Design and Development of a Named Entity Based Question Answering System for Malayalam Language

This is a Named Entity Based Question Answering System for Malayalam Language. Although a vast amount of information is available today in digital form, no effective information access mechanism exists to provide humans with convenient information access. Information Retrieval and Question Answering systems are the two mechanisms available now for information access. Information systems typically return a long list of documents in response to a user's query which are to be skimmed by the user to determine whether they contain an answer. But a Question Answering System allows the user to state his/her information need as a natural language question and receives most appropriate answer in a word or a sentence or a paragraph.

This system is based on Named Entity Tagging and Question Classification. Document tagging extracts useful information from the documents which will be used in finding the answer to the question. Question Classification extracts useful information from the question to determine the type of the question and the way in which the question is to be answered. Various Machine Learning methods are used to tag the documents. Rule-Based Approach is used for Question Classification.

Malayalam belongs to the Dravidian family of languages and is one of the four major languages of this family. It is one of the 22 Scheduled Languages of India with official language status in the state of Kerala. It is spoken by 40 million people. Malayalam is a morphologically rich agglutinative language and relatively of free word order. Also Malayalam has a productive morphology that allows the creation of complex words which are often highly ambiguous.

Document tagging tools such as Parts-of-Speech Tagger, Phrase Chunker, Named Entity Tagger, and Compound Word Splitter are developed as a part of this research work. No such tools were available for Malayalam language. Finite State Transducer, High Order Conditional Random Field, Artificial Immunity