Perspectives on Chinese Question Answering Systems

Cheng-Wei Lee[12], Cheng-Wei Shih[1], Min-Yuh Day[1], Tzong-Han Tsai[1], Tian-Jian Jiang[1],
Chia-Wei Wu[1], Cheng-Lung Sung[1], Yu-Ren Chen[1], Shih-Hung Wu[3], Wen-Lian Hsu[1§]

[1]Institute of Information Science, Academia Sinica, Taiwan, R.O.C
[2]Department of Computer Science, National Tsing-Hua University, Taiwan, R.O.C
[3]Department of CSIE, Chaoyang University of Technology, Taiwan, R.O.C

§corresponding author

{aska,dapi,myday,thtsai,tmjiang,cwwu,clsung,yrchen}@iis.sinica.edu.tw,
shwu@cyut.edu.tw, hsu@iis.sinica.edu.tw

Abstract

Question Answering (QA) is becoming an increasingly important research area in natural language processing. Since 1999, many international question answering contests have been held at conferences and workshops, such as TREC, CLEF, and NTCIR. Thus far, eleven languages – Bulgarian, Dutch, English, Finnish, French, German, Indonesian, Italian, Japanese, Portuguese, and Spanish – have been tested on monolingual or cross-lingual question answering tasks. Although Chinese is the second most popular language in the world, NTCIR only conducted the first QA contest in Chinese this year. The results reveal that there seems to be a performance gap between Chinese question answering systems and some systems of other languages.

In this paper, we review previous works on Chinese question answering, including our two systems on frequently asked questions and factoid questions. Comparing Chinese with other languages, word segmentation is a key problem in Chinese question answering. We review studies on word segmentation and discuss important issues, such as part-of-speech tagging, named entity recognition, deep and shallow parsing, semantic role and relation labeling etc., which are helpful for building QA systems. Machine learning approaches currently represent the main stream on many QA research issues, we believe, by efficiently utilizing the above resources, the performance of machine learning approaches can be improved further in Chinese question answering.

Keywords: Chinese Question Answering, QA Systems, FAQ system, Factoid Question Answering

1. Introduction

In recent years, question answering (QA) has become a key research area in several of the world's major languages due to the information overload caused by the rapid growth of the Internet. The first QA system was built by Green [25] in 1961 to provide answers to questions about the American Baseball League. Green's BASEBALL system was written in IPL, used IPL lists as storage structures, and ran on an IBM 7090 platform. The computing power of IBM 7090 was very limited, but its architecture was very complex – even by today's standards. Advances in computer hardware design, such as CPU speeds and memory capacity have made the implementation of QA systems much easier. In this section, we address general issues about QA systems, and discuss Chinese QA system and two implementations, namely, FAQ and Factoid QA in Sections 2 and 3 respectively.

From a human perspective, the behavior of QA systems is very straightforward. Usually, natural language questions are taken as input to the systems. Answers can be in many forms such as a word, a phrase, a sentence, or a document according to the system's capacity and the user's request. For example, in response to the question, "Who is the president of the United States?", a good QA system would analyze the question, search information sources, and then output "George W. Bush" as the answer directly. This kind of system would be an ideal way for people to search for information. However, current QA systems are far below human expectations. This is because natural language is not easy to process and interactive QA is even harder to model. Behind the simplistic behavior of QA systems, there are often complex mechanisms at work.

QA research is a multidisciplinary field, and systems are usually integrated from many techniques and resources. Information Retrieval (IR), Natural Language Processing (NLP), Information Extraction (IE), Machine Learning, and even Software Engineering techniques are all needed to build a QA system. IR is the most common component used in QA systems to retrieve relevant documents. It is a rather shallow technique compared to NLP and IE, but IR technology is very mature and there are many commercial or open source products that can deal with huge amounts of data efficiently and effectively. In contrast, the processing times of NLP and IE are longer, but they provide deeper analysis of questions and documents, which may be needed to answer some difficult questions.

Machine Learning is another approach to QA. Since natural language is full of noise, it is important to deal with real data. Using Machine Learning, computers can learn from tagged data sets. Noisy or inconsistent information could be filtered out to achieve better performance.

To integrate these multidisciplinary techniques in a QA system, a great deal of software development effort is needed, since the components may come from many research groups and may be created in different programming standards. Therefore, Software Engineering or good software development experience, would be helpful in building QA systems.

Information retrieval researchers have long recognized the need for a forum to compare different systems. The first Text REtreival Conference (TREC) was held in 1992. The motivation of TREC is:

> *In the 30 or 50 years of experimentation there have been two missing elements. First, although some research groups have used the same collections, there has been no concerted effort by groups to work with the same data, use the same evaluation techniques, and generally compare results across systems. The importance of this is not to show any system to be superior, but to allow comparison across a very wide variety of techniques, much wider than only one research group would tackle.* [31]

Only twenty-five groups participated in the first TREC, but the number has increased to more than one hundred in recent years. In addition to addressing traditional IR issues, many interesting Tracks (contests) have been held at the annual conference, including the first QA Track in 1999, which focused on providing short passage answers in fixed length.

The degree of difficulty of the TREC QA Track has gradually increased since the first contest. From 1999 [50] to 2004 [51], the answer corpus grew from 558,000 documents to 1,033,000 documents. Answers have become more complex from passage answers to exact list answers and definition answers. Some of the questions are even presented without answers. The creation of the questions has also changed from manually created questions by all participants to collection from real search engine logs.

TREC focuses on English QA evaluation, while the Cross Language Evaluation Forum (CLEF), which is held in Europe, deals with other major languages. In terms of the number of participating languages, QA@CLEF is the biggest evaluation forum. In 2005, ten languages – Bulgarian, Dutch, English, Finnish, French, German, Indonesian, Italian, Portuguese and Spanish – were included in the contest.

Although Chinese is the world's second most popular language, a Chinese QA contest was not held until this year. The first such contest was the Cross-Language Question Answering (CLQA) task at the *NII-NACSIS Test Collection for IR Systems* (NTCIR) Workshop. CLQA also contains other contests between Chinese, Japanese, and English. Besides CLQA, the NTCIR Workshop holds other tasks, namely, Japanese QA evaluation and Question Answering Challenge (QAC) [23; 33]. Section 2 will provide detailed information about the first CLQA contest.

QA systems can be categorized in many ways, such as by the application domain, answer sources, or target and source languages. In this section, we discuss some of these categories and introduce some important systems.

(a) Open-Domain QA and Restricted-Domain QA

Restricted-domain or closed-domain QA is only concerned with a specific domain, such as medicine or knowledge about some corporation. The possible questions are limited by the domain, therefore it is possible to encode all the domain knowledge or ontology in the system to analyze questions or answer sources. The answer sources can be fully structured data, which is easier to process. Obviously, Green's BASEBALL system is a restricted-domain QA system that only answers questions about one season's baseball data.

