

Informatics Institute of Technology

In Collaboration With

University of Westminster, UK

Agrisage

An HMM and Ontology based Cross-lingual Question
Answering system for the Agricultural Domain

A dissertation by

Ms. Niveditha Karmegam

Supervised by

Mr. Kaneeka Vidanage

Submitted in partial fulfillment of the requirements for the

BEng (Hons) Software Engineering degree

Department of Computing

April 2019

© The copyright for this project and all its associated products resides with Informatics
Institute of Technology.

Abstract

Although internet is the predominant source of information, the diversity and high heterogeneity of data makes finding relevant data a complex task. English is the predominant language that is being used in the internet, however only 26.8% of the internet speakers are said to be native English speakers. In developing countries like India and other Asian African countries, the most significant barrier in agricultural information dissemination is found out to be as Illiteracy and Diversity. With the globalized economy increasing, the need to find information in other languages is a necessity. Although cross-lingualism has been considered in many researches they mostly support popular languages like French, Chinese, Spanish, Hindi, etc. And most of the existing solutions are in the field of Information Retrieval. This research identifies the need for an expert system in this domain and introduces an Ontology based cross-lingual question answering system for Tamil Language.

The proposed system uses a hybrid of knowledge-based and machine translation, since machine translation alone will not be adequate for a domain restrictive system targeting a lowly-resourced language. Sense overlap based disambiguation is performed to reduce ambiguity during translations. And an Ontology based query reformulation is carried out to support short queries. The most prominent work carried out in this research is in ontology resource mapping component. To identify the most suitable Ontology resource for each user-supplied query a mathematical model is proposed, the Hidden Markov Model parameters are bootstrapped using the topology of the underlying graph, and the semantic relationship between the resources. This system is evaluated against 100 Tamil queries and gives an overall precision and recall of 68.33 and 65.95 respectively. The combination of sense-based disambiguation and the mathematical model for resource mapping makes Agrisage perform well and produce satisfactory results

Key words:

Expert System, Cross-lingual Retrieval, Hidden Markov Model, Natural Language Processing, Ontology, Agriculture