

Abstract

A Question Answering System can be interpreted as an information retrieval system where an answer is expected to be obtained or responded directly based on the questions that are given correctly. One type of application contained in the Question Answering System is Reading Comprehension. Reading Comprehension can do a process in which from a text reading material, questions, and a set of answer choices provided as input data, the computer can determine which answer is correct. This system allows someone to be able to do the question and answer process with the aim to get a certain set of information. In this research, we propose a method on tackling the reading comprehension task based on Indonesian Open Information (Open IE) output representation. An Open IE system will extract important parts of the input sentence into tuple relations in the format (Argument, Relation, Argument). This Open IE output representation has been studied on English dataset and an English Open IE system. In this work, we explore the use of Indonesian Open IE system, applied to the translated dataset. As an evaluation, a baseline method consisting of a sliding window and distance based algorithm on lexical matching was implemented. The experimental result shows that on the raw text representation, baseline method has better performance, while after sentence rewriting was performed the Open IE representation outperformed the raw text representation.

Keywords: question answering system, reading comprehension, open information extraction, indonesian language, baseline