International Journal of Digital Information and Wireless Communications (IJDIWC) 6(4): 292-298
The Society of Digital Information and Wireless Communications, 2016 ISSN: 2225-658X (Online); ISSN 2412-6551 (Print)

A Question-Answering Inferencing System Based on Definition and Acquisition of Knowledge in Written English Text

Kenta Hiratsuka and Hiroki Imamura
Graduate school of Engineering, Soka University
1-236 Tangi-machi, Hachioji-shi, Tokyo, Japan
e15m5221@soka-u.jp, imamura@soka.ac.jp

ABSTRACT

In recent years, question-answering systems have become a major area of research and development in information search technologies. We have proposed a new question answering system that can produce new knowledge by the logical inferencing. The proposed system is based on two characteristics that are logical inferencing using the knowledge and dealing with diversity in natural language. To realize this system, in this paper, we construct a question-answering processing section that can perform logical inferencing in the proposed system. This system performs inferencing based on "knowledge sentences" and "definition sentences". It can perform a more complicated logical inference by using definition sentences and performing the inference repeatedly. In this way, new knowledge is not only acquired by humans, but may be produced by the logical combination of knowledge. In the future, this system has the potential to become an important aid to human intellectual activity.

KEYWORDS

Nature Language Processing, Question Answering System, Logical Inference, Knowledge Sentence, Definition Sentence

1 INTRODUCTION

In recent years, question-answering systems have become a major area of research and development in information search technologies[1]. For example, question answering systems have been developed such as "Power-Answer", which can perform question answering based on logic[2] or "WATSON" that can deal with the diversity of natural language[3]. These conventional systems were intended to answer the question of the user by using the information extracted from documents and can generate highly accurate answers. However, these systems require that information to become the direct answer is listed on documents.

In contrast, we have proposed a new type question answering system that can produce new knowledge by logical inferencing in natural language[4]. Our system extracts not only knowledge but also definition which is a rule of the inference from web documents and can generate answers about the knowledge that is not listed in web documents by logical inference using the definition. Therefore, this system has the potential to become not only as simple question-answering system but also an important aid to human intellectual activity.

As shown in figure 1, we have been developing the question-answering system having two characteristics of logical inference using knowledge and dealing with the diversity of natural language. The proposed system normalizes all input sentences having the same meaning to just one sentence. It can deal with the diversity of the natural language by using normalized sentences as knowledge. In addition, this system can answer questions even if it does not have direct knowledge by performing logical inference using knowledge sentences.

In this paper, we construct a question answering processing section among other sections of the proposed system (within the dotted frame in figure 1). In this process, this system performs inferences based on "knowledge sentences" and "definition sentences". These two kinds of sentences are acquired from text on the internet or inputted by the users. "Definition sentences" are sentences in the form of “if..., then...” and these sentences express rules of inference that combine knowledge to produce different knowledge. For example, users may input definition sentences such as "If X is a father of Y and Y is a father of Z, then X is a grandfather of Z" and knowledge such as "Michael is a father of Daniel" and "Daniel is a father of David". Users can also input a question such as “Who is a grandfather of David?”. In this case, even if this system does not have direct knowledge, the system can answer such as "Michael is a grandfather of David" by the inference based on the definition sentence. In addition, this system can perform a more complicated logical inference by using definition sentences, and performing the inference repeatedly. Thereby, new knowledge is not only acquired by humans, but may be produced by logical inference.

