

Descriptive Question Answering in Encyclopedia

Hyo-Jung Oh, Chung-Hee Lee, Hyeon-Jin Kim, Myung-Gil Jang

Knowledge Mining Research Team

Electronics and Telecommunications Research Institute (ETRI)

Daejeon, Korea

{ohj, forever, jini, mgjang} @ etri.re.kr

Abstract

Recently there is a need for a QA system to answer not only factoid questions but also descriptive questions. Descriptive questions are questions which need answers that contain definitional information about the search term or describe some special events. We have proposed a new descriptive QA model and presented the result of a system which we have built to answer descriptive questions. We defined 10 Descriptive Answer Type(DAT)s as answer types for descriptive questions. We discussed how our proposed model was applied to the descriptive question with some experiments.

1 Introduction

Much of effort in Question Answering has focused on the ‘short answers’ or factoid questions, which answer questions for which the correct response is a single word or short phrase from the answer sentence. However, there are many questions which are better answer with a longer description or explanation in logs of web search engines(Voorhees, 2003). In this paper, we introduce a new descriptive QA model and present the result of a system which we have built to answer such questions.

Descriptive question are questions such as “Who is Columbus?”, “What is tsunami?”, or “Why is blood red?”, which need answer that contain the definitional information about the search term, explain some special phenomenon.(i.e. chemical reaction) or describe some particular events.

At the recent works, definitional QA, namely questions of the form “What is X?”, is a developing research area related with a subclass of descriptive questions. Especially in TREC-12 conference(Voorhees, 2003), they had produced 50 definitional questions in QA track for the competition. The systems in TREC-12(Blair et al, 2003; Katz et al, 2004) applied complicated technique which was integrated manually constructed definition patterns with statistical ranking component.

Some experiments(Cui et al, 2004) tried to use external resources such as WordNet and Web Dictionary associated with a syntactic pattern. Further recent work tried to use online knowledge bases on web. Domain-specific definitional QA systems in the same context of our works have been developed. Shiffman et al(2001) applied on biographical summaries for people with data-driven method.

In contrast to former research, we focus on the other descriptive question, such as “why,” “how,” and “what kind of”. We also present our descriptive QA model and its experimental results.

2 Descriptive QA

2.1 Descriptive Answer Type

Our QA system is a domain specific system for encyclopedia¹. One of the characteristics of encyclopedia is that it has many descriptive sentences. Because encyclopedia contains facts about many different subjects or about one particular subject explained for reference, there are

¹ Our QA system can answer both factoid questions and descriptive questions. In this paper, we present only sub system for descriptive QA

many sentences which present definition such as “X is Y.” On the other hand, some sentences describe process of some special event(i.e. the 1st World War) so that it forms particular sentence structures like news article which reveal reasons or motives of the event.

We defined Descriptive Answer Type (DAT) as answer types for descriptive questions with two points of view: *what kind of descriptive questions are in the use’s frequently asked questions?* and *what kind of descriptive answers can be patternized in the our corpus?* On the view of question, most of user’s frequently asked questions are not only factoid questions but also definitional questions. Furthermore, the result of analyzing the logs of our web site shows that there are many questions about ‘why’, ‘how’, and so on. On the other side, descriptive answer sentences in corpus show particular syntactic patterns such as appositive clauses, parallel clauses, and adverb clauses of cause and effect. In this paper, we defined 10 types of DAT to reflect these features of sentences in encyclopedia.

Table 1 shows example sentences with pattern for each DAT. For instance, “A tsunami is a large wave, often caused by an earthquake.” is an example for ‘Definition’ DAT with pattern of [X is Y]. It also can be an example for ‘Reason’ DAT because of matching pattern of [X is caused by Y].