In contrast to restricted-domain QA, open-domain QA tries to answer almost anything. Ask Jeeves [1] is the most well-known open-domain QA system. According to The Nielsen/NetRatings MegaView Search report this year, Ask Jeeves is the fifth most popular search engine and the only natural language search engine on the list. Posing the question "*Who is the president of the United States*" to Ask Jeeves, elicits "*The Chief of State of the United States is President George W. Bush, who is also Head of State*" as the answer. To answer unrestricted questions, a general ontology or commonsense knowledge would be useful. WordNet [4] and Cyc [35] are two popular general resources used in many systems [17; 19; 30; 42].

(b) Database/FAQ/Newswire/Web QA

The answer source is an important factor in designing a QA system. Databases are the most popular answer source that can store structured data. Traditionally, databases have used Structured Query Language (SQL) to retrieve data. Although SQL is descriptive and is much easier than other programming languages, some systems provide natural language interfaces that are more intuitive for people to use. LUNAR [52] was probably the most successful database QA system in the early days with answers to questions about the rock samples brought back from the moon. Although LUNAR was built in the 1970s, its performance was impressive. It could answer 70% of questions correctly. Microsoft English Query [27] is a more recent commercial product, which uses knowledge of the English language and the underlying database structure to convert English questions into SQL statements. Although the performance of such database QA systems is acceptable, the technique is not widely used. Most developers still use SQL as the database query language and most end users still cannot retrieve data only by natural language.

Frequently Asked Questions (FAQs) represent another important answer source on the Internet or in business customer service systems. Unlike other QA systems that focus on the analysis of questions and the generation of the answers, FAQ systems only focus on processing input questions and matching them with FAQs. If we can find an appropriate FAQ for an input question, the answer can be output directly by looking up a list of question-answer pairs.

The answer source used in the QA contests mentioned in this paper are mostly newswire documents. For example, the AQUAINT Corpus used in TREC QA Track, consists of newswire text data drawn from three sources: the Xinhua News Service (People's Republic of China), the New York Times News Service, and the Associated Press Worldstream News Service. These kinds of corpora are good sources for QA system research, because both the quality and the quantity of the data are good. All major newspapers now provide digitalized versions of their publications for online viewing or searching. Therefore, the quality can be controlled and the amount of data increases daily. In addition, the content of newspapers is so general, which makes them good sources for open-domain QA research.

The success of the Google search engine shows that the Internet is a promising realm for QA systems. It is estimated that the Internet contains more than 7,500 terabytes of digital data. And according to ClickZ Stats' list of the latest statistics on Internet users, the worldwide Internet population is 1.08 billion. This data shows the potential of QA on Web. Therefore, some systems treat the Web as the answer source and some systems try to combine the web information with other answer sources to achieve better QA performance.

(c) Monolingual QA and Multilingual QA

Since questions and answers are represented by language. We can characterize QA systems by the source language, which represents questions, and the target language, which represents answers. Systems that participated in the TREC QA Track and NTCIR QAC are all monolingual QA systems, which use the same source and target languages. Monolingual QA is good for people who speak one of the popular languages, and researchers have paid a great deal of attention to monolingual QA research.

Multilingual QA, which has only emerged in the last few years as a complementary research task, represents a promising direction for two reasons. First, it allows users to interact with machines in their native languages, thus providing easier, faster, and more equal information access. Second, cross-lingual capabilities enable QA systems to access information stored in language-specific text collections.

(d) Factoid/List/Description QA

In addition to problem domains, answer sources, and languages, question types can be used to categorize QA. Different question types may require different strategies to deal with them. There are three question types: Factoid QA, List QA, and Description QA. Factoid QA is the simplest, as the answers are named entities, such as personal names, organization names, location names, etc. Some Factoid QA systems return fixed size short passages as answers, while others provide exact answers. The QA contests mentioned in this paper have Factoid QA tasks.

List QA is very similar to Factoid QA, except that a question may have more than one answer and the evaluation of List QA is based on the completeness of the answers. Description QA, on the other hand, is much complex because the answer is usually a paragraph describing the question focus and a summarization technique is needed to minimize the answer size. More detailed information about these QA types can be found in [50].

2. Chinese Question Answering

There are only a handful research teams working on Chinese QA, which may explain why the first Chinese QA contest was only held at this year’s NTCIR workshop (2005).

The NTCIR workshop held the first Chinese QA contests in CLQA, which is a rather simple QA task compared to recently held tasks, such as QAC, QA@CLEF, and TREC QA Track. The answers to CLQA questions are restricted to named entities, for example, proper nouns, such as the name of a person, an organization, and various artifacts; and numerical expressions, such as money, size, date, etc. The Chinese document corpus for finding answers is CIRB 4.0 which consists 901,446 Chinese news articles from United Daily News, United Express, Ming Hseng News, and Economic Daily News. The participants received a development set containing 200 sample questions and answers. The development set was used to build the system. After months of development, the set of test questions was given for the contest. Participating QA systems had to return one exact answer with the corresponding document’s id. The evaluation was carried out by collecting all the participants’ answers for human evaluators to analyze. The correct answers became the gold standard to evaluate participant systems. The evaluators marked three possible labels on each question when evaluating it. If the answer was not contained in the gold standard, the question was marked *wrong (W)*. If the answer was correct, the document containing it was examined to check whether or not it logically supported the answer, which was then marked *correct (R)* or *unsupported (U)* accordingly. After the evaluation, two metrics, *R Accuracy* and *R+U Accuracy*, were reported for each system.

For example, given the sample question we used previously, "*Who is the president of the United States?*" if a system returned "*Jordan*" as the answer, it would obviously be marked wrong. However, an answer from a document that says "*Bush is the president of USA*", would be treated as correct. But if a document does not contain any words related to "*president*", but contains a vague sentence like "*Bush leads American to ……*", the answer would be treated as unsupported.

Building a Chinese QA system is not an easy task due to some characteristics of the Chinese language, such as word segmentation, and insufficient basic research components. Also, there are very few works on such systems in the literature. Chinese QA systems are very similar to those for English, except that they usually contain an additional layer for Chinese word segmentation. Obviously, the ontology and linguistic resources are also different from English systems.

Chinese FAQ QA systems usually use semantic information. Hsu et al. [32] proposed a knowledge representation scheme, called InfoMap, that can be used to answer Chinese FAQs. InfoMap is a tree-like knowledge structure that represents the relations between concepts and syntactic patterns of concepts. (Knowledge) Nodes in InfoMap are fired if they match something in the question sentence, and the fired nodes are collected to calculate the score for each FAQ. We give more details of the system in Subsection 3.1. Qin et al. [44] compare two similarity measures of a question and an FAQ. The traditional vector space model (VSM), which is usually keyword-based, is very common in IR systems. Qin found that the performance of the keyword-based approach is not as good as their semantics-based approach, because of the much smaller data size of FAQ QA systems. The new approach incorporates HowNet to measure the distance between a question and an FAQ. The accuracy of the keyword-based and semantics-based approaches is 73.2% and 80.78% respectively. Wu et al. [53] created a system for answering medical questions. The system is much more complex than the ones we have mentioned previously. It has a question classification component that can classify questions into ten types, ie. what, when, where, why, how, degree, quantity, whether, relation, and capability. Question classification is usually employed in a factoid QA system to understand the users' intentions. Wu et al use a probabilistic mixture model, which incorporates question types and other aspects derived from WordNet and HowNet, to interpret the question and QA pairs. Their results show that the semantics-based approach is also good for medical domain FAQ QA.