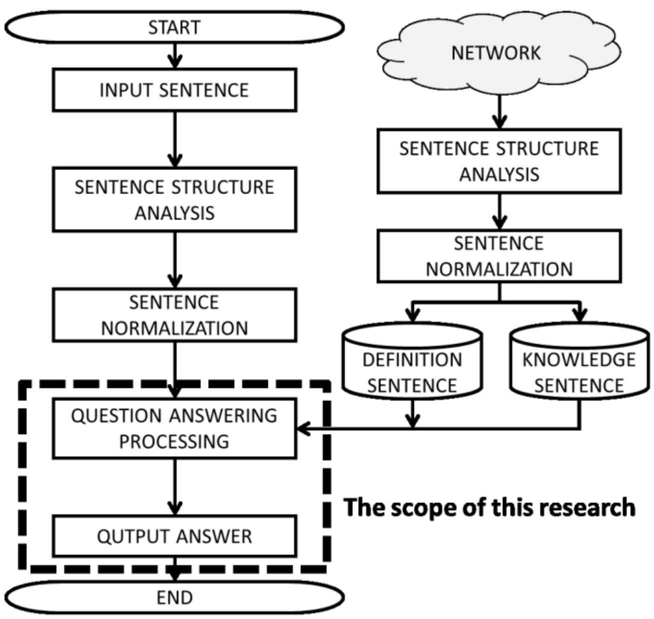

Figure1. Concepts of the proposed system

2 PROPOSED SYSTEM

Figure 2 shows a flowchart of the system from the dotted frame in figure 1. Here users manually input the knowledge and definition sentences for the system. At first, this system performs morphological analysis of an input sentence to provide parts of speech information of each word. This system classifies the analyzed sentences as any one of knowledge sentences, definition sentences or question sentences. When this system classifies input sentences as knowledge sentences or definition sentences, this system saves it into database as text data. In contrast, when this system classifies the input sentences as question sentences, this system performs the question answering processing. It matches knowledge data, inferences and multiplex inferences sequentially and outputs all the answers that are produced for a question. If this system does not produce answers, it outputs information necessary to allow successful inferences.

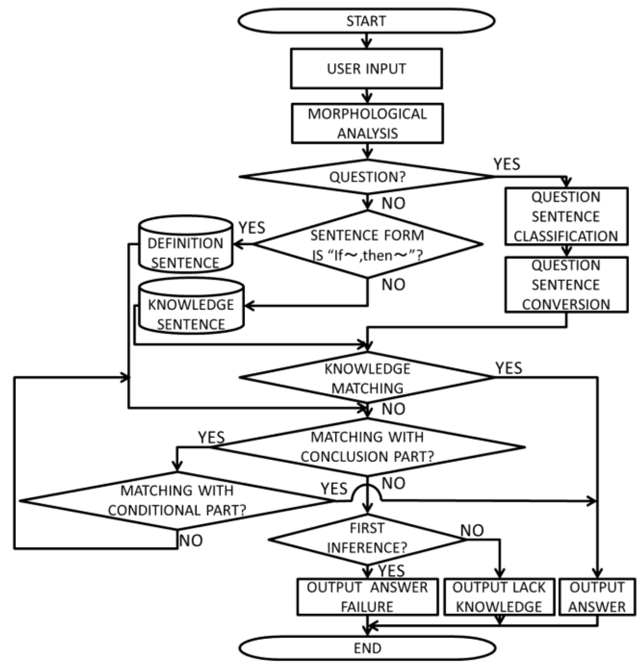

Figure2. The system flowchart

2.1 Morphological Analysis

This system performs morphological analysis of an input sentence using “Stanford CoreNLP” [1] which is a morphological analyzer for English language. This system classifies input sentences by the part of speech of the head word of a sentence. When the head word of an input sentence is a "be verb", "general verb", "auxiliary verb" or "interrogative", it is classified as a question sentence. For example, sentences such as "Does Saburo like coffee?" are classified as question sentences. The other input sentences are classified as knowledge sentences or definition sentences. For example, sentences such as "Taro is a man" are classified as knowledge sentences. In

[1] http://nlp.stanford.edu/software/corenlp.shtml

contrast, sentences are classified as definition sentences if they match the "if... then..." structure. The sentences such as "If Taro is a man and Taro has a son then Taro is a father" are classified as definition sentences. Thus, this system classifies input sentences into three types, "knowledge sentences", "definition sentences" and "question sentences". We regard "knowledge sentence" and "definition sentence" as declarative sentence.

When a declarative sentence is inputted, this system classifies the sentence as either a knowledge or a definition sentences and saves it as text data. If a similar sentence has already been inputted into the database, it is not saved. In addition, in the case of knowledge sentences, old knowledge is deleted when contradictory knowledge is inputted; that is the old knowledge is overwritten by the new inputted knowledge.

Based on information given by morphological analysis when a question sentence is inputted, this system performs question answering processing. Based on the part of speech information of the input sentence, question sentences are classified into 13 patterns according to English rules of grammar as follows.

