and trigrams features. All the features are binary. Table 5 lists the accuracies of the proposed classifiers. Using triggers and relation keywords is better than using 2-grams and 3-grams in all the evaluation experiments. The classification performance in 5-fold cross validation on the student corpus with the triggers and relation keyword approach is better than that in 5-fold cross validation on the MCQ corpus. There are performance drops from cross validation on the MCQ corpus to testing on the student corpus. The drops by the approach using the latter approach is much larger than those by the former approach. It shows the wordings used and topics mentioned in these two corpora are somewhat different. Referencing to the external linguistic recourse such as Chinese Proposition Bank (Xue et al., 2013) will be explored. We can introduce semantic roles into Chinese predicates (Xue, 2008).

Table 6 further shows precision, recall and F1 score of the approach using triggers and relation keywords on the student corpus. Types of the top five F1 scores are reason (T9), people (T1), things (T2), numerical value (T6) and location (T4). Their F1 scores are larger than 0.66. In contrast, the numbers of instances for phenomena (T8), title (T10), and language (T11) types are very few. All of them are not correctly classified in the experiments.

Table 5: Accuracies of the proposed question type classifiers.

	Triggers and relation keyword	2-grams and 3-grams
5-fold cross validation on the MCQ corpus	0.7707	0.7673
5-fold cross validation on the student corpus	0.8608	0.7468
Test on the student corpus	0.5848	0.4734

Table 6: Precision, recall, and F1 score using trigger words and relation keywords.

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
Precision	0.8448	1	0.9737	0.5374	0.8636	0.6667	0.4286	0	0.9474	0	0
Recall	0.7538	0.6176	0.4302	0.8681	0.2676	0.6667	0.2000	0	0.7826	0	0
F1 Score	0.7967	0.7636	0.5968	0.6639	0.4086	0.6667	0.2727	0	0.8571	0	0

5.3 Discussion

The analyses in Section 5.1 demonstrate some type distributions of the MCQ corpus and the student corpus are different. We further analyze which question types are confused with one another. Table 7 show a confusion matrix of the approach trained with the MCQ corpus and tested on the student corpus. Here the triggers and relation keywords are adopted. Numbers in bold denote how many instances in type A is correctly recognized as A. Numbers underlined show the numbers of bad recognitions. People (T1), time (T3), and matter (T5) types, which are three major types in the student corpus, are confused with the location type (T4), which is the largest type in the MCQ corpus. Time (T3) and matter (T5) types are also easy to be recognized as phenomena type (T8). Moreover, six instances of the language type (T11) are misclassified into the application type (T7). (E10) is an example. The relation keyword 寫 (write) does not appear in the training corpus, thus the trigger word 怎麼 (how) determines the question type. Considering the word dependencies, such as the noun modifier of 英文 (English) is 蒸汽機 (steam engine), and the object of 寫 (write) is 英文 (English), and linguistic resource will be helpful to deal with this problem.

(E10) 蒸汽機的英文怎麼寫? (How do we write steam engine in English?)

Table 7: Confusion matrix of prediction using trigger words and relation keywords.

		Prediction										
		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
Ground Truth	T1	**49**	0	0	13	1	0	0	2	0	0	0
	T2	0	**21**	0	5	1	0	0	7	0	0	0
	T3	4	0	**37**	22	0	1	0	20	1	1	0
	T4	4	0	0	**79**	0	0	0	8	0	0	0
	T5	0	0	1	17	**19**	0	2	32	0	0	0
	T6	0	0	0	1	0	**2**	0	0	0	0	0
	T7	0	0	0	2	0	0	**6**	2	0	0	0
	T8	0	0	0	3	1	0	0	**0**	0	0	0
	T9	0	0	0	5	0	0	0	0	**18**	0	0
	T10	1	0	0	0	0	0	0	0	0	**0**	0
	T11	0	0	0	0	0	0	6	1	0	0	**0**

6. Conclusion

This paper proposes a question answering system as a virtual tutor for students in the 5th and 6th grades. We train a question type classification system with the MCQ corpus. To understand how children formulate their questions, we design a procedure to collect students' questions and analyze the qualities of the questions. We found that 155 of 683 students' questions, i.e., 22.69%, contain some incomplete, vague, ambiguous or erroneous information to QA systems. The experimental results show that the classification system using triggers and relation keywords achieves accuracies 0.7707 and 0.8608 when 5-fold cross validation on the MCQ corpus and the student corpus are adopted, respectively. The system trained with the MCQ corpus and tested on the student corpus has accuracy 0.5848.

