Proceedings of the Sixth International Workshop on Information Retrieval with Asian Languages, July 2003, pp. 92-99.

AnyQ: Answer Set based Information Retrieval System

Hyo-Jung Oh, Myung-Gil Jang

Electronics and Telecommunications Research Institute (ETRI)
Daejeon, Korea
{ohj, mgjang}@etri.re.kr

Moon-Soo Chang

Department of Software
Seokyeong University
Seoul, Korea
cosmos@skuniv.ac.kr

Abstract

The accuracy of IR result continues to grow on importance as exponential growth of WWW, and it is therefore increasingly important that appropriate retrieval technologies be developed for the web. We explore a new type of IR, "answer set based IR", and its operational experience. Our proposed approach attempts to provide high quality answer documents to user by maintaining a knowledge base with expected queries and corresponding answer document. We will elaborate on our architecture and the experimental results.

Keywords: *answer set driven IR, attribute-based classification, automatic knowledge base construction,.*

1. Introduction

The goal of Information Retrieval (IR) is finding answer suited to user question from massive document collections with satisfied response time. With the exponential growth of information on the Web, user is expecting to find answer more fast with less effort. Current IR systems especially focus on improving precision the result rather than recall. A notable trend in IR is to provide more accurate, immediately usable information as in Question Answering systems(Q/A) [1] or in some systems using pre-constructed question/answer document pairs [2, 3], known "answer set driven" system. While traditional search engine uses term indexing, i.e. tf*idf, answer approaches use syntactic, semantic and pragmatic knowledge provided expert, i.e. WordNet[4]. Another difference comes from the fact that answer approach returns "answer set" distilled information need of user as retrieval result, not just document appeared query terms.

The TREC Q/A track [1, 5, 6] which has motivated much of the recent work in the field focuses on fact-based, short-answer question type, e.g. "*Who is Barbara Jordan?*" or "*What is Mardi Gras?*. The Q/A runs find an actual answer in TREC collection, rather than a ranked list of documents, in response to a question. On the other hand, user queries in answer set driven system, like AskJeeves[2], are more implicit and conceptual. These system was developed targeting the Web [7, 8], is larger than the TREC Q/A document collection. Whereas the user gives incomplete query to system, they need not only answers but related information. Sometimes the user even has uncertainty what exactly they need. For example, the user query just "*Paris*" is answered by gathering information including Paris city guide, photographs of Paris, and so on. To catch information need of user, these system have pre-defined query pattern and prepared correct answers belonging to each question. Since it is still considered difficult, if not impossible, to capture semantics and pragmatics of sentences in user queries and documents, such systems require knowledge bases built manually so that a certain level of quality can be guaranteed. Needless to say, this knowledge base construction process is labor-intensive, typically requiring significant and continuous human efforts [9].

This paper rests on the both directions: a new type of IR and its operational experience. Our system, named "*AnyQ*"[1], attempts to provide high quality answer documents to user queries by maintaining a knowledge base consisting of expected queries and corresponding answer document. We defined the semantic category of the answer as *attributes* and the