Type1 sentence: be verbs

Type2 sentence: general verbs

Type3 sentence: auxiliary verbs

Type4 sentence: interrogative as the subject

Type5 sentence: which + be verbs

Type6 sentence: which + general verbs

Type7 sentence: which + auxiliary verbs

Type8 sentence: why + be verbs

Type9 sentence: why + general verbs

Type10 sentence: why + auxiliary verbs

Type11 sentence: interrogative + be verbs

Type12 sentence: interrogative + general verbs

Type13 sentence: interrogative + auxiliary verbs

For example, "Does Saburo like coffee?" is classified as type2 when this sentence is inputted. Because "Does", the head word of this sentence, is classified as a general verb. The classified sentence is converted into the form of the declarative sentence to perform the matching with knowledge sentences. For examples, the previously described example sentence is converted into "Saburo likes coffee".

2.2 Matching with Knowledge Data

This system matches a converted sentence from the question sentence and saved knowledge data. The matching starts from head words of each sentence, and the matching is shifted to next word one by one. When this matching arrives at the last of each sentence, the matching becomes the success. Conversely, if this system fails in matching in the middle of the process, it cannot output an answer. If matching is successful, it outputs the answer to the question.

2.3 Omission of the Modifier

When this system matches the knowledge sentences including modifier and the question sentences not including modifier, it can perform the matching that against these omitted a modifier. Figure 3 shows the rule of the modifier omission using an example sentence.

For example, we assume that this system is inputted a question "Is Mary a student?" when the system saves a knowledge "Mary is a serious student" such as (a) of figure 3. In this case, this system cannot output correct answer because it failed in matching "student" and "serious". To deal with this problem, when a word of the knowledge sentence side is the modifiers this system matches with a next word once again. This system can output the answer to the question if this system succeeds in the matching.

In contrast, in the case of a knowledge expressed as a negative sentence such as (b) of figure 3, this system does not omit the modifier in the above mentioned. Because the sentences that omitted a modifier are not necessarily to be true in the case of negative sentence. For example, in the case of (b) of figure 2, this system cannot determine whether Mary is not a student or not. Thus, this system does not omit the modifier about the knowledge expressed as a negation sentence.

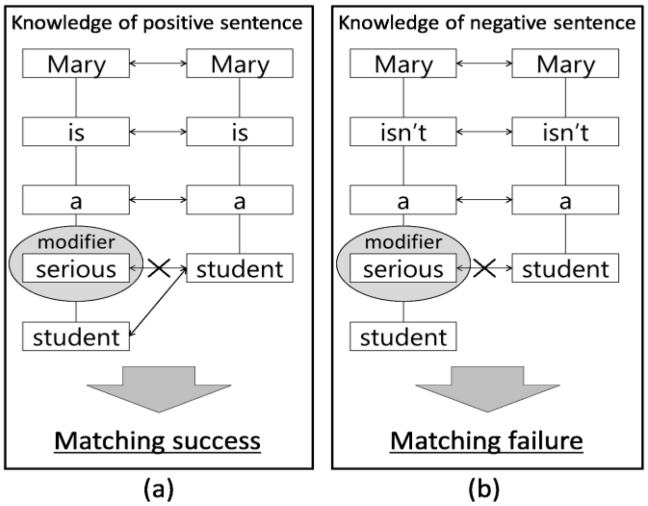

Figure3. The matching rule in the modifier omission

2.4 Inference

This system performs the inference using the definition sentences. At first, this system needs to find the definition sentence that can arrive at the appropriate answer in conclusion. Therefore, it matches a question sentence with the conclusion part of the definition sentence, because a conclusion produced from the definition sentence must be the answer for the question. A conclusion part is the sentence after “then" in definition sentences. If an appropriate definition sentence is not found, it finishes processing because the answering to the question is impossible. If an appropriate definition sentence is found, it performs the matching knowledge sentences with a precondition part of the definition sentence. A precondition part is the sentence before “then" in definition sentences. It outputs the answer to the question if matching of precondition parts of sentences surround “and" or “or" are successful. If the matching is not successful, it outputs the necessary knowledge to succeed in the inference.