Table 1: Descriptive Answer Type

DAT	Example/Pattern
DEFINITION	<i>A tsunami is a large wave, often caused by an earthquake.</i> [X is Y]
FUCTION	<i>Air bladder is an air-filled structure in many fishes that functions to maintain buoyancy or to aid in respiration.</i> [X that function to Y]
KIND	<i>The coins in States are 1 cent, 5 cents, 25 cents, and 100cents.</i> [X are Y ₁ , Y ₂ ,... and Y _n]
METHOD	<i>The method that prevents a cold is washing often your hand.</i> [The method that/of X is Y]
CHARCTER	<i>Sea horse, characteristically swimming in an upright position and having a prehensile tail.</i> [X is characteristically Y]
OBJECTIVE	<i>An automobile used for land transports.</i> [X used for Y]
REASON	<i>A tsunami is a large wave, often caused by an earthquake.</i> [X is caused by Y]
COMPONENT	<i>An automobile usually is composed of 4 wheels, an engine, and a steering wheel.</i> [X is composed of Y ₁ , Y ₂ ,... and Y _n]
PRINCIPLE	<i>Osmosis is the principle, transfer of a liquid solvent through a semipermeable membrane that does not allow dissolved solids to pass.</i> [X is the principle, Y]
ORIGIN	<i>The Achilles tendon is the name from the mythical Greek hero Achilles.</i> [X is the name from Y]

2.2 Descriptive Answer Indexing

Descriptive Answer indexing process consists of two parts: *pattern extraction* from pre-tagged corpus and *extraction of DIU*(Descriptive Indexing Unix) using a pattern matching technique.

Descriptive answer sentences generally have a particular syntactic structure. For instance, definitional sentences has patterns such as “X is Y,” “X is called Y,” and “X means Y.” In case of sentence which classifies something into sub-kinds, i.e. “Our coin are 50 won, 100 won and 500 won.” it forms parallel structure like “X are Y₁, Y₂,... and Y_n”.

To extract these descriptive patterns, we first build initial patterns. We constructed pre-tagged corpus with 10 DAT tags, then performed sentence alignment by the surface tag boundary. The tagged sentences are then processed through part-of-speech(POS) tagging in the first step. In this stage, we can get descriptive clue terms and structures, such as “X is caused by Y” for ‘Reason’, ‘X was made for Y” for ‘Function’, and so on.

In the second step, we used linguistic analysis including chunking and parsing to extend initial patterns automatically. Initial patterns are too rigid because we look up only surface of sentences in the first step. If some clue terms appear with long distance in a sentence, it can fail to be recognized as a pattern. To solve this problem, we added sentence structure patterns on each DAT patterns, such as appositive clause patterns for ‘Definition’, parallel clause patterns for ‘Kind’, and so on.

Finally, we generalized patterns to conduct flexible pattern matching. We need to group patterns to adapt to various variations of terms which appear in un-training sentences. Several similar patterns under the same DAT tag were integrated into regular-expression union which is to be formulated automata. For example, ‘Definition’ patterns are represented by [X<NP> be called/named/known as Y<NP>].

We defined DIU as indexing unit for descriptive answer candidate. In DIU indexing stage performed pattern matching, extracting DIU, and storing our storage. We built a pattern matching system based on Finite State Automata(FSA). After pattern matching, we need to filtering over-generated candidates because descriptive patterns are naive in a sense. In case of ‘Definition’, “X is Y” is matched so many times, that we restrict the

pattern when “X” and “Y” under the same meaning on our ETRI-LCN for Noun ontology². For example, “Customs duties are taxes that people pay for importing and exporting goods[X is Y]” are accepted because ‘custom duty’ is under the ‘tax’ node so they have same meaning.

DIU consists of Title, DAT tag, Value, V_title, Pattern_ID, Determin_word, and Clue_word. Title and Value means X and Y in result of pattern matching, respectively. Determin_word and Clue_word are used to restrict X and Y in the retrieval stage, respectively. V_title is distinguished from Title by whether X is an entry in the encyclopedia or not. Figure 1 illustrated result of extracting DIU.