Like other open-domain QA systems, Chinese open-domain QA systems usually employ pipeline architecture comprising four components: question processing, passage retrieval, answer extraction, and answer ranking. In 2001, Zhang et al. [24] conducted a large-scale experiment on Chinese open-domain QA. They collected a 1-gigabyte corpus containing data from encyclopedias, news, and the Web. There were 298 test questions produced by six students. The experiment achieved 43.62% accuracy, which is a promising result for an open-domain Chinese QA system. Although the experiment scale was big and the result was good, the corpus has not been released and detailed information about their system is not available. Meng et al. [40] proposed another open-domain newswire QA system that returns answers from 900 Chinatimes news articles. They use AutoTag to segment words and employ a word similarity measure based on HowNet to expand question words. The answer extraction of Meng's system is rule-based, which utilizes the information provided by the AutoTag result. Their system also achieves an excellent performance, as the MRR (Mean Reciprocal Rate) value is as high as 0.84. Although English is obviously very different to Chinese, some research on English QA has been adapted to Chinese QA. Guo's team [29] has participated in a number of TREC tracks, such QA Track, Robust Track, and Genomics Track. They have modified their MultiText system to deal with Chinese factoid questions. Like Zhang and Meng, Guo's team also collected their own corpus, which comprised a 17 GB web corpus, and the People's Daily newspaper and Xinhua newswires from 1991 to 1995. They conducted several experiments on the passage retrieval component, which showed that about ninety percent of answers could be found in the top 5 passages. There are performance differences between question types. Lin [36] conducted thorough experiments on different question types. The performance results of different question types are: MRR 0.446 for *short-answer* questions; MRR 0.443 for *DEFINITION* questions; and 0.418 for *PERSONDEF* questions.

After reviewing these Chinese QA systems, we would like to introduce some resources and techniques that would be useful for building a Chinese QA system. Thesaurus and ontology are two kind of important resources. Rule-based QA components need these resources for symbolic computation. And statistic-based QA components use them to calculate features for training and testing. Cilin and Sinica BOW are two important resources of this kind. Cilin is a Chinese thesaurus with a hierarchy structure which contains valuable semantic information of Chinese. It is a good knowledge source for Chinese processing. Sinica BOW was evolved from WordNet which is an lexical resource. Sinica BOW groups synonyms as *synsets*, provides short definitions and various semantic relations between these *synsets*. HowNet is another well-known common-sense knowledge base. Different from the approach adopted by WordNet, HowNet [2] tries to decompose concepts into basic concept components called *sememes*, Many researches have demonstrated the usefulness of these resources in QA systems.

In the following paragraph, we will describe some useful techniques for QA:

(a) Word Segmentation and Part-of-Speech (POS) Tagging

Chinese text is different from Western language text because it lacks explicit word boundaries. Therefore, word segmentation, which identifies delimiters for meaningful words from a Chinese sentence, is a necessary step in Chinese text processing. Researchers usually perform word segmentation by using techniques such as a statistical method [38], dictionary-based method [55], or syntax-based method [13]. A combination of these approaches is also popular [41]. The report in [22] gives the latest competition results for Chinese word segmentation. Many participants achieved over 95% in the F-score.

Part-Of-Speech (POS) tagging is another basic, well-known technique in natural language processing (NLP). Providing a proper morpho-syntactic tag for each word is very useful for many QA component. In Chinese QA, many POS tag systems, such as the Penn Treebank tagset [3] and the CKIP AutoTag tagset [5], have been introduced. Recent research has also achieved good performance in automatic POS tagging [11].

(b) Named Entity Recognition

Named entity recognition (NER) is very important in most QA applications. Named entities (NEs) are good candidates for answering factoid questions and are used in many QA modules.

The concept of a named entity (NE), which was first introduced at the Message Understanding Conference (MUC), defines some informative words, such as personal names, locations, organization names, and time and number expressions as information units in text [26]. Tracking these units is a helpful preprocessing step in information extraction (IE), especially for dealing with unstructured text like news articles. Recently, NE recognition tasks with different NE definitions have become one of the hottest topics in NLP. Many machine learning algorithms like the Hidden Markov Model (HMM) [9], Support Vector Machine (SVM) [6], Maximum Entropy (ME) [10], and Conditional Random Fields (CRF) [39] are applied in NE tracking and detection processing. The work in [45] introduced the shared task of NE recognition and the best performance reached 88% for English and 72% for German.

In Chinese, the word segmentation problem makes the NER problem harder. Previous works [46] on Chinese NER rely on the word segmentation module. However, an error in the word segmentation step might lead to errors in NER results. Some studies [28; 49] try to deal with this problem. They demonstrated the potential of char-based NER that doesn't take the word segmentation step.

(c) Full and Shallow Parsing

Parsing, which deals with syntactic analysis in text processing, merges several adjacent words into one syntactic unit. It then assigns a proper phrase category by matching some grammar rules or patterns and provides the syntactic structure of a text. The difference between deep and shallow parsing is that the latter, also known as chunking, focuses on local syntactic structures, whereas the former provides a thorough analysis of the text. Both rule-based [20] and statistics-based [14] approaches are used in parsing, which is a hot topic in NLP [8; 15; 18].

Chinese full parsing is very challenging [56; 61], because it is difficult to achieve high accuracy, and the performance is not suitable for online applications. Shallow parsing of Chinese, on the other hand, is promising and desirable in terms of efficiency. Researchers have developed related techniques [34; 37; 48; 47; 54; 60]. Most of these works use machine learning approaches, instead of the rule-based approach used in full parsing. Popular machine learning methods such as SVM, CRF, and ME, have been tested. Although full parsers can provide thorough information about sentences, shallow parsers are more suitable for QA systems, because the output of a shallow parser is more reliable and the processing time is shorter.

(d) Semantic role/relation labeling

Semantic roles are used to represent the semantic functions and relationships of the constituents of a sentence. Unlike parsing, semantic role labeling uses predicate-argument structures, which represent the core meaning of a sentence, for annotation. Various types of semantic role labeling systems, such as [16], [7], [57] and their annotated corpora have been introduced in recent years to help researchers annotate semantic roles automatically. The shared task of CONLL 2005 on Semantic Role Labeling is a good example of the progress that has been made [12]. The best result in that competition was 77% for the F-score [43]. Chinese semantic role labeling is a growing research area. Xue et al. and You et al. [58; 59] provides many viewpoints and experiences on it.

