Abstract

The lack of accessible information about herbal medicine poses challenges for users seeking reliable answers in this domain. This research proposes an ontology-based question answering system for herbal medicine products that integrates classification, entity recognition, and large language models (LLMs) to provide accurate and natural responses. The system classifies Indonesian-language questions into 12 categories using a Support Vector Machine (SVM) and extracts 6 key entity types using a Named Entity Recognition (NER) model. To improve classification performance, the study compares TF-IDF, FastText, and IndoBERT vectorization, as well as the use of text preprocessing. The classification results and keywords are used to build SPARQL queries to search for information from the ontology. This information then becomes context for a Large Language Model (LLM) to generate more natural and complete answers. This integration of SVM, NER, ontology, and LLM represents a novel approach in the domain of herbal medicine. The test results show that the best SVM accuracy was achieved using the IndoBERT model without text preprocessing, reaching 96 percent. The performance test results of the NER model varied for each entity. Furthermore, this question-answering system is implemented through a Telegram bot to improve user accessibility of information on herbal medicine products.

Keywords: question answering system (QAS), herbal medicine, ontology, SVM classification, Named Entity Recognition (NER), Large Language Model (LLM)