[1] http://anyq.etri.re.kr in korean

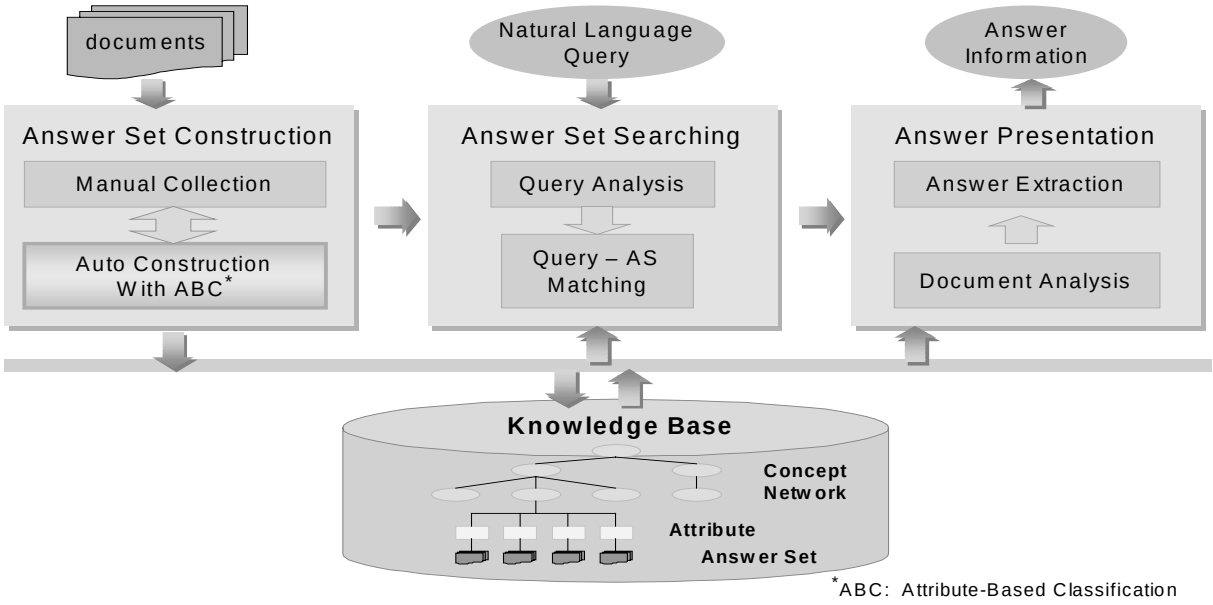

Figure 1. System architecture of Answer Set based IR

documents associated with each attributes as *answer set*. In order to reduce the cost of manually constructing and maintaining answer sets, we have devised a new method of automating the answer document selection process by using the automatic text categorization, reported ABC(Attribute-Based Classification)[10].

The rest of the paper is organized as follows. Section 2 presents overviews of our answer set driven retrieval system and knowledge base. In Section 3 and 4 elaborates on answer set construction and its retrieval process. Section 5 details experiment results for our method. After discussing the limitations of our approach in Section 6, we conclude by summarizing our contributions and describing future works.

2. Answer Set based IR System

2.1. System Overview

Several approaches to find answer using informative knowledge from expert were reported [1, 2, 3]. Most recent research proposed a new method of capturing the semantics of the question and then presents the document as answer, named “answer set driven IR. The goal of these systems is to explore how does map user question into answer document that might be contain pertinent information. In these systems, it is crucial to devise a method to construct a high-quality knowledge base. In our system, we take a hybrid approach of using a human-generated concept hierarchy and automatic classification techniques to make it more feasible to build an operational system.

Our system analyzes a user query to extract concept and attribute terms that can be matched against the knowledge base where a set of answer documents can be found. As such, it has three parts: *answer set construction*, *answer set search*, *answer presentation*, as illustrated in Figure 1. The *answer set construction* part, which is seemed indexing part in traditional IR system, employees both manual and automatic methods to build the knowledge base. The *answer set search* part processes a natural language query, extracts concepts and attributes, and maps them to the knowledge base so that the answer documents associated with the <concept, attribute> pairs can be retrieve. In the *answer presentation* part, the search result is presented with highlighted paragraphs considered to contain the answer to the query.