3. Academia Sinica Chinese Question Answering Systems

The Intelligent Agents System Laboratory of Academia Sinica has been researching Chinese QA systems for many years. Two systems have been developed. The first is an FAQ system, the Academia Sinica Question Answering System for FAQs (ASQA-FAQ), which can answer frequently asked questions of a restricted domain. FAQs are collected and analyzed to build a domain MAP, which can be treated as an informal ontology. The MAP is organized to reflect the hierarchy of domain concepts and the FAQs are scattered within the MAP.

The second system is an open-domain system for factoid questions. Most current QA systems, including those in Chinese, deal with factoid questions. Though there are many sophisticated QA systems, our factoid question answering system only employs shallow natural language processing techniques. We utilized available software modules and effectively integrated human knowledge and machine learning methods to achieve a good performance in the first Chinese QA contest held by NTCIR in Japan. The structure and performance of these two systems are described in the following subsections.

3.1 Answering Frequently Asked Questions (FAQs)

The FAQ system we created attempts to answer questions about small domains, such as an organization like Academia Sinica. The knowledge needed for such a small domain is controllable and FAQ lists usually exist. Therefore, very little effort is needed to collect FAQ pairs. The main goal of an FAQ QA system is to match an input question with the most appropriate FAQ, then output the corresponding answers. In our system, we try to parameterize the FAQ list. After generalization, some FAQs can be categorized into the same parameterized FAQ. This process has two major benefits. First, it gives semantic analysts an overview of the FAQs and helps generate our tree-like knowledge base, InfoMap, which we describe later. Also, since every question matches a parameterized FAQ, it could be used to generate related information, not just the answer the user wants. We use InfoMap to organize the knowledge for answering FAQs. Three kinds of knowledge, namely, concepts, relations, and syntactic templates, are stored in InfoMap. After receiving a question, a mechanism fires the templates and the related concepts through the relations between them. A scoring method is then applied to rank the matched FAQs.

3.1.1 Knowledge Representation

InfoMap has a tree-like structure though this is only a deceivingly simple statement since it does contain "references" that connect nodes on different branches. The root node is usually the name of a domain or a subject such as passport or department store. Following the root node, the first level nodes down are topics that users may be interested in. These topics have sub-categories that list related sub-topics.

There are some nodes, called function nodes, to label the relation between two other nodes in InfoMap. The basic function nodes are: category, attribute, example, synonyms and event. There are some function nodes to build the QA system, such as FAQ, FAQ condition, test query and some other infrequently used function nodes. These function nodes help to represent and identify query concepts. The synonym of a concept is listed under the function node synonym of each concept. Under the function node "example" are the examples of a concept. For example, if the concept is "hotel", then its examples can be the actual hotel names. Function node "FAQ" gives a typical question associated with the concept. Function node "FAQ conditions" are items in a query that can be substituted by examples such as cities, hotels and etc.

Figure 1 shows the related information of library (in Academia Sinica) and some FAQs. It can be seen from the figure that the information of a library is divided into three functions nodes: event, category and attributes at the first branch under the library agent. The category of library consists of a list of libraries such as chemistry library, earth sciences library, Chinese literature library, European and American studies library, life science library, economics library, and Information Science library. The attributes of library include admission,

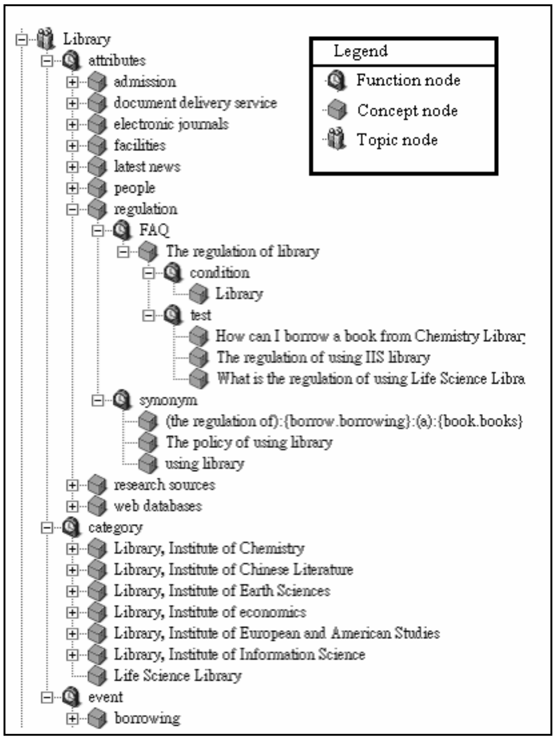
Library
attributes
admission
document delivery service
electronic journals
facilities
latest news
people
regulation
FAQ
The regulation of library
condition
Library
test
How can I borrow a book from Chemistry Librar
The regulation of using IIS library
What is the regulation of using Life Science Libra
synonym
(the regulation of):{borrow.borrowing}:(a):{book.books}
The policy of using library
using library
research sources
web databases
category
Library, Institute of Chemistry
Library, Institute of Chinese Literature
Library, Institute of Earth Sciences
Library, Institute of economics
Library, Institute of European and American Studies
Library, Institute of Information Science
Life Science Library
event
borrowing
Legend
Function node
Concept node
Topic node

Figure 1. FAQs and concepts

research sources, document collection, latest news, web databases, regulation, electronic journals and so on. The event of library can be “borrowing books from the library”. Nodes under the “Category” form a hierarchy, which is taxonomy of library. Each node under “attribute” forms a new hierarchy, which is not part of the taxonomy hierarchy of library, but is relevant to it. The nodes under “event” are similar. We classify the relevant concept into “attribute” and “event” based on the concept belong to noun or verb. This is a special criterion of Chinese, since the verb and noun have no morphological difference in Chinese.

3.1.2 The Firing Mechanism

In order to understand the meaning of an open query, we designed a firing mechanism to identify the most probable context and the most likely FAQ. We say that an concept node or template node is fired when it is related to the input string in some way. Concept nodes and template nodes would be fired if the syntactic pattern described by the concept name or by the template rule appears in the question. Other concepts can be fired if they are connected to the fired nodes through specific function nodes, which is called propagate firing or reference firing. Once the firing process ends, then the system collects all the fired nodes. A fired node in the target topics will be assigned a score according to the length of the string that fired this node. If this node is in a predefined “non-context” area, then its score is one per Chinese character. Otherwise, its score is ten per Chinese character. The non-context area contains nodes, which represent common concepts. These concepts are less helpful to identify the context of an open query. Afterward, collecting all the FAQ nodes of which the parent is fired, and calculating a path score by sum up all the fired node scores from root to this FAQ node. Finally, sort FAQs according to the total score.