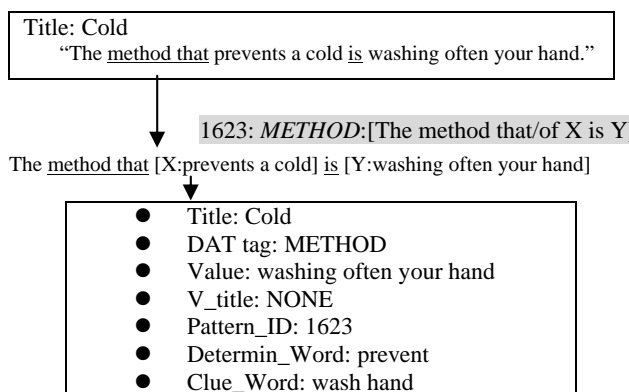


Figure 1: Result of DIU extracting

2.3 Descriptive Answer Retrieval

Descriptive answer retrieval performs finding DIU candidates which are appropriate to user questions through query processing. The important role of query processing is to catch out <QTitle, DAT> pair in the user question. QTitle means the key search word in a question. We used LSP pattern³ for question analysis. Another function of query processing is to extract Determin_word or Clue_Terms in question in terms of determining what user questioned. Figure 2 illustrates the result of QDIU(Question DIU).

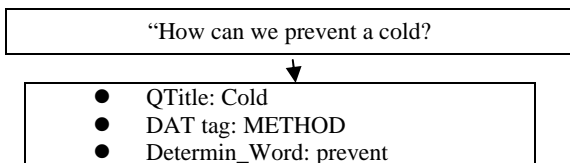


Figure 2: Result of Question Analysis

² LCN: Lexical Concept Network. ETRI-LCN for Noun consists of 120,000 nouns and 224,000 named entities.

³ LSP pattern: Lexico-Syntactic Pattern. We built 774 LSP patterns.

3 Experiments

3.1 Evaluation of DIU Indexing

To extract descriptive patterns, we built 1,853 pre-tagged sentences within 2,000 entries. About 40%(760 sentences) of all are tagged with ‘Definition, while only 9 sentences were assigned to ‘Principle’. Table 2 shows the result of extracted descriptive patterns using tagged corpus. 408 patterns are generated for ‘Definition’ from 760 tagged sentences, while 938 patterns for ‘Function’ from 352 examples. That means the sentences of describing something’s function formed very diverse expressions.

Table 2: Result of Descriptive Pattern Extraction

DAT	# of Patterns	DAT	# of Patterns
DEFINITION	408(22)	OBJECTIVE	166(22)
FUNCTION	938(26)	REASON	38(15)
KIND	617(71)	COMPONENT	122(19)
METHOD	104(29)	PRINCIPLE	3(3)
CHARCTER	367(20)	ORIGIN	491(52)
		Total	3,254(279)

* The figure in () means # of groups of patterns

Table 3: Result of DIU Indexing

DAT	# of DIUs	DAT	# of DIUs
DEFINITION	164,327(55%)	OBJECTIVE	9,381(3%)
FUNCTION	25,105(8%)	REASON	17,647(6%)
KIND	45,801(15%)	COMPONENT	12,123(4%)
METHOD	4,903(2%)	PRINCIPLE	64(0%)
CHARCTER	10,397(3%)	ORIGIN	10,504(3%)
		Total	300,252

Table 3 shows the result of DIU indexing. We extracted 300,252 DIUs from the whole encyclopedia⁴ using our Descriptive Answer Indexing process. As expected, most DIUs(about 55%, 164,327 DIUs) are ‘Definition’. We assumed that the entries belonging to the ‘History’ category have many sentences about ‘Reason’ because history usually describes some events. However, we obtained only 25,110 DIUs(8%) of ‘Reason’ because patterns of ‘Reason’ have lack of expressing syntactic structure of adverb clauses of cause and effect. ‘Principle’ also has same problem of lack of patterns so we only 64 DIUs.

3.2 Evaluation of DIU Retrieval

To evaluate our descriptive question answering method, we used 152 descriptive questions from our ETRI QA Test Set 2.0⁵, judged by 4 assessors.

⁴ Our encyclopedia consists of 163,535 entries and 13 main categories in Korean.

⁵ ETRI QA Test Set 2.0 consists of 1,047 <question, answer> pairs including both factoid and descriptive questions for all categories in encyclopedia

For performance comparisons, we used Top 1 and Top 5 precision, recall and F-score. Top 5 precision is a measure to consider whether there is a correct answer in top 5 ranking or not. Top 1 measured only one best ranked answer.

