ABSTRACT

SUCI ARMELIA SANUR. Passages Selection in Question Answering System for Indonesian Language Documents. Supervised by JULIO ADISANTOSO.

The first step on Question Answering System was the user enter question query. The used question query is limited to question type: WHO, WHERE, WHEN, and HOW MANY or HOW MUCH. The question word on query is used to obtain an answer candidate, while other words beside the question word are used to analyze the question. Question analysis process is started by parsing into keyword become tokens. The question sentence that has parsed is used to retrieve document and top passage. Top passage is obtained of question from passages that has highest point. Passages was done by three scoring method : rule-based, heuristic, and combination of rule-based with heuristic. The answer extraction is conducted by calculating the nearest distance between each answer candidate in top passage and each word in keyword. Answer correction is evaluated by using these criteria: right, unsupported, wrong, and null.

The evaluation of the research was seen on the set of question and document, also the accuracy for each answer. The result of rule-based scoring used 10 top documents was 77.5 % for criteria right, 2.5 % for criteria unsupported, 17.5 % for criteria wrong, and 2.5 % for criteria null. The result of heuristic scoring was 75 % for criteria right, 2.5 % for criteria unsupported, 20 % for criteria wrong, and 2.5 % for criteria null. The result of rule-based and heuristic scoring was 72.5 % for criteria right, 2.5 % for criteria unsupported, 22.5 % for criteria wrong, and 2.5 % for criteria null. The result of heuristic scoring used 2 top documents was 75 % for criteria right, 22.5 % for criteria wrong, and 2.5 % for criteria null. The result of rule-based scoring was 60 % for criteria right, 37.5 % for criteria wrong, and 2.5 % for criteria null. The result of rule-based and heuristic scoring was 75 % for criteria right, 22.5 % for criteria wrong, and 2.5 % for criteria null.

Keywords: question answering, rule-based scoring, heuristic scoring