If Noah has feathers then Noah is a bird. (1)

Noah has feathers (2)

Is Noah a bird? (3)

For example, we assume that this system has one definition sentence such as (1) and one knowledge sentence such as (2). In this time, if this system is asked a question such as (3), this system performs inference by using above definition sentence. Next, this system performs the matching with conditional parts and knowledge sentences. In this case, this system matches the conditional part of (1) such as “Noah has feathers" with a knowledge sentence such as (2). As a result, this system outputs the answer that “Yes.".
When wild cards (the part expressed in capital letters of the alphabet such as "X" and "Y" shown in (4) are included in the definition sentence, this system substitutes applicable nouns for wild cards.

If X is father of Y and Y is father of Z
then X is grandfather of Z. (4)

Bob is father of Daniel. (5)

Daniel is father of John. (6)

Who is grandfather of John? (7)

For example, we assume that this system has one definition sentence such as (4) and two knowledge sentences such as (5) and (6). In this time, if this system is asked a question such as (7), this system performs inference by using above definition sentence. And, the word "John" is substituted for Z by matching with conclusion part of above definition sentence and question sentences. Next, this system performs the matching with conditional parts and knowledge sentences, and this system substitutes words that "Bob" and "Daniel" for X and Y each. As a result, this system outputs the answer that ``Bob is grandfather of John.".

2.5 Multiplex Inference

Figure 4 shows a flow of the multiplex inference using an example sentence. This system performs a more complicated inference by following up the associated definition sentence.
For example, we assume that this system was inputted the question sentence when this system is inputted the sentences that are shown in figure 4.

At first, this system performs the matching with knowledge sentences. In this case, the matching is not successful, therefore this system matches a question sentence with the conclusion part of the definition sentence. Next, this system succeeds in the matching with conclusion part of a definition sentence that “If B is a father of D and C is a mother of D, then D is a son of B”. After that, this system matches knowledge and the precondition parts. If this system failed in the matching, it matches the precondition part of a definition sentence and the conclusion part of other definition sentences again. In this figure, this system has some knowledge sentences that ``Taro is a man.", “Jiro is a man.”, “Mary is a woman.” and “Jiro is twelve years old.”. However, these knowledge sentences cannot match each conditional part. Accordingly, this system performs the matching with conditional parts of first definition sentence and conclusion parts of other definition sentences. The second sentence and the third sentence from the top succeed in matching with each condition part of first definition sentence that "B is a father of D" and "C is a mother of D". After that, this system succeeds in the matching each conditional part of the second and third sentences to four saved knowledge sentences. Thus, this system can performs the inference repeatedly until it succeeds in the producing the answer or cannot perform the inference.

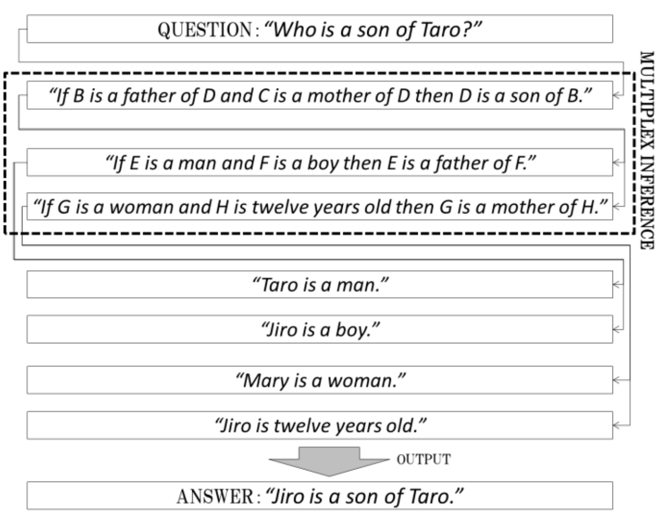

Figure4. An example of Multiplex Inference

2.6 Output Answer

When this system succeeds in obtaining answers, this system outputs the answer and the process of logical inference. The reason why this system output a process of logical inference is to know the inference process that arriving at an answer or new knowledge. In addition, if new knowledge is produced, users can know the process of logical inference of this system. When this system does not acquire answers, this system outputs information necessary to let the inference succeed. If there are multiple answers to questions, this system outputs all answers and process of inferences.

3 EVALUATION EXPERIMENTS

To evaluate the effectiveness of this system, we experimented to evaluate whether this system can answer to questions including following items.

