

Abstract– Despite the severe impact of COVID-19 on humans has already decreased, people still need to be aware of the recent disease information. A continually updated Frequently Asked Questions (FAQ) system could help the public get valid and relevant information. To maintain a FAQ system manually needs much effort, hence an approach to develop the system automatically is needed. Question Answering System (QAS) is a system that can accept input in question sentences and produces an answer quickly, concisely, and relevantly, and could be used to provide COVID-19 information to the public. One method on developing a QAS is Recognizing Question Entailment (RQE). RQE is a form of relationship based on a cause-and-effect relationship between two pieces of text called text (T) and hypothesis (H). We present a study on developing Covid-19 QAS in Bahasa Indonesia using RQE. The datasets are collected from reputable sources and consist of 725 pairs of questions and answers. The experimental results show that the best performance results were obtained using the Logistic Regression model in training set 1, which contains 54.2% of positive question pairs and 45.8% of negative question pairs with an f-measure value of 83.65%. These results indicate that the RQE method can identify the entailment between new questions and questions in the dataset well.

Keywords: Covid-19; Indonesia; Question Answering System; Question Entailment; Supervised Learning