3.1.3 Academia Sinica FAQ System

Academia Sinica FAQ System is an online demonstrative system of our FAQ QA system. Academia Sinica is a government funded research organization, which has 26 independently running institutes and several thousand employees. The amount of information in the Websites is very large. In 2001, The total number of Web pages is well over 80,000. It is hard to find information from the Web pages of 26 institutes in a uniform manner. Therefore, we constructed the Academia Sinica QA System to retrieval information from them. We collected and identified 626 distinct FAQ question types. With the possible combination of different conditions, the number of query concepts can be as much as 12,876. Among them, 10100 query concepts have associated URLs that answers the query. The remaining query concepts have no answer yet.

3.2 Answering Factoid Questions

ASQA for Factoid Questions (ASQAFQ) can answer factoid questions, such as “Who is the president of the United States?”, returning “Bush” as the exact answer. The architecture of ASQAFQ is shown in Figure 2. The architecture of ASQA comprises four main components: Question Processing, Passage Retrieval, Answer Extraction, and Answer Ranking. Questions are analyzed to obtain question types (QTypes), segments, focuses (QFocuses), and other limitations (QLimitations). Through a simple mapping table, question types are used to constrain possible answer types. Documents are segmented and indexed by both characters and words. After question analysis, we extract query terms from the question segments and construct queries from the terms to retrieve possible document passages, which are then sent to a named entity recognition system to obtain answer candidates. Finally, answers are ranked according to the needs of the question focuses.

3.2.1 System Description

When ASQA receives a question, it is analyzed by the Question Processing module to obtain question segments, question types, question focuses, and other question limitations. We can identify 6

Table 1. Taxonomy of Question Types.

Coarse-grained	Fine-grained	Coarse-	Fine-grained
PERSON|人	APPELLATION|稱謂	ARTIFACT|物	COLOR|顏色
	DISCOVERERS|發現者		CURRENCY|貨幣
	FIRSTPERSON|第一人		ENTERTAINMENT|娛樂
	INVENTORS|發明者		FOOD|食物
	OTHER|人其他類		INSTRUMENT|工具
	PERSON|人名		LANGUAGE|語言
	POSITIONS|職位		OTHER|物其他類
LOCATION|地	ADDRESS|地址		PLANT|植物
	CITY|城市		PRODUCT|產品
	CONTINENT|大陸、大洲		SUBSTANCE|物質
	COUNTRY|國家		VEHICLE|交通工具
	ISLAND|島嶼		ANIMAL|動物
	LAKE|湖泊		AFFAIR|事件
	MOUNTAIN|山、山脈		DISEASE|疾病
	OCEAN|大洋		PRESS|書報雜誌
	OTHER|地其他類		RELIGION||宗教
	PLANET|星球	NUMBER|數	AGE|年齡
	PROVINCE|省		AREA|面積
	RIVER|河流		COUNT|數字
ORG.|組織	BANK|中央銀行		LENGTH|長度
	COMPANY|公司		FREQUENCY|頻率
	OTHER|組織其他類		MONEY|金額
	POLITICALSYSTEM|政治體系		ORDER|序數
	SPORTTEAM|運動隊伍		OTHER|數值其他類
	UNIVERSITY|大學		PERCENT|比例
TIME|時間	DATE|日期		PHONENUMBER|電話號碼、郵遞區
	DAY|日		RANGE|數字範圍
	MONTH|月		SPEED|速度
	OTHER|時間其他類		TEMPERATURE|溫度
	RANGE|時間範圍		WEIGHT|重量
	TIME|時間		
	YEAR|年		

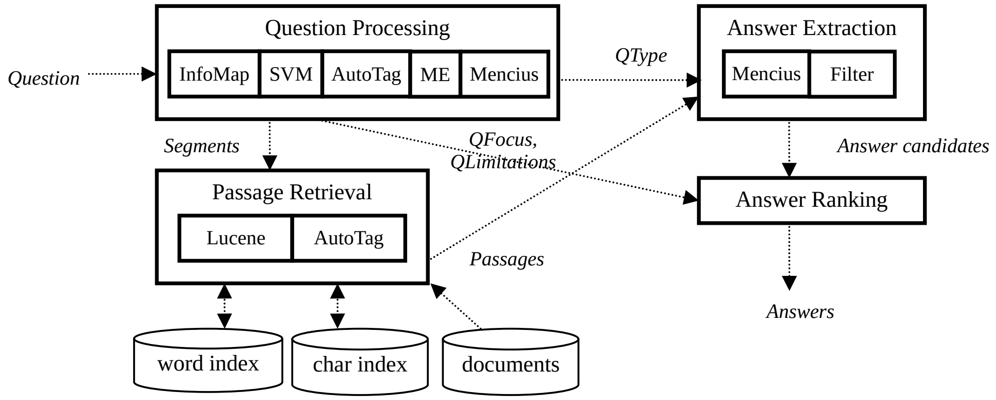

Figure 2. System architecture and data flow of ASQA. The outer rectangles are the four main modules. The inner rectangles are important sub-modules. The dashed arrows indicate the data flow

coarse-grained question types (PERSION, LOCATION, ORGANIZATION, ARTIFACT, TIME, and NUMBER) and 62 fine-grained question types as shown in Table 1. We adopt an integrated knowledge-based and machine learning approach for Chinese question classification.

We use InfoMap [3], which uses template rules to model Chinese questions as the knowledge-based approach, and adopt SVM (Support Vector Machines) [7] as the machine learning approach for a large collection of labeled Chinese questions.

Table 2. Examples of QFocus Analysis. All question focuses and limitations are in parentheses; “QF” means the question focus, “QFD” is the description of the question focus, “TI” represents time, and “NE” denotes the named entities in the sentence.

請問 [西元2000年7月/**TI**] [美方/**NE**] 派何人前往 [北京/**NE**] 對TMD以及其他全球戰略佈局與中方展開對話? July, 2000 USA Beijing *Who is the delegate of United States visiting Beijing to negotiate the TMD issue in July, 2000?*
請問 [2000年/**TI**] 的 [G8高峰會/**NE**] 在 [日本/**NE**] 何地舉行? Year 2000 G8 summit Japan *Which city in Japan hosted the G8 summit in 2000?*
請問 [芬蘭第一位女總統/**QF**] 為誰? Finland's first woman president *Who is the Finland's first woman president?*
請問 [2000年/**TI**] [沉沒於北極圈巴倫支海/**QFD**] 的 [俄羅斯核子潛艇/**QF**] 的名字? Year 2000 sank in the Barents Sea Russian nuclear submarine *Which Russian nuclear submarine sank in the Barents Sea in 2000?*
請問 [涉嫌竊取美國洛薩拉摩斯實驗室核武機密/**QFD**] 的 [華裔科學家/**QF**] 為誰? accused of violating……National Laboratories Chinese scientist *Which Chinese scientist was accused of violating Atomic Energy Act because of his purportedly mishandling restricted data of Los Alamos National Laboratories?*

In addition to question segments and types, we conduct QFocus analysis to extract the question focus and other QLimitations to fully capture the main purpose of the question. Table 2 shows some manually annotated examples of QFocus analysis. All the examples in this paper are taken from the NTCIR-5 CLQA development set or test set. The second example has no QFocus, but has three QLimitations: one of time and two of named entities. In contrast, the third example has only one QFocus and no QLimitations.