Item1: The system can answer to the question about knowledge sentences

Item2 The system can answer to the question needs to omit a modifier

Item3 The system can answer to the question about definition sentences

Item4 The system can answer to the question about definition sentences including wild cards

Item5 The system can answer to the question about multiplex inference

Item6 The system can answer to the question about multiplex inference including wild cards

In this experiment, we had the thirteen subjects asking two ways of questions of Yes/No and 5W1H about each item. Accordingly, this system was asked twelve questions by each subject.
The subjects evaluated whether the answer to each questions of this system is correct or not. In advance, knowledge and definition sentences

necessary for question answering processing were inputted by subjects.

3.1 Results

Table 1 shows a correct answer rate in each question. On average, we could see that this system was able to answer correctly more than 80% of all questions in each experimental item.

Table1. Experimental results

Experiment Items	Correct / Question(rates)
Item1	24/26 (0.923)
Item2	21/26 (0.808)
Item3	25/26 (0.962)
Item4	22/26 (0.846)
Item5	23/26 (0.885)
Item6	19/26 (0.732)

3.2 Discussions

There were two reasons why this system could not answer correctly about some questions. The first reason is that a morphological analyzer failed in analysis of some sentences and thereby a question sentence was not given right part of speech information. For example, this system was asked question "Is Mike a kind teacher?". The morphological analysis result of "kind" is output with "noun". However, the analysis result of "kind" should become "adjective". Therefore, this system becomes not able to appropriate matching.
Another reason is that this system was not able to make a right conversion for a question sentence because of not having performed semantic analysis. Figure 5 shows an example of the false conversion that is caused by not doing semantic analysis.
When this system was asked question “Is the policeman Manabu?", this system needs to find a main clause to convert the question into an assertive sentence. If this system is able to find a main clause, this system can convert the question sentence into the sentence of right form such as the left side of figure 5. However, this system cannot find a right main clause because this system could not distinguish a certain semantic difference in the range of nouns “the policeman Manabu". Consequently, question sentence is converted into the sentence of wrong form such as the right side of figure 5. This system fails in matching of knowledge sentences and question sentences.
Accordingly, we consider that this system can answer the question given right parts of speech information. However, the content of the question is limited to the question that semantic analysis is unnecessary.

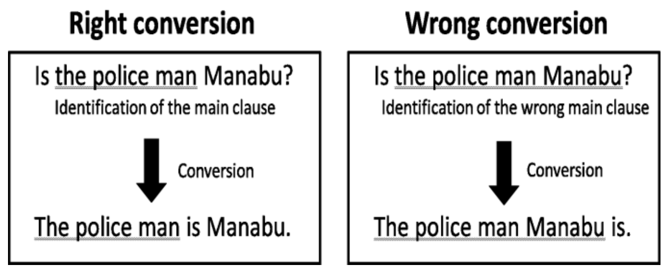

Figure5. Cause of wrong conversion

4 CONCLUSIONS

In this paper, we have worked on the construction for the question answering processing section that can perform the logical inference in the proposed system. We experimented to evaluate the effectiveness of this system.
From the experimental result, we could see that this system was able to answer the almost question sentences of the assumed form. In the future, we attempt for implementation of the process of semantic analysis and the process of discrimination of proper noun.

REFERENCES

1. Hirschman L and Gaizauskas R, “Nature Language Qusetion Answering: the View from Hereg”, volume 1, Cambridge University Press, 2001.
2. Moldovan, Dan I., Christine Clark, and Moldovan Bowden. "Lymba's PowerAnswer 4 in TREC 2007." TREC. Vol. 1. No. 5.3. 2007.
3. David Ferrucci, Eric Brown, Jennifer Chu-Carroll, James Fan, David Gondek, Aditya A. Kalyanpur, Adam Lally, J. William Murdock, Eric Nyberg, John Prager, Nico Schlaefer and Chris Welty, “Building Watson: An Overview of the DeepQA Project", AI Magazine, pp. 59-79, 2010.

4. Hiroki Imamura, "A Dialog System being able to Infer based on Definition and Acquisition of Knowledge", The 27th Annual Conference of the Japanese Society for Artificial Intelligence, 1K3-OS-17a-1, 2013.