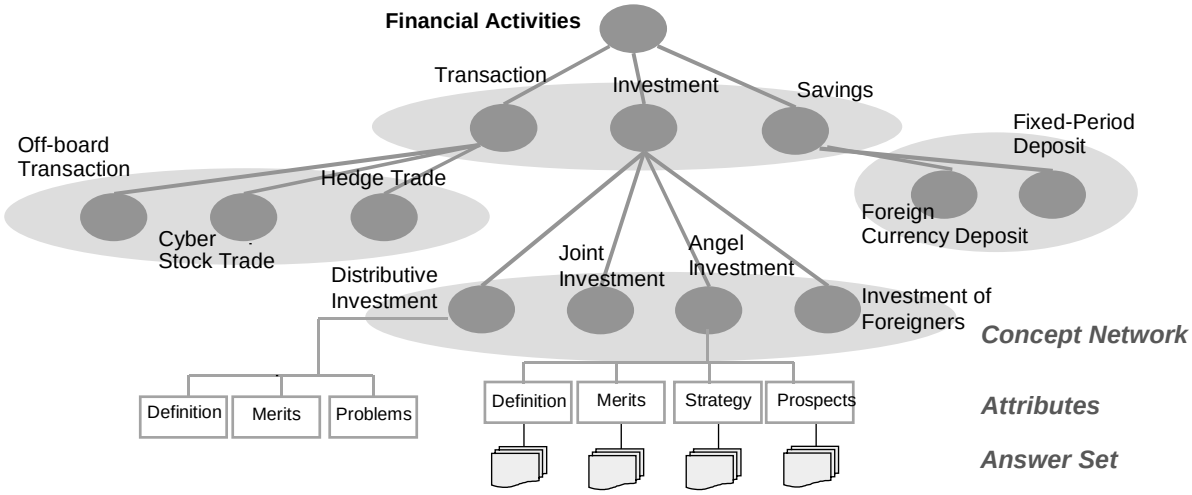

Figure 2. Concept, Attribute, and Documents
Group of concepts in concept network hierarchy and Distribution of attributes for concepts

Table 1. Distribution of attributes for concepts in an equivalence class(α-relation)

concepts \ attributes	Definition	Policy	Pay	Consultation case	Problems	Kinds	Purpose	Current situation	Regulations	Merits	Calculation methods	Negotiation	# attributes
Incentives	O	O	O	O	O	O	O	O	O	O			**10**
Hourly wage	O		O	O					O	O	O		**6**
Basic salary		O	O						O			O	**4**
Service allowance	O		O		O	O	O	O	O	O			**8**
Ability allowance	O	O	O		O			O	O	O			**7**
Bonus		O	O	O	O			O	O		O	O	**8**

2.2. Knowledge Base

Our knowledge base consists of three parts: *a concept network, attributes* associated with each concept, and *answer set* belonging to each <concept, attribute> pair.

The *concept network* contains about 50,000 conceptual word[2] as in WordNet [11] with 6 lexico-semantic relations that are used to form a synset hierarchy. By using the concept network, as already well-know in the WordNet-related research [1, 4, 6], a semantic processing of questions becomes possible. The information mined from concept network guides process of bridging inference between the query and the expected answer. Finding a place, i.e. concept node, in the network for a query can be construed as understanding the meaning of the query. *Attribute* set an intermediary as connecting concept network with answer documents. The answer-set driven retrieval system maps a user query into one or more concepts and further down to one or more attributes where associated documents can be picked up as the *answer set.* Attributes play the role in subcategorizing the documents belonging to the concept node. A set of attributes chosen for a particular concept specifies various aspects often mentioned in the documents bearing the concept and serves as an intermediary between a concept and high-precision answer documents. It should be noted that attributes are not inherently associated with a concept, but found in the documents addressing the concept. For instance, as in Figure 2, the concept node for “angel investment” would have attributes, “definition”, “strategy”, “prospects” and “merits”, that are aspects or characteristics of “angel investment” often mentioned in relevant documents.

Figure 2 and table 1 represent groups of concepts from different levels of the concept network and the distribution of attributes, showing that some

[2] Include 14,700 conceptual word in economy domain

attributes are shared by some concepts while others are unique to a concept. The fact that some attributes are shared by all or most of the concepts belonging to a higher level concept allow us to assume that related concepts share the same set of attributes. Another assumption we employ is that because of the observation that attributes tend to be found in the neighborhoods of some concept, the training data for a particular attribute under a given concept can be used for the same attribute under another concept.

With these assumptions, we devised a method of minimizing the training data construction efforts required for attribute-based classification, which is essential to select documents to be associated with <concept, attribute> pairs. In order to re-use the training data constructed for a particular <concept, attribute> pair, we define *α-relation* between two concepts. Two concepts are said to have an α-relation when the sets of associated attributes are sufficiently similar to each other. A later section describes how this relation is used for the knowledge base construction process.

3. Answer Set Construction

Attributes are defined to exist for concepts corresponding to categories in subject-based classification. Documents classified to a concept are considered to possess one or more attributes that reveal some characteristics or aspects of a document. Considering attributes as a different type of categories, we can define an attribute-based classifier for documents. While we employ the same learning-based and rule-based classification techniques for attribute-based classification, we underscore the way it is used for automatic knowledge-base construction, together with traditional subject-based classification. It works on reducing human efforts dramatically to knowledge base construction.

