

APQA System for CCLQA

Song Liu

The University of Tokushima
Department of Information Science &
Intelligent Systems
Faculty of Engineering

Beijing University of Post and
Telecommunications

Liusong @is.tokushima-
u.ac.jp

Xiaobo Feng

The University of Tokushima
Department of Information Science &
Intelligent Systems
Faculty of Engineering

Beijing University of Post and
Telecommunications

Fengxiaobo @is.tokushima-
u.ac.jp

Fuji Ren

The University of Tokushima
Department of Information Science &
Intelligent Systems
Faculty of Engineering

Ren @is.tokushima-u.ac.jp

ABSTRACT

The Question Answering System can be divided into four components as Question Analysis, Document Retrieval, Answer Extraction and Answer Generation. And because of the requirement of the Complex Cross-Lingual Question Answering(CCLQA) task the first three parts are necessary. In question Analysis we use the templates and rules to classify the questions and acquire the keywords. In Document Retrieval part Lucene was used as the search engine. And in Answer Extraction, the retrieved documents are divided into sentences. We use the keywords matching to find the candidate sentences. For different types of questions particular strategy are attempt to extract the answers. Some questions are returned good answers, while some questions' results are not satisfied. We try to analyze the results to find the success and the mistake reasons.

1. INTRODUCTION

The Question Answering System can be divided into four components as Question Analyzsis, Document Retrieval, Answer Extraction and Answer Generation. And because the CCLQA task requires that the answers returned by the system must be the phrases or sentences that can be find in the corpus, the Answer Generation part is not necessary. Our question answering system APQA only contains Question Analysis, Document Retrieval and Answer Extraction. We use the rule and statistical methods. The multi-level Retrieval and Information Extraction are introduced.

There are totally 100 questions that belong to nine types: DEFINITION, BIOGRAPHY, RELATIONSHIP, EVENT, WHY, PERSON, ORGANIZATION, LOCATION, and DATE [8]. In the process of Question Analysis the types of the questions should be confirmed. For different types of questions the follow-up processes are different. Then the key words of the questions are extracted. The next part is Document Retrieval. In this part the search queries are created depending on the types of questions and the key words. Based on the query The documents are returned by the search engine. In the Answer Extraction process, the returned documents are divided into sentences, form which the keywords matching is used to select the candidate sentences. There are particular strategies to confirm the answers for each type of questions. At last the answers are scored.

In the remaining document, section 2 describes the APQA system. Section 3 discusses the evaluation results of our system. Section 4 is the future works. And conclusion is in section 5.

2. System

There are three components in the APQA system, which are Answer Analysis, Document Retrieval and Answer Extraction. In Answer Analysis, The question type is confirmed by the rules. And the keywords are extracted by using different templates from different types of questions. In Document Retrieval, Lucene is the base search engine. And there are different types of search queries for different types of question. In Answer Extraction, because each type of question has its particular interest information, there are also different extraction strategies for them.

2.1 Question Analysis

There are two main task of question analysis. The first is question classification. For different types of questions the following processing is significant different. The second is getting the keywords for all types of questions. The type and keywords guide the following Document Retrieval and Answer Extraction processes.

2.1.1 Question classification

For the questions classification the rule based classifying is used. There are 9 types of questions. Five types are complex questions including DEFINITION, BIOGRAPHY, RELATIONSHIP, EVENT, and WHY; and four types are factoid questions including PERSON, ORGANIZATION, LOCATION, and DATE [8]. All of the nine type of questions have the particular features to distinguish their types, for example the Why questions are beginning with “Why” most of time. We draft the rules to determine the types based on these features. Table 1 describes the rules in details.