For our experimental evaluations we constructed an operational system in the Web, named “AnyQuestion 2.0.” To demonstrate how effectively our model works, we compared to a sentence retrieval system. Our sentence retrieval system used vector space model for query retrieval and 2-poisson model for keyword weighting.

Table 4 shows that the scores using our proposed method are higher than that of traditional sentence retrieval system. As expected, we obtained better result(0.608) than sentence retrieval system(0.508). We gain 79.3% (0.290 to 0.520) increase on Top1 than sentence retrieval and 19.6%(0.508 to 0.608) on Top5. The fact that the accuracy on Top1 has dramatically increased is remarkable, in that question answering wants exactly only one relevant answer.

Whereas even the recall of sentence retrieval system(0.507) is higher than descriptive QA result(0.500) on Top5, the F-score(0.508) is lower than that(0.608). It comes from the fact that sentence retrieval system tends to produce more number of candidates retrieved. While sentence retrieval system retrieved 151 candidates, our descriptive QA method retrieved 98 DIUs under the same condition that the number of corrected answers of sentence retrieval is 77 and ours is 76.

Table 4: Result of Descriptive QA

	Sentence Retrieval		Descriptive QA	
	Top 1	Top 5	Top 1	Top 5
Retrieved	151	151	98	98
Corrected	44	77	65	76
Precision	0.291	0.510	0.663	0.776
Recall	0.289	0.507	0.428	0.500
F-score	0.290	0.508	0.520 (+79.3%)	0.608 (+19.6%)

We further realized that our system has a few weak points. Our system is poor for inverted retrieval which should answer to the quiz style questions, such as “What is a large wave, often caused by an earthquake?” Moreover, our system depends on initial patterns. For the details, ‘Principle’ has few initial patterns, so that it has few descriptive patterns. This problem has influence on retrieval results, too.

4 Conclusion

We have proposed a new descriptive QA model and presented the result of a system which we have built to answer descriptive questions. To reflect characteristics of descriptive sentences in encyclopedia, we defined 10 types of DAT as answer types for descriptive questions. We explained how our system constructed descriptive patterns and how these patterns are worked on our indexing process. Finally we presented how descriptive answer retrieval performed and retrieved DIU candidates. We have shown that our proposed model outperformed the traditional sentence retrieval system with some experiments. We obtained F-score of 0.520 on Top1 and 0.680 on Top5. It showed better results when compared with sentence retrieval system on both Top1 and Top5.

Our Further works will concentrate on reducing human efforts for building descriptive patterns. To achieve automatic pattern generation, we will try to apply machine learning technique like the boosting algorithm. More urgently, we have to build an inverted retrieval method. Finally, we will compare with other systems which participated in TREC by translating definitional questions of TREC in Korean.

References

- S. Blair-Goldensohn, K. R. McKeown, and A. H. Schlaikjer. 2003. *A Hybrid Approach for QA Track Definitional Questions*, Proceedings of the twelve Text REtrieval Conference(TREC-12), pp. 336-342.
- H. Cui, M-Y. Kan, T-S. Chua, and J. Xian. 2004. *A Comparative Study on Sentence Retrieval for Definitional Question Answering*, Proceedings of SIGIR 2004 workshop on Information Retrieval 4 Question Answering(IR4QA).
- B. Katz, M. Bilotti, S. Felshin, et. al. 2004. *Answering Multiple Questions on a Topic from Heterogeneous Resources*, Proceedings of the thirteenth Text REtrieval Conference(TREC-13).
- B. Shiffman, I. Mani, and K.Concepcion. 2001. *Producing Biographical Summaries: Combining Linguistic Resources and Corpus Statistics*, Proceedings of the European Association for Computational Linguistics (ACL-EACL 01).
- Ellen M. Voorhees. 2003. *Overview of TREC 2003 Question Answering Track*, Proceedings of the twelfth Text REtrieval Conference(TREC-12).