

An Overview of Question and Answering Challenge (QAC) of the next NTCIR Workshop

Jun'ichi FUKUMOTO

Department of Computer Science, Ritsumeikan University
1-1-1 Noji-higashi, Kusatsu-shi, Shiga 525-8577, Japan
fukumoto@cs.ritsumei.ac.jp

Tsuneaki KATO

Language and Information Science, University of Tokyo
3-8-1 Komaba, Meguro-ku, Tokyo 153-8902 Japan
kato@boz.c.u-tokyo.ac.jp

Abstract

In this paper we will propose the question and answering task, called Question and Answering Challenge (QAC), and its first evaluation. This will be carried out as a sub-task of NTCIR3 Workshop scheduled in October 2002. In QAC, we aimed to encourage the development of practical Q&A systems in a general domain and focus on research of user interaction and information extraction. User interaction technology leads to actual interaction between computer and person. In actual Q&A between people, there will typically be several interactions in order to confirm the intention of the questions and so on. Information extraction that works in general domain is also an important technology in order to realize real Q&A system.

Keywords: Q&A, Information Extraction, user interaction.

1 Introduction

We propose the question and answering task, called Question and Answering Challenge (QAC), and its first evaluation QAC-1. This will be carried out as a sub-task of NTCIR3 Workshop scheduled in October 2002.

Question and answering is a task to obtain appropriate answers for given domain independent questions written in natural language from a large document collection. Question and answering has been considered as an interesting research topic for many years and there has been many results published up to now. Most of Q&A systems can get answers for given questions as a result of interaction between the system and users although these interactions are limited in some specific domains. In the current Q&A research, the goal of sys-

tems is to retrieve an answer to a given question from very large amount of information source in general domain.

Q&A technology is related to information extraction, information retrieval, natural language interaction and other NLP research. We will focus on user interaction among several research topics in QAC because it will lead actual interaction between computer and person in general domain. We are also interested in how to extract answer expressions from information source to a given question. Information extraction that works in general domain will be an important technology in order to get such answer expressions.

Question and answering is also related to text summarization technology. If a question is why or how type question, answer of the question may consist of several sentences or some sort of summary of the original text. Moreover, in evaluation of question and answering, system is required to provide support information that is important to know why answer is obtained or generated. In order to generate support information, summarization technology can be sometimes used.

2 Issues of QAC

We will describe several technical points to conduct evaluation of Q&A task. Currently we are planning to run QAC evaluation exercises over five years. In QAC we aim to encourage the development of practical Q&A systems in a general domain by providing an evaluation bases such as a question answering task set and evaluation criterion.

2.1 Goal of QAC evaluation

As we have already described in the previous section, our main goal of QAC is to encourage user interaction technology that works in a real domain. When user wants to know something from database, s/he makes a question to the database. If s/he can get insufficient information after the first question, s/he makes more questions according to the response of questions until the requiring answer is obtained. Moreover, in a real situation of user interaction user generally uses minimum expression of answer to given question. The user also uses a series of questions related to the first question, if they want to know more about the topic of the question.

In Q&A task, it is important to evaluate how correct and valid the answer is and how appropriate the support information for each answer is. Support information means some part of a text that will be evidence why the answer is correct. It will sometimes be a part of text that includes correct answer elements. It is also important to generate answers for no answer question and multiple answer questions. If there is no answer to a question, the system has to respond, for example, there is no answer.

In the Q&A task of TREC[2][3], systems are required to retrieve sentences or a part of text. However, in actual interaction between persons, several interactions such as question answering dialogues frequently occur. When one person can not understand what questioner intends in his/her question, the person make another question to get the questioner's intention to answer the question. There is a case that a person may formulate the second question according to the answer to the first question. Moreover, these answers by persons consist of simple words or sentences, which express contents of answer and do not include meaningless expressions.

2.2 How to define Q&A task

There have been already some evaluation of question and answering[4][5]. Several points of Q&A task definition have to be considered as follows:

1. number of question

How many questions are sufficient for evaluation of Q&A system?

2. no answer question

If there is no answer in database, how a system reply to user?

3. multiple answer question

How to evaluate multiple answers: all the answers should be correct or some of them should be correct

4. information level of answer

A system has to reply an answer in detail level or general level such as specific town name or city name.

5. an amount of answer information

Generated answer express is word, phrase, sentence, generated passage and so on.

6. question type

There are several question types such as 5WH type question and yes/no question.

