Resource Creation Using Natural Language Processing Techniques for Malay Translated Qur'an

Authors : Nor Diana Ahmad, Eric Atwell, Brandon Bennett

Abstract : Text processing techniques for English have been developed for several decades. But for the Malay language, text processing methods are still far behind. Moreover, there are limited resources, tools for computational linguistic analysis available for the Malay language. Therefore, this research presents the use of natural language processing (NLP) in processing Malay translated Qur'an text. As the result, a new language resource for Malay translated Qur'an was created. This resource will help other researchers to build the necessary processing tools for the Malay language. This research also develops a simple question-answer prototype to demonstrate the use of the Malay Qur'an resource for text processing. This prototype has been developed using Python. The prototype pre-processes the Malay Qur'an and an input query using a stemming algorithm and then searches for occurrences of the query word stem. The result produced shows improved matching likelihood between user query and its answer. A POS-tagging algorithm has also been produced. The stemming and tagging algorithms can be used as tools for research related to other Malay texts and can be used to support applications such as information retrieval, question answering systems, ontology-based search and other text analysis tasks.