Table 1. The classification rules

Question type	Rules
DEFINITION	1 The question contains 什么是 / 是什么 / 何为 / 何谓 / 定义. 2 The question has one objective.
BIOGRAPHY	1 The question contains “谁”. 2 the main clause of question contains person's name.
RELATIONSHIP	The question contains “关系”
EVENT	The question contains “举出 / 列举 / 事件 / 影响”
WHY	The question contains “为什么 / 为何 / 何

	以”
PERSON,	1 The question contains “谁”. 2 the main clause of question does not contains person’s name.
ORGANIZATION,	The hypernoms of the word after the interrogative includes “组织”.
LOCATION	The question contains the word “哪里 / 位于”.
DATE	The question contains the word “时候 / 时间 / 日期 / 年 / 月 / 日”.

And if the question does not match all the conditions the question is classified into Event.

2.2.2 The key words

Almost all the questions just have one simple sentence. It is suitable to extract the keywords by template. For the nine types of questions we construct several templates to get the key words from them.

For example the templates of DEFINITION questions:

“什么是 <Object>”

“<Object> 是什么”

“<Object> 的定义是什么”.

Another example some templates of WHY question:

“为什么 <Object1> 要 / 会 / 能 <Action> <Object>”.

“为什么 <Object1> 要 / 会 / 能 <Action>”.

A template for RELATIONSHIP question :

"<Objiet1> 和 <Object2> 是什么关系".

Templates for EVENT:

"列举 <Event> 的 <Attribute>".

"列举 <Event> 的 <Impact>".

Templates for BIOGRAPHY question:

"<Name> 是谁".

"谁是 <Name>".

For some questions such as some Event questions, some Date Questions and some Organization questions all the substantive are also important part of the keys.

2.2. Retrieval

In the Retrieval part the search engine is Lucene. After Question Analysis we get the question types and the keywords. We construct the queries for different types of questions.

2.2.1 Keywords expansion

It is impossible that all the key words we get from the questions appear in the documents in the same form. In order to cover all the situations the key words are expanded .

For Name Entities we expand them by the encyclopedia to get the entire aliases. And we construct a geography shortened form list to find the geography nouns. For the verbs and common nouns we get the synonyms from the hownet [12].

2.2.3 Query Generation

In this part we will describe the queries for each type of question. The DEFINITION and BIOGRAPHY questions have the same type of query that only contains the keyword and the alias expanded from the encyclopedia. For DEFINITION question the keyword is the object. And for BIOGRAPHY question the keyword is the name. For example “谁是华特·迪斯尼？”，the name “华特·迪斯尼” and the alias “沃尔特·埃利亚斯·迪士尼” compose the query, "华特·迪斯尼 || 沃尔特·埃利亚斯·迪士尼".

The RELATIONSHIP question contains two objects. The query is made by the objects and their aliases. The question, "巴厘岛爆炸和本·拉丹的关系" the two objects are "巴厘岛爆炸" and "本·拉丹". The object "本·拉丹" has the aliases "乌萨马", "拉丹" and "本·拉登" etc. The query is "巴厘岛爆炸 && (本·拉丹 || 乌萨马 || 拉丹 || 本·拉登)".

For the EVENT question and WHY question the nouns, verbs and their synonyms also the name entities and their aliases are included by the query.

The rest questions are factoid questions: DATE, LOCATION, PERSON and ORGANIZATION. First the words that show the type information of the question are removed. For the DATE question the words are “时候”, “时间” and “日期” etc. For the LOCATION question the words are “哪里”, “位于” and “在”. For the ORGA-ZNATIION question the hypernoms of the words after the interrogative includes “组织”. Then the nouns, verbs and name entities are expanded to construct the queries.

2.2.3 Document ranking

Although the returned documents are ranked by the search engine, the re-rank is necessary for question answering. We use the document-question nouns relevance, paragraph-question nouns relevance and sentence-question similarity to rank the retrieval documents.

The document-question relevance and paragraph-question relevance are measured by the proportion of keywords and their expansions in the documents and paragraphs. (Equation 2.1)

$$Re = \frac{Rnum}{Allnum} \quad (2.1)$$

Re is the relevance between the question and a document or a paragraph. Rnum is the number that the keywords and their expansions appear in the document or paragraph. Allnum is the number of the document or paragraph words. The bigger Re the more relevance is.