2.3 Target information sources

What kinds of information sources are suitable for question and answering task? Newspaper articles are suitable for answering general question and are also good database because newspaper articles are widely used for NLP communities. Question and answering on WWW is very real situation and is very useful but is very difficult to evaluate generated answers at the same evaluation criterion. On the other hand, for evaluation using the same WWW environment, it is necessary to gather a number of copyright-free WWW pages.

2.4 Multilingualism

At the first stage of QAC evaluation, question and answering is conducted only in Japanese. However, Japanese Q&A for English text database or English Q&A for Japanese is very attractive. Use of other languages is also useful in actual simulations of WWW search.

3 Task definition of QAC-1

There are many evaluation points in Q&A evaluation. In the first evaluation of question and answering, QAC subtask of NTCIR 3 Workshop, we tentatively chose two kinds of evaluation points according to our research focus of QAC. In the first QAC evaluation, we will emphasize two kinds of technical points in the first QAC evaluation at NTCIR 3 Workshop, that is, information extraction and user interaction technologies.

1. IE technology

Information Extraction like MUC evaluation[6] aimed to extract information elements from documents using template. For the extraction of template elements, detailed definitions of Named Entity, Template Element, Template Relation, and Scenario Template are required, that is, IE technology is domain specific one. In Q&A, every element and contents can be extracted from documents as an answer of a given question. One

of our aims is to make IE technology available in general purpose. In NE, the extraction elements are person name, organization name, location name and so on[7], however, answering elements are not limited to these types.

2. user interaction

Our main purpose in Q&A research is to develop user interaction technology in real domain. At the first stage of QAC, we will propose a series of questions for Q&A evaluation. In these questions, it is necessary to resolve reference between questions and there will be ellipsis in the following questions, which frequently occurs in Japanese.

We will describe a tentative QAC task definition of the NTCIR 3 Workshop as follows:

- Question type

Questions are short answer questions. Answer will be a noun or noun phrase which indicates person names, organization names, name of various artifact, money, size, date and so on.

- no answer or multiple answer

There will be no answer questions or multiple answer questions. That is, there is a case that there is no answer object in documents to a given question or there are many answer objects. In this case, it is necessary to define how to describe these answers as system responses, for example, a list of top five answers or all the answers in some order.

- a series of questions

There will be one or more follow-up questions to the first question. When enough information is not obtained in the first question or more information related to the first question is needed, follow-up questions are generally given. For example, if the first question is a question of person name and the second question is a question of his/her age, In Japanese, there will be ellipsis in the follow-up question.

- support information

When a system give an answer, the system is required to provide evidential information why this answer is obtained from document collection. Such information is support information which is taken from original document and includes the system's answer.

- Target documents

We are now planning to use one year Japanese newspaper articles to get answers to given questions.

We are planning to prepare two tools to encourage beginners of this research area: NE tagger and IR system. It will also be helpful for NLP researchers to share dictionaries, extraction rules and some other resources from the use of these tools.

Finally, we describe the Current tentative schedule as follows:

| | |
|------------------------|----------------|
| Call for Participation | June, 2001 |
| dry run | winter of 2001 |
| formal run | spring of 2001 |
| NTCIR3 Workshop | October, 2002 |

4 Conclusions

We have described our aim and the current evaluation plan as above. Although the details of evaluation have not yet been determined, we will discuss task definition and evaluation criterion in the later meetings. We hope this QAC evaluation will attract a lot of research groups that are interested in question and answering research.

References

- [1] Burger, J., Cardie, C. et al. Issues, Tasks and Program Structures to Roadmap Research in Question & Answering (Q&A) NIST DUC Vision and Roadmap Documents <http://www-nlpir.nist.gov/projects/duc/roadmapping.html>, 2001.
- [2] Voorhees, E.M. and Harman, D.K. (eds.) Proceedings of the Eighth Text REtrieval Conference (TREC-8) http://trec.nist.gov/pubs/trec8/t8_proceedings.html, 2000.
- [3] QA Track Overview TREC9, <http://trec.nist.gov/presentations/TREC9/qa/index.htm>, 2000.
- [4] Sasaki, Y., Isozaki, H et al., An evaluation and Comparison of Japanese Question Answering Systems (In Japanese) Technical Report of IEICE, NLC 2000-24, 2000.
- [5] Voorhees, E.M. and Tice, D.M. Building a Question Answering Test Collection in Proceedings of SIGIR2000, pp. 200-207, 2000.
- [6] Proceedings of 7th Message Understanding Conference (MUC-7), DARPA, 1998.
- [7] Information Retrieval and Extraction Exercise (IREX) <http://cs.nyu.edu/cs/projects/proteus/irex/>, 1999.