Title	
Author(s)	,
Citation	
Issue Date	2005-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1911
Rights	
Description	Supervisor: , ,

Studies on an Interactive Question Answering System which can handle users' ambiguous questions

Hidenori Tokue (310073)

School of Information Science,
Japan Advanced Institute of Science and Technology

February 10, 2005

This paper describes a concept of an open-domain interactive question answering system. When an user inputs an ambiguous question, this system asks him/her back and selects an appropriate answer according to his/her response. Ambiguous questions in our studeis indicate ones that can't choose only one answer, because word meaning in question is ambiguous. For example, a question "worldcup no yuusyoukoku ha dokodesuka" is ambiguous, because there are some meaning in "worldcup" as soccor and rugby. So, system can't choose only one answer. When a user asks such a question, system asks "dono kyogi no worldcup desuka" to user, and proper answer is selected by user asking back. In our studies, we mainly describe the way to extract ambiguous part in user question. Outline of our proposal techique is described following. In the previous example, when system extracts two answers as "soccor" and "rugby", we pay attention to the expression which specializes keyword meaning(specializing expression). For example, the phrase "soccor no" or "rugby no" which modifies "worldcup" can be considerd to be expression that specialize meaning of worldcups . Thus, about keyword in question sentence, we extract specializing expression of the keyword, and it is considered that the keyword is ambiguous when a different specializing expression at each answer exists for the same keyword.

The flow of the process of the system is shown below. First, keywords, answer type and a type of keyword are extracted from user question and text including keywords is extracted. When answer candidates are extracted from those text, in our studies, noun that meets two requirement, morpheme information matches to the answer type, it is in the neighborhood of the keyword, is extracted as answer candidates. Thus, it sets priorities to answer candidates. The score is given by the follwoing equation.

$$S_{ans} = w_{pat} \times S_{pat} + w_{mor} \times S_{mor} + w_{dis} \times S_{dist}$$

S_{pat} is score of type of keyword, S_{mor} is score of morpheme infomation, S_{dist} is distance between each keyword and answer candidate. w_x are weights of each score.

Next, for detection of question ambiguous, the expression that modifies keyword is extracted from text extracted answer candidates. Pattern of specializing expression to extract is, noun modifier, case-marking particle of keyword as pattern of modifing keyword, "previous word" and "succeeding word" when a keyword is compound noun, and "date expression". For example, when keyword is "kinmedarisuto", "judo" is extracted from "judo no kinmedarisuto" as specializing expression of noun modifier. And, the group of specializing expression of each keyword and type of specializing expressions, and in this group, ambigugous is detected by reserching each answer have different specializing expression. But, in this group, inadequate specializing expressions are often included. So, in our studies, we introduce concept of attribute and try to remove such a inadequate specializing expressions. Attribute is feature of specializing expression. In our studies, we deal with 5 type, "number + suffix", "parentheses" , "semantic class of thesaurus" , "N characters at ending" and "date expression" as the attribute. For example, When there is the group of specializing expression"5 daime" "3 daime" "mito", "5daime" and "3daime" is extracted in attribute of "number + suffix" and "mito" is removed, because "mito" doesn't have the attribute "number + suffix". Thus, this group is considered new group, if each specializing expression have common attribute in initial specializing expression group and in this group, specializing expression that doesn't have attribute is removed. Next, several group such a specializing expression is extracted, so it sets priorities by scoring and we

choose proper group for response to a user. Score is decided by how times specializing expression is appeared in some answer candidates, frequency of specializing expression with attribute, whether same specializing expressions are extracted for different answers. Using 33 ambiguous question, we attempt experiment for detecting ambiguous above mentioned. As a result, the ratio of question judged to be ambiguous was the about 80%, and the ratio to which specializing expression was extracted by proper attribute was also about 80%.