The similarity between sentences and questions is measured by the edit distance[13]. (Equation 2.2)

$$Sim = \frac{Ed}{Qlen} \quad (2.2)$$

The Sim is the similarity. Ed is the edit distance between the question and the sentence. Qlen is the length of the questions. The smaller the Sim the more similarity is.

And the weights of the three similarities are different. The sentence similarity and paragraph similarity is more important. (Equation 2.3)

$$Rs = \alpha Re d + \beta Re p - \gamma Simm \quad (2.3)$$

The R_s is the rank score of the document. The R_{ed} is the document-question relevance and the R_{ep} is the biggest value of the paragraph-question relevancies in a document. The S_{imm} is the minimum of all the sentence-question similarities. α , β and γ is coefficient and $\alpha < \beta < \gamma$.

2.3 Answer extraction

First we orient the candidate sentences from the documents. There are two standards. First is the matching of key words and their expansions. The second is the similarity between the question and sentence. The next step is selecting the answers from the candidate sentences.

Different types of question focus on distinctness information. We should develop strategies for each type of question. The DEFINITION and BIOGRAPHY questions need the mainly brief information to a person or a concept. The knowledge from the encyclopedia is an important reference. Based on the concept and introduction from the encyclopedia we choose the answers from the candidate sentences. Comparing the sentences in the document and the sentences in the encyclopedia we select several most similarity sentences as answers.

The DATE question wants to know the date when an event happened. First we find the event by keywords and syntactic analysis. Second the words that indicate the date are selected. The characters "年", "月" and "日" indicate the date information. Similarly for other factoid questions the features that describe the demanded information are important. For example the word "位于" indicate a location information. For RELATIONSHIP and EVENT questions we use the keywords matching to select the answers. WHY question is the most difficult question. It wants to know the reasons of an event but event itself. We find the answers surround the sentences that indicate the event. And the words indicate reason such as "因为", "原因", "所以" and "为了" etc. are important sign to the answers.

3. Result

From the Official per-topic F-score definition based on nugget pyramid method we can find there are three points, the length of the answers, and the precision of matched nuggets and the recall of the nuggets.

And in the NTCIR-8 ACLIA, C value depends on the question type (see Table 2, column Avg Char Length). For the factoid questions the number of nuggets is just little more than 1. And the average length of the nuggets is less than 10 characters. It means that we should find one answer for each factoid question, which is usually a phrase. For the complex questions almost the numbers of nuggets is from 4 to 7. And the length of the nuggets is from 20 to 30 characters. The answers are sentence level. For this point we should cover all the nuggets and avert introducing the not relevant contents.

Let	
r	sum of weights over matched nuggets
R	sum of weights over all nuggets
a_{HUMAN}	# of nuggets matched in SRs by human
L	total character-length of SRs
C	character allowance per match
$allowanc$	$a_{HUMAN} \times C$
e	
Then	
$recall$	$= \frac{r}{R}$
$precision$	$= \begin{cases} 1 & \text{if } L < allowance \\ \frac{allowance}{L} & \text{otherwise} \end{cases}$
$F(\beta)$	$= \frac{(\beta^2 + 1) \times precision \times recall}{\beta^2 \times precision + recall}$

Figure 1 Official per-topic F-score definition Based on nugget pyramid method. [10]

Table 2 Macro-average nugget statistics over topics in CS [8]

type	Avg Nuggets	Avg Char Length	Avg Weight
DEFINITION	4.1	24.9	0.5
BIOGRAPHY	6.5	18.3	0.34
RELATIONSHIP	4.7	30.6	0.42
EVENT	6.0	24.1	0.44
WHY	4.0	29.8	0.48
PERSON,	1.0	3.4	0.93
ORGANIZATION,	1.0	9.6	0.87
LOCATION	1.2	9.6	0.97
DATE	1.4	4.7	0.87

3.1 The average length of answers

There two kinds of nuggets. For the complex questions (DEFINITION, BIOGRAPHY, RELATIONSHIP, EVENT, and WHY) the nuggets are sentence level. And for factoid questions (PERSON, ORGANIZATION, LOCATION, and DATE) the nuggets are phrase level. Unfortunately our answers are all the sentences in the document (see Table 3). It seems too long for the nuggets especially for the nuggets of factoid questions.