Our attribute-based classification method [10], at least as it is now, is no different from traditional text classification methods in that it uses training documents. However, the task of knowledge base construction for the answer set based retrieval system calls for unusual requirements. Whereas categories are pretty much fixed in traditional classification systems, the number of attributes (i.e. categories) for a given concept may change in our context. Another difference comes from the fact that it is not easy to have a sufficient number of training documents for each category since there are so many <concept, attribute> pairs that correspond to categories.

In order to address the issues mentioned above, we decided to add two additional steps to the ordinary statistical classification:

- use of pattern rules in conjunction with the usual learning-based classification
- selective use of words that may not be specific for attributes
- use of α-relation

While the first and second were chosen to improve precision of the classifier, the third was devised to widen the coverage of <concept, attribute> pairs for which training documents are provided. In other words, the use of α-relation allow us to re-use a classifier learned from a set of training documents belonging to a concept for the same attribute class under a different concept.

To improve upon accuracy of our attribute-based classifier, we have employed both rule-based and learning-based approaches. Unlink the case of subject-based classification, attribute class boundaries are sometimes hard to detect if only words are used as features. As such, we decided to use patterns of word sequences, not just single words. We have defined rules from train documents, which express the characteristics of a given attribute class. Rules may include single words, phrases, sentences, or even paragraphs. The pattern rules[3] are used to complement the errors made by the machine learning method, and further it is reused in query processing. We take the approach of a hybrid system combining rule-based classification and learning-based classification, with different weights are applied to different attributes. Besides we detect that some terms are not good at discriminate attribute since they are too specific to concept, whereas these terms are helpful to classify in concept. Therefore we eliminate the terms that concept-dependent word which frequently appearing in a certain concept area.

Another challenging problem in an operational setting is to define useful attributes to each of the concept nodes and collect training documents for each attribute under a concept node. It would be too expensive and time-consuming to collect a sufficient number of training documents for all the classes represented by <concept, attribute> pairs. Currently the number of classes is more than 250,000[4]. We need a method by which we can assign an attribute to a new document without separate training documents for that particular <concept, attribute> pair. Our approach to this problem is to use a special kind of relation, named *α-relation*, defined over the concepts

[3] Currently, we define 83 kind of attributes and 1,938 attribute pattern rule.

[4] More precisely, 14700*18, the number of concept nodes times the average number of attribute number of attributes under each concept.