In the passage retrieval stage, the documents are preprocessed by AutoTag. Empirical rules are used to combine short Chinese terms into meaningful longer terms. We use an off-the-shelf IR engine, Lucene, to index the documents. The question is also segmented and combined by the same empirical rules. After filtering by a stop word list, the remaining question terms are treated as keywords to form Lucene queries.

Two Lucene queries are constructed for each question. All the query terms are connected and weighted via Lucene's boosting operator. In the initial query, quoted terms and nouns are set as required. If this query does not return a result, we retry a relaxed version of it that does not assign any query term as required.

Taking the question:「請問台灣童謠「天黑黑」是由哪位作曲家所創作？」(Who was the composer of the Taiwanese nursery rhyme "Dark Dark Sky?") as an example, the segmentation result is:

請問(VE) 台灣(Nc) 童謠(Na) 「(PARENTHESISCATEGORY) 天(Nc) 黑黑(VH) 」(PARENTHESISCATEGORY) 是(SHI) 由(P) 哪(Nep) 位(Nf) 作曲家(Na) 所(D) 創作(VC) ？(QUESTIONCATEGORY)

After combination and filtering the result, five keywords are extracted, i.e. “作曲家”,”台灣”,”創作”,”童謠”, and “天黑黑”. We then use those keywords to create Lucene queries such as:

+"作曲家"^1.2 +"台灣"^1.2 "創作"^0.7 +"童謠"^1.2 +"天黑黑"^2

To retrieve document passages for answer extraction. We perform a two-step answer extraction process. First, an online named entity recognition (NER) system is used to retrieve passages and obtain answer candidates. Second, the extracted named entities are filtered based on the expected answer types derived in the question processing phase.

In the answer ranking phase, we use QFocus and QLimitations to sort the answer candidates derived from the Answer Extraction module. An answer candidate is given a ranking score if it fits the answer focus or limitations of the question. The candidate with the highest score is the one that fits the most clues of the question, and is therefore regarded as the top 1 answer to the question. The ranking score of answer candidate a_{ij} in passage p_i is calculated as follows:

$$\mathrm{Score}(a_{ij}) = \frac{\sum_{k=1}^{m} \mathrm{Exist}(p_i, ne_k)}{NE\_Number} + \frac{\sum_{l=1}^{o} \mathrm{Exist}(p_i, cue_l)}{CUE\_Number} + \mathrm{QFI}(a_{ij}) + \mathrm{QFA}(a_{ij}),$$

where

- p_i is the selected passage and a_{ij} is the ***j-th*** answer candidate extracted from p_i;
- $NE = \{ne_1, ne_2, \cdots, ne_m\}$ is the named entity set appearing in the question;
- $CUE = \{cue_1, cue_2, \cdots, cue_o\}$ are other question limitations, except named entities.
- $\mathrm{Exist}(p_i, ne_k) = \{1,0\}$, which represents the matching bonus score of related named entities. If the source passage p_i of answer candidate a_{ij} contains $ne_k \in NE$, then $\mathrm{Exist}(p_i, ne_k) = 1$; otherwise $\mathrm{Exist}(p_i, ne_k) = 0$.
- $\mathrm{Exist}(p_i, cue_l) = \{1,0\}$, which is the answer cue's matching bonus score. The calculation of $\mathrm{Exist}(p_i, cue_l)$ is similar to that of $\mathrm{Exist}(p_i, ne_k)$.
- NE_Number and CUE_Number are the number of named entities in NE and the number of cues in CUE respectively.
- $\mathrm{QFI}(a_{ij})$ indicates the extra score if answer candidate a_{ij} comprises the question focus string.
- $\mathrm{QFA}(a_{ij})$ indicates the extra score if answer candidate a_{ij} is adjacent to the question focus string.

3.2.2 Performance

This architecture was applied to the NCTIR CLQA task. According to the evaluation results, the *R-Accuracy* of our system is 37.5 and the *R+U Accuracy* is 44.5, both of which are better than other

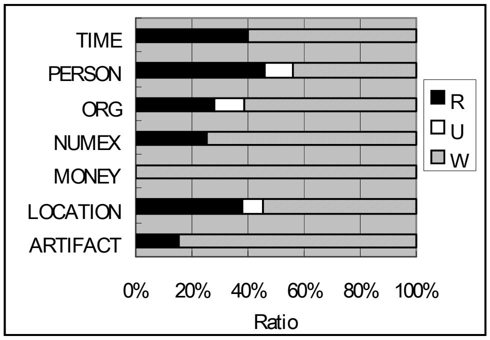

Figure 3. Accuracy of QA system by question type. R: correct answers, U: unsupported answers, W: wrong answers

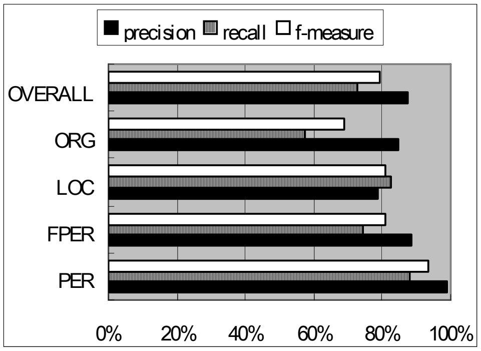

Figure 4. NER overall performance, plus the performance of organization names, location names, person names and foreign person names

Chinese QA systems that participated in CLQA. However, the results are not comparable with other Chinese QA systems mentioned in Section 2, because the scale of CLQA is much larger than previous research evaluation efforts.

R-Accuracy or *R+U Accuracy* represent a single global index that is useful for comparing the performance of different systems. We find that the performance varies from QType to QType. As Figure 3 shows, PERSON is the best performing QType in our CLQA result. The primary reason is that our high NER accuracy for the PERSON type entity. By comparing Figure 3 and Figure 4, we can observe the relation between NER performance and QA performance for each QType.

There was no significant performance difference between Chinese and the other languages' first attempts in QA@CLEF and QAC. The state-of-the-art performance of some languages, such as French and Portuguese, achieved more than sixty accuracy, which is close to the performance of English systems in TREC. There is still much work to be done to improve the Chinese Factoid QA system.

3.2.3 Answering Factoid Questions from Web

The architecture we designed is not only for answering questions about off-line newswire corpora like the one in NTCIR CLQA. By replacing the passage retrieval module with a search engine, we can turn ASQA into a Web QA system (ASQAWeb). Various types of information are provided by search engine results. The resulting web page is certainly important for QA, but the contents are noisy and maybe unavailable. Therefore, we use snippets from the web pages returned by the search engine. The quality of snippets is better than the web pages.