Table 3 the average length of answers

type	Average length of answers
DEFINITION	57.76
BIOGRAPHY	53.68
RELATIONSHIP	46.46
EVENT	59.26

WHY	158.19
PERSON,	63.82
ORGANIZATION,	69.34
LOCATION	57.73
DATE	54.72

We can find that the length of answers for the complex questions is twice to the length of nuggets. And the answers for factoid questions in the sentence level are too much exceeded. This is a serious error. We should extract more accurate information for the factoid questions.

3.2 The recall of matched nuggets

Compare with the results and nuggets we find that for the complex questions some of the results do not cover all the nuggets. But more than a half results cover all the nuggets.

Table 4 the precision and recall for complex questions

	recall	precision
Complex questions	64.5%	29.1%

The mainly related work with the recall of nuggets is Document Retrieval. The complex questions contain several nuggets. Some of the nuggets appear in many documents and some of them only appear in fewness documents. This kind of disequilibrium debases the recall.

3.3 The precision of matched nuggets

The mainly related work with the precision of nuggets is Answer Extraction. There is a lot of information that relates to the questions besides the correct answers. Some of them are even ambivalent. To find the answers from the anfractuositities is a difficult work.

4. Discussion

We find there are some differences between the answers that we submit and the evaluated result returned by the organizer. The returned result is IR4QA version but recognized as the CCLQA result. The CCLQA version is not evaluated. There are about half questions are different. And most of them have very low account even some are zero.

And the less of deep information extraction for the factoid questions is a critical error. This two points make APQA's performance doesn't satisfactoriness.

There are two key points of question answering system. First is the document retrieval. It should cover all the nuggets in the

limited documents. The second is that the information extraction finding the correct answers in all the related information.

In feature work, we should pay more attention on the two points.

5. Conclusion

This paper introduced the APQA System for CCLQA task. The three parts of the System are Question Analysis, Document Retrieval and Answer Extraction. The Question Analysis extracts the keywords of questions and confirms the types of the questions. Because of the particular interesting of information for different quotations there are different strategies in Document Retrieval and Answer Extraction.

6. REFERENCES

- [1] Diane Kelly, Jimmy Lin. Overview of the TREC 2006: ciQA Task. ACM SIGIR Forum 2006.
- [2] Che Wanxiang, Liu Ting, etc. Similar Chinese Sentence Retrieval based on Improved Edit-Distance. Chinese High Technology Letters, 2004
- [3] Liu qun, Chinese word similarity computing based on HowNet, Computational Linguistics and Chinese Language Processing, 2002
- [4] Buckley C, Salton G, etc. Automatic query expansion using SMART. Proceedings of the 3rd TREC, 1995
- [5] K. Hammouda and M. Kamel, "Incremental Document Clustering Using Cluster Similarity Histograms", The IEEE/WIC International Conference on Web Intelligence (WI 2003), pp: 597-601, 2003.
- [6] Andrew Hickl, Kirk Roberts, etc, Question Answering with LCC's CHAUCER-2 at TREC 2007
- [7] Jeongwoo Ko, etc, A Probabilistic Graphical Model for Joint Answer Ranking in Question Answering, ACM SIGIR Forum 2007.
- [8] Teruko Mitamura Overview of the NTCIR-8 ACLIA Tasks: Advanced Cross-Lingual Information Access
- [9] <http://Wikipedia.org>
- [10] Teruko Mitamura Overview of the NTCIR-7 ACLIA Tasks: Advanced Cross-Lingual Information Access
- [11] <http://aclia.lti.cs.cmu.edu/ntcir8>
- [12] <http://www.kenae.com/>
- [13] RA Wagner, MJ Fischer, The string-to-string correction problem, Journal of the ACM (JACM), 1974