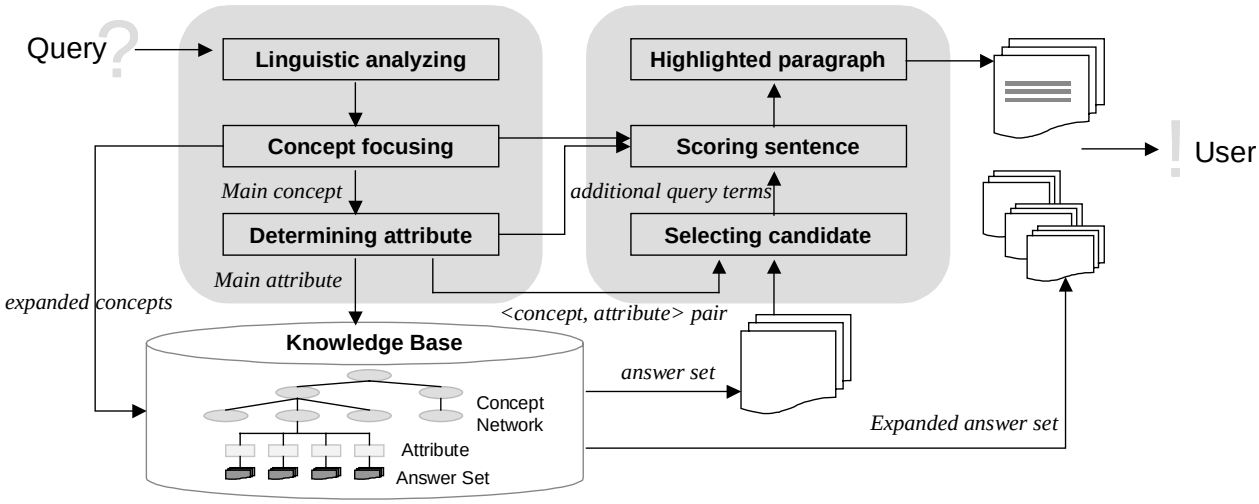

Figure 3. Retrieval Process

in the network. The main idea is to build a classifier for only one concept among the many belonging to an equivalence class based on the α-relation, and use it for other concepts. That is, we only need training documents for the representative class. Table 1 shows a distribution of attributes among the concept in *α-relation*. Once the attributes are identified for the first two concepts, "Incentives" and "Hourly wage", and training documents are selected for them, all the attributes except for the last one, "negotiation", can be considered having training documents. If we define attributes up to the third concept, "Basic salary", all the attributes are covered. The classifier learned from a single set of documents belonging to the concept would have the capability of classifying documents to the attribute classes belonging to other concepts if the concepts are all α-related.

We first construct a training document set[5] only for a single concept node representing all those nodes with the same attributes, using meta-search engine and document clustering. So we can build a classifier for those attributes that manifest themselves in the training set. If some documents fail to be assigned to an attribute, they are assigned to one of the remaining attributes. If a concept node other than the representative node in the equivalence class needs a. new attribute, we simply look for training documents for that attribute only. This kind of incremental process is based on our assumption that although attributes are associated with individual concept nodes, they share the common characteristics regardless of their parent concept nodes. Undoubtedly, however, this assumption does not always hold.

4. Retrieval Process

4.1. Answer Set Search

The main task of answer set search process is capturing a <concept, attribute> pair from natural language query, and mapping them to the knowledge base so that the answer documents can be retrieve. User query is represented as natural language so that imply semantic information need, not just single term. The query processing distinguishes between the main and additional terms from query. The former covey the essence of the query, reflected <concept, attribute> pairs. The latter help to convey the meaning of the query but can be omitted without changing the essence of the meaning. The secondary terms are useful clue for extracting answer sentence in answer document. Predefined patterns are also important for query processing. As noted earlier, we defined attribute pattern rules for improving accuracy of attribute-based classification. Then we rebuild these patterns as query-attribute pattern[6], expecting appeared pattern in interrogative form. Query processing consists of following part: 1) linguistic analyzing, 2) concept focusing, and 3) determining attribute, as illustrated in Figure 3.

Given a query, "*what is the problem of angel investment?"*, we analyze the sentence structure, such as conjunction structure and parallel phrases. We segment complex query into simple sentence. We distinguish the main terms in query by matching the longest term in concept network for focusing concept. To determine attribute of query we first plainly look

[5] It is only 3% of total amount of training set we needed.

[6] It was extended 2,170 query-attribute pattern.

for the term in attribute synset, which included title attribute representing all those synonym, i.e. “Problems” is title of set {Warning, danger, abuse, damage}. If not, we classify the question into one of 83 categories, each of mapped to a particular set of query-attribute pattern. Our example query map <angel investment, problem> pair.

After extracting appropriate <concept, attribute> pair, query expansion is generated, connecting with related concept in concept network. This expansion is based on the assessment of similarity between distances of concept network. The main advantage of connecting related concept is that the user can be traverse concept network through semantic path. Thus continuous search feedback can be possible. Expanded query map to knowledge base so that the documents corresponding the <concept, attribute> pairs can be retrieve as answer set. The results were ranked using attributed-based classification score in answer set construction processing.

4.2. Result Presentation: Highlighting Answer Sentence

Finding answer to a natural language question involves not only knowing what the answer is but also where the answer is. The answer set that produced initial searching step is considered to include the candidate answer sentence. For detecting answer sentence, we extract all the possible <concept, attribute> pairs each sentence in answer document. The sentence was not include query pairs was discard so that we can get candidate sentences where answer is appeared. Similarly query expansion, candidate sentences calculate score of match additional query words, which is generated in query processing. Highest scoring sentence was highlighted including its former and latter sentence. Right side box in Figure 3 shows our retrieval process.