We experiment with four configurations, using the same test set provided by CLQA. The purpose of using four configurations is to determine the best strategy for applying question keywords to a search engine. Search engines usually provide a *required operator*, which lists specific words that have to appear in the returned web pages. The *required operator* is applied to every query and the maximum number of retrieved snippets is restricted. The configurations are as follows:

- **Full-keyword search**: All the keywords of the question are used.
- **Required-keyword search**: Only required keywords are used
- **Loose two-step keyword search**: a full-keyword search is applied first. If it does not return any result, a required-keyword search is applied.
- **Strict two-step keyword search**: a full-keyword search is applied first. If there is not enough result according to the maximum number of snippets, a required-keyword search is applied.

After searching, we extract the snippets from web pages returned by search engine. The snippets are then split into sentences and passed to the answer extraction module.

Table 3. Performance of ASQAWeb configurations

Snippet Amount	Configuration	R+U Accuracy	MRR
30	Full-keyword search	24.0%	0.27
	Loose two-step keyword search	26.0%	0.29
	Strict two-step keyword search	28.0%	0.32
	Required-keyword search	27.0%	0.30
50	Full-keyword search	23.0%	0.27
	Loose two-step keyword search	24.0%	0.28
	Strict two-step keyword search	26.0%	0.30
	Required-keyword search	26.5%	0.30

Table 3 shows the evaluation results of the four configurations. We note that the results of 30 snippets are better than those for 50 snippets, which may indicate that too many web results confuse the answer extraction and answer ranking modules. Based on our experiment, a total of 30 snippets is the most suitable number for searching in our system.

4. Conclusion

QA systems are extremely complex, but their potential is unlimited. In this paper, we have reviewed three international QA contests: TREC QA Track, QA@CLEF, and NTCIR CLQA. Among the three contests, TREC QA Track is the only one dedicated to monolingual QA. The other two contests deal with both monolingual and multilingual QA. TREC QA Track has become so successful that more than twenty groups have participated in recent years. We have discussed some ways to categorize QA systems. In addition to categorizing QA systems as monolingual or multilingual according to the language of questions and corpus, we also divide them into open-domain QA and restricted-domain QA systems. The answer source is another way to classify a QA system.

One of the major differences between Chinese QA and English QA, is that word segmentation problem is an important issue in the former, since it is the basis for other deep processing techniques. There is a great deal of ongoing research into Chinese named entity recognition, parsing, and semantic role labeling, which will help improve QA systems. Some QA systems have demonstrated the potential of incorporating semantic information in Chinese QA.

We have implemented two kinds of QA system; the first is a restricted-domain FAQ system, while the other is an open-domain newswire and web QA system. The Academia Sinica FAQ system demonstrates the concept of matching questions and FAQs through the InfoMap knowledge representation scheme; and the ASQA system for Factoid Questions produced good results in the NTCIR-5 CLQA task.

Although the accuracy of English QA is over seventy percent, it does not mean that current QA technology is good enough for real applications. There are still many shortcomings in evaluating QA systems that may affect their application to domains. Diekema et la. [21] discuss five evaluation dimensions of restricted-domain QA systems: system performance, answers, database content, display, and expectations. These dimensions would also be useful for evaluating open-domain QA systems.

5. References

1. "Ask Jeeves," http://www.ask.com/
2. "HowNet," http://www.keenage.com/
3. "Penn Treebank," http://www.cis.upenn.edu/~treebank/
4. *WordNet: An Electronic Lexical Database* The MIT Press, 1998.
5. "Autotag, CKIP, Academia Sinica," http://ckipsvr.iis.sinica.edu.tw/
6. Asahara, M., and Matsumoto, Y. "Japanese Named Entity Extraction with Redundant Morphological Analysis," HLT-NAACL, 2003.
7. Baker, C., Fillmore, C., and Lowe, J. "The Berkeley framenet project," Proceedings of COLINGACL, Singapore, 1998.
8. Bikel, D. "Design of a multi-lingual, parallel-processing statistical parsing engine," Human Language Technology Conference (HLT), 2002.
9. Bikel, D., Miller, S., Schwartz, R., and Weischedel, R. "Nymble: a High-Performance Learning Name Finder," 1997.
10. Borthwick, A., Sterling, J., Agichtein, E., and Grishman, R. "Exploiting Diverse Knowledge Sources via Maximum Entropy in Named Entity Recognition," WVLC98, 1998.
11. Brill, E. "Transformation-Based Error-Driven Learning and Natural Language Processing: A Case Study in Part-of-Speech Tagging," *Computational Linguistics*) 1995.
12. Carreras, X., and Màrquez, L. "Introduction to the CoNLL-2005 Shared Task: Semantic Role Labeling," CONLL2005, 2005.
13. Chang, C.-H., and Chen, C.-D. "A study of integrating Chinese word segmentation and part-of-speech tagging," *Communications of COLIPS* (3:1) 1993, pp 69-77.