The question type distributions in the MCQ corpus and the student corpus affect the performance. The methodology to guide students to prepare their questions determines the question types. Thus, people, location, time, things, and matter types dominate the top 5 questions in the student corpus. Comparatively, reason and phenomena, which instructors usually ask in the assessment, belong to the top 5 question types in the MCQ corpus. The wordings and the topics difference between these two corpora are the major problems of the performance drop.

Introducing the dependency relations of parsing and the outside linguistic resource will be explored in the future task. Besides, students' questions containing errors are removed in this study. How to deal with those types of errors in QA systems will be investigated. Wrong characters in the students' questions can be dealt by n-gram language models. Translation mechanism can be considered to resolve questions in both Chinese and English.

Acknowledgements

This research was partially supported by the Project D352B23100 (ITRI) under the sponsorship of the Ministry of Economic Affairs, Taiwan, R.O.C.

References

Chang, C.-C. and Lin, C.-J. (2011). LIBSVM: A Library for Support Vector Machines. ACM Transactions on Intelligent Systems and Technology, 2(3), 27:1–27:27.

Ferrucci, D.A. (2012). Introduction to "This is Watson". *IBM Journal of Research and Development*, 56(3), 1.

Giampiccolo, D., Froner, P., Pen, A., Ayache, C., Cristea, D., Jijkoun, V., Osenova, P., Rocha, P., Sacaleanu, B., & Suteliffe, R. (2007). Overview of the CLEF 2007 multilingual question answering track. In , C. Peters, V. Jijkoun, T. Mandl, M. Henning, D. W. Oard, P. Anselmo, V. Petras, and D. Santos (Eds), *Proceedings of Cross Language Evaluation Forum, Advances in Multilingual and Multimodal Information Retrieval* (pp. 200-236), vol. 5152, Lecture Notes In Computer Science, Berlin, Germany: Sringer-Verlag.

Lally, A., Prager, J.M., McCord, M.C., Boguraev, B.K., Patwardhan, S., Fan, J., Fodor, P., & Chu-Carroll, J. (2012). Question analysis: How Watson reads a clue. *IBM Journal of Research and Development*, 56(3/4), 2:1-2:14.

Ravichandran, D. & Hovy, E. (2002). Learning surface text patterns for a question answering system. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics* (pp. 41-47).

Sasaki, Y., Chen, H.-H., Chen, K.H., & Lin, C.J. (2005). Overview of the NTCIR-5 cross-lingual question answering task. In *Proceedings of the 5th NTCIR Workshop Meeting* (pp. 175-185).

Voorhees, E. (2002). Overview of the TREC 2002 question answering track. In *Proceedings of the 11th Text REtrieval Conference* (pp. 115–123).

Wang, C.C., Hung, J.C., Yang, C.Y., & Shih, T.K. (2006). An application of question answering system for collaborative learning. In *Proceedings of the 26th IEEE International Conference on Distributed Computing Systems Workshops*.

Wen, D., Cuzzola, J., Brown, L. & Kinshuk (2012). Instructor-aided asynchronous question answering system for online education and distance learning. *The International Review of Research in Open and Distance Learning*. http://www.irrodl.org/index.php/irrodl/article/view/1269/2369.

Xue, N. 2008. Labeling Chinese predicates with semantic roles. Computational Linguistics, 34(2):225-255.

Xue, N., Bai, X., Lu, J., Zhang, J., Palmer, M., Chang, M., & Zhong, H. (2013). Chinese Proposition Bank 3.0. http://catalog.ldc.upenn.edu/LDC2013T13.

Zhou, G., Liu, Y., Liu, F., Zeng, D., & Zhao, J. (2013). Improving question retrieval in community question answering using world knowledge. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence* (pp. 2239-2245).