5. Experiments

Whereas traditional Q/A and IR system have competition conference, like TREC, so that they can start with standard retrieval test collection, to explore how useful the proposed approach, we evaluate performance of answer document and candidate answer sentence. Another difference comes from the fact that result units for these systems are different. That is Q/A system returns exactly relevant answer (50 byte or 250 byte), while IR system returns document scored by ranking mechanism. Our system returns “answer set” distilled semantic knowledge as retrieval result

Table 2. Result of Answer Set Construction

Attribute sets	All attributes (no α-relations used)	Pre-selected attributes (with α-relation)
Precision	.5025	.6020
Recall	.4662	.6696
F-score	.4835	.6358(+31.4%)
Time	4	1

Table 3. Result of Answer Set Retrieval

	AS based IR		Web IR	
	Total	Top 5	Top 5	Top 10
Precision	0.584	0.769	0.291	0.2864
Recall	0.391	0.655	0.291	0.315
F-score	0.468	0.797	0.291	0.3
Highlighting MRR		0.78		

5.1. Automatic Answer Set Construction

Before evaluating our retrieval system, we were interested in knowing how effective and efficient the proposed knowledge base construction method. We tested the attribute-based classification for automatic construction method with 4,599 documents, 120 concepts, and 83 attributes. For performance comparisons, we used the standard precision, recall, and F-score [12]. Table 2 shows that the scores for using α-relation are higher than that of not using the relation. We gain a 31.4% increase in F-score and 400% in speed by using the knowledge. The potential advantage of using the α-relation is the ability to minimize the efforts not only required training set construction but also new answer set construction. On the other hand, a disadvantage is that it has a less chance to assign new attributes. The result our automatic answer set construction was to establish a ground work for further experiments.

5.2. Performance of Answer set retrieval

For our experimental evaluations we constructed operational system in Web, named “AnyQ”. Our AnyQ system currently consists of 14,700 concepts, 83 unique attributes, and more than 1.8 million web documents in the economy domain for Korean. The average number of document under each concept is 43.4, the average number of answer document is 25, and the average number of attribute is 18. To measure performance of retrieving answer set, we build 110 query-relevant answer set, judged by 4 assessor. Our

assessors team with 2 people. For performance comparisons, we used the P, R, F-score and MRR[5] for highlighted sentence. All retrieval runs are completely automatic, starting with queries, retrieve answer documents, and finally generating a ranked list of 5 candidate answer sentence.

We build traditional Web IR system on the same document set for baseline system. The Web IR system uses 2-poisson model for term indexing and vector space model for document retrieving. Table 3 summarized the effectiveness of our initial search step, answer set search. As expected, we obtained better results progressively as answer set based approach. The accuracy of Web IR become higher top10(0.31) to top5(0.291) when we determine more number of documents retrieved. By contrast, AS based IR has improvement both precision(0.769) and recall(0.655) when we assess less number of documents on top ranked. Even when all documents was considered(0.468) is higher than Web IR top 10(0.3). It comes from the fact that Web IR retrieves massive documents appeared term query. But AS based IR handled prepared answer set. That is, AS based IR tend to set highly relevant documents on top result. In other words, answer set based approach can be easier for user to find information they need with less effort.

To evaluate highlighted paragraphs, we generate a ranked list of 5 candidate answer sentences considered to contain the answer to the query. The score is 0.78 MRR. As mentioned before, our result is not the same type as TREC answer. But we can say that highlighted sentences are helpful to satisfy user information need.

We further realized that the query pattern as attribute was not sufficient for finding answer. Moreover, Korean has characteristic, various variation of same pattern, its duplicate over the attributes. It brings the fact that query processing has ambiguity. Another weakness of our system is that the accuracy of retrieval depends on knowledge base granularity. That is, the effectiveness of attribute-based classification influences whole process of our approach.

Unfortunately, Our experience cannot compare with other commercial system since there is no standard test collection. By the way AskJeeves was published their accuracy of retrieval is over 30~40%[7], however, this is not absolute contrast.