14. Charniak, E. "Statistical Parsing with a Context-Free Grammar and Word Statistics," AAAI/IAAI, 1997.
15. Charniak, E. "A maximum-entropy-inspired parser," Proceedings of NAACL, 2000.
16. Chen, K.-J., Huang, C.-R., Chen, F.-Y., Luo, C.-C., Chang, M.-C., and Chen, C.-J. "Sinica Treebank: Design Criteria, Representational Issues and Implementation," in: *Building and Using Parsed Corpora,* A. Abeill´e (ed.), Kluwer, 2004.
17. Chu-Carroll, J., Prager, J., Welty, C., Czuba, K., and Ferrucci, D. "A Multi-Strategy and Multi-Source Approach to Question Answering," Text REtrieval Conference (TREC), 2002.
18. Collins, M., and Marcus, M. "Head-driven statistical models for natural language parsing," *Computational Linguistics*) 2003.
19. Curtis, J., Matthews, G., and Baxter, D. "On the Effective Use of Cyc in a Question Answering System," IJCAI Workshop on Knowledge and Reasoning for Answering Questions, 2005.
20. Daelemans, W., Buchholz, S., and Veenstra, J. "Memory-based shallow parsing," Proceedings of CoNLL, 1999.
21. Diekema, A.R., Yilmazel, O., and Liddy, E.D. "Evaluation of Restricted Domain Question-Answering Systems," ACL, 2004.
22. Emerson, T. "The second international Chinese word segmentation bakeoff," The 4th SIGHAN Workshop, IJCNLP'05, 2005.
23. Fukumoto, J.i., Kato, T., and Masui, F. "Question Answering Challenge for Five ranked answers and List answers - Overview of NTCIR4 QAC2 Subtask 1 and 2," NTCIR Workshop, 2004.
24. Gang, Z., Ting, L., Shi-Fu, Z., Wan-Xiang, C., Bing, Q., and Sheng, L. "Research on Open-domain Chinese Question-Answering System," 20th Annual Meeting of the Chinese Information Processing Society of China, 2001. (written in Chinese)
25. Green, B., Wolf, A., Chomsky, C., and Laughery, K. "BASEBALL: an automatic question answerer," in: *Readings in natural language processing*, Morgan Kaufmann Publishers Inc., 1986, pp. 545-549.
26. Grishman, R., and Sundheim, B. "Message Understanding Conference - 6: A Brief History," Proceedings of COLING-96, 1996.
27. Gunderloy, M., and Sneath, T. *SQL Server Developer's Guide to OLAP with Analysis Services* Sybex, 2001.
28. Guo, H., Jiang, J., Hu, G., and Zhang, T. "Chinese Named Entity Recognition Based on Multilevel Linguistic Features," International Joint Conference on Natural Language Processing (IJCNLP), 2004.
29. Guo, Y. "Chinese Question Answering with Full-Text Retrieval Re-Visited," Waterloo, 2004.
30. Harabagiu, S., Moldovan, D., Clark, C., Bowden, M., Williams, J., and Bensley, J. "Answer Mining by Combining Extraction Techniques with Abductive Reasoning," Text REtrieval Conference (TREC), 2003.
31. Harman, D. "Overview of the First Text REtrieval Conference (TREC-1)," Text REtrieval Conference (TREC), 1992.
32. Hsu, W.-L., Wu, S.-H., and Chen, Y.-S. "Event Identification Based on the Information Map - INFOMAP," IEEE International Conference on Natural Language Processing and Knowledge Engineering (IEEE NLP-KE), 2001.
33. Kato, T., Fukumoto, J.i., and Masui, F. "Question Answering Challenge for Information Access Dialogue - Overview of NTCIR4 QAC2 Subtask 3," NTCIR Workshop, 2004.
34. Le, Z., Xue-qiang, L., Yan-na, S., and Tian-shun, Y. "A Statistical Approach to Extract Chinese Chunk Candidates from Large Corpora," 20th International Conference on Computer Processing of Oriental Languages, 2003.
35. Lenat, D. "Cyc: A Large-Scale Investment in Knowledge Infrastructure," *Communications of the ACM* (38:11), November 1995.
36. Lin, C.-J. "A Study on Chinese Open-Domain Question Answering Systems," in: *Department of Computer Science and Information Engineering*, National Taiwan University, 2004.
37. Lu, Q., Zhou, J., and Xu, R.-F. "Machine Learning Approaches for Chinese Shallow Parsers," International Conference On Machine Learning And Cybernetics, 2003.
38. Lua, K.T., and Gan, G.W. "An application of information theory in Chinese word segmentation," *Computer Processing of Chinese & Oriental Languages* (8:1) 1994, pp 115-124.
39. McCallum, A., and Li, W. "Early Results for Named Entity Recognition with Conditional Random Fields: Fetures Induction and Web-Enhanced Lexicons," CoNLL 2003, 2003.

40. Meng, I.H., and Yang, W.P. "The design and implementation of chinese question and answering system," Computational Science and Its Applications - Iccsa 2003, Pt 1, Proceedings, 2003, pp. 601-613.
41. Nie, J.Y., Hannan, M.L., and Jin, W.Y. "Unknown word detection and segmentation of Chinese using statistical and heuristic knowledge," *Communications of COLIPS* (5:1&2) 1995, pp 47-57.
42. Prager, J., Chu-Carroll, J., and Czuba, K. "Question Answering Using Constraint Satisfaction: QA-By-Dossier-With-Contraints," The 42nd Annual Meeting of the Association for Computational Linguistics, 2004.
43. Punyakanok, V., Koomen, P., Roth, D., and Yih, W.-T. "Generalized Inference with Multiple Semantic Role Labeling Systems," CONLL2005, 2005.
44. Qin, B., Liu, T., Wang, Y., Zheng, S.-F., and Li, S. "Chinese Question Answering System Based on Frequently Asked Questions," *Journal of Harbin Institute of Technology*) 2003. (written in Chinese)
45. Sang, E.F.T.K., and Meulder, F.D. "Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition," Proceedings of CoNLL-2003, Edmonton, Canada, 2003, pp. 142-147.
46. Sun, J., Gao, J.F., Zhang, L., Zhou, M., and Huang, C.N. "Chinese Named Entity Identification Using Class-based Language Model," the 19th International Conference on Computational Linguistics, 2002.
47. Tan, Y., Yao, T., Chen, Q., and Zhu, J. "Chinese Chunk Identification Using SVMs plus Sigmoid," The First International Joint Conference on Natural Language Processing, 2004.
48. Tan, Y., Yao, T., Chen, Q., and Zhu, J. "Applying Conditional Random Fields to Chinese Shallow Parsing," CICLing, 2005.
49. Tsai, T.-H., Wu, S.-H., Lee, C.-W., Shih, C.-W., and Hsu, W.-L. "Mencius: A Chinese Named Entity Recognizer Using Maximum Entropy-based Hybrid Model," *Computational Linguistics & Chinese Language Processing* (9) 2004, pp 65-82.
50. Voorhees, E.M. "The TREC-8 Question Answering Track Report," Text REtrieval Conference (TREC), 1999.
51. Voorhees, E.M. "Overview of the TREC 2004 Question Answering Track," Text REtreival Conference (TREC), 2004.
52. Woods, W.A. "Progress in Natural Language Understanding - an application to lunar geology," American Federation of Information Processing Societies, 1973, pp. 441-450.
53. Wu, C.-H., Yeh, J.-F., and Chen, M.-J. "Domain-specific FAQ Retrieval Using Independent Aspects," *ACM Transaction on Asian Language Information Processing* (4:1) 2005, pp 1-17.
54. Wu, S.-H., Shih, C.-W., Wu, C.-W., Tsai, T.-H., and Hsu, W.-L. "Applying Maximum Entropy to Robust Chinese Shallow Parsing," Proceedings of ROCLING2005, 2005.
55. Wu, Z., and Tseng, G. "ACTS: An automatic Chinese text segmentation system for full text retrieval," *Journal of the American Society for Information Science* (46:2) 1995, pp 83-96.
56. Xix, X., and Wu, D. "Parsing Chinese with an Almost-Context-Free Grammar," Conference on Empirical Methods in Natural Language Processing (EMNLP), 1996.
57. Xue, N., and Palmer, M. "Annotating the Propositions in the Penn Chinese Treebank," Proceedings of the 2nd SIGHANWorkshop on Chinese Language Processing, Sapporo, Japan, 2003.
58. Xue, N., and Palmer, M. "Automatic Semantic Role Labeling for Chinese Verbs " Proceedings of the 19th International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, 2005.
59. You, J.-M., and Chen, K.-J. "Automatic Semantic Role Assignment for a Tree Structure," Proceedings of the 3rd SigHAN Workshop on Chinese Language Processing, ACL-04, Barcelona, 2004.
60. Zhao, T.-J., Yang, M.-Y., Liu, F., Yao, J.-M., and Yu, H. "Statistics Based Hybrid Approach to Chinese Base Phrase Identification," 2001.
61. Zhou, M. "A Block-based Robust Dependency Parser for Unrestricted Chinese Text," The second Chinese Language Processing Workshop attached to ACL2000, 2000.