6. Conclusion

The accuracy of IR result continues to grow on importance as exponential growth of WWW, and it is therefore increasingly important that appropriate retrieval technologies be developed for the web. We have introduced a new type of IR, “Answer Set based IR”, attempts to provide high quality answer documents to user queries.

In the context of answer set-driven text retrieval, it is crucial to capture semantics and pragmatics of sentences in user queries and documents. In our case, we defined the semantic category of the answer as *attributes*, the documents associated with each attributes as *answer set*. We attempted to provide more accurate answers by attaching attributes to individual concepts in concept network. In order to construct knowledge bases, a certain level of quality is guaranteed, we developed a new method for attributed-based classifier(ABC) and built attribute pattern for improving accuracy of ABC and query processing both. In retrieval, we process a natural language query, extract concepts and attributes, and map them to the knowledge base so that the answer documents associated with the <concept, attribute> pairs can be retrieve.

Our proposed IR ranked highly relevant document on top result, thus it helps reducing human efforts dramatically to find answer. By established operational system, named “AnyQ”, our experiment showed realistic possibility of our approach systematically.

While our experiments were designed carefully, and comparisons made thoroughly, it has limitations. Our current work depends on the domain of the concept network. It is not clear how the proposed method can be extended to other domains. Our assumption, reflecting semantics in sentence to <concept, attribute> pairs, needs to be tested further. More fundamentally, we need a certain amount of manual work to initially construct the knowledge base such as the concept hierarchy and the initial training documents. We will have to see how the initial manual process influences the latter processes and what kind of performance degradation occurs when smaller efforts are used for the initial construction.

7. Reference

[1] Dan Moldovan, Sanda Harabagiu, et al, “LCC Tool for Question Answering”, *Proc of Text Retrieval Conference (TREC-11)*, November, 2002.

[2] Ask Jeeves[tm], http://www.jeevessolutions.com/technology/

[3] M. G. Jang, H. J. Oh, M. S. Chang, et al, “Semantic Based Information Retrieval”, *Korea*

Information Science Society review, 19(10):7-18, October 2001.

[4] Marius Pasca and Sanda M. Harabagiu, "The Informative Role of WordNet in Open-Domain Question Answering", *Proc of the NAACL 2001 Workshop on WordNet and Other Lexical Resource*, pp 138-143, CMU, Pittsburge PA, June 2001

[5] Ellen M. Voorhees, "Overview of the TREC 2000 Question Answering Track", *Proc of Text Retrieval Conference (TREC-11)*, November, 2002

[6] Eduard Hovy, Ulf Hermjakob, and Chin-Yew Lin, "The Use of External Knowledge in Factoid QA", *Proc of Text Retrieval Conference (TREC-10)*, November, 2001

[7] Cody C. T. Kwok, Oren Etzioni, and Daniel S. Weld, "Scaling Question Answering to the Web", , *Proc. of the 10th annual international ACM WWW10*, pp. 150-161, 2001

[8] Susan Dumais, Michele Banko, Eric Brill, Jimmy Lin, and Andrew Ng, "Web Question Answering: Is More Always Better?", *Proc. of the 25th annual international ACM SIGIR '2002*, pp. 291-298, Tampere, Finland, 2002

[9] Andrew McCallum, Kamal Nigam, et al., "A Machine Learning Approach to Building Domain-Specific Search Engines", *Proc. of the 16th IJCAI Conference*, pp 662-667, 1999

[10] Hyo-Jung Oh, Moon-Su Chang, Myung-Gil Jang, and Sung-Hyon Myaeng, "Integrating Attribute-Based Classification for Answer Set Construction", *Proc. of the 25th annual international ACM SIGIR '2002 2nd workshop on Operational Text Classification*, Tampere, Finland, 2002

[11] Christiane Fellbaum, "WordNet : An Electronic Lexical Database", *The MIT press*, 1998

[12] Ricardo Baeza-Yates, Berthier Ribeiro-Neto, "Modern Information Retrieval", Yiming Yang and Xin Liu, "A Re-examination Of Text Categorization Methods", pp. 73~98, Addition-Wesley Published, ACM press New York, 1999.