

Chapter 10

Tamil Question Answering System Using Machine Learning

Ashok Kumar L.

PSG College of Technology, India

Karthika Renuka D.

PSG College of Technology, India

Shunmugapriya M. C.

PSG College of Technology, India

ABSTRACT

Tamil question answering system (QAS) is aimed to find relevant answers in in the native language. The system will help farmers to get information in Tamil related to the agriculture domain. Tamil is one of the morphologically rich languages. As a result, developing such systems that process Tamil words is a difficult task. The list of stop words in Tamil has to be collected manually. Parts of speech (POS) tagging is used to identify suitable POS tag for a sequence of Tamil words. The system employs Hidden Markov Model (HMM)-based viterbi algorithm, a machine learning technique for parts of speech tagging of Tamil words. The analyzed question is given to the Google search to obtain relevant documents. On top of Google search, locality sensitive hashing technique (LSH) is utilized to retrieve the five relevant items for the input Tamil question. Jaccard similarity is used to obtain the response from the retrieved document items. The proposed system is modelled using a dataset of 1000 sentences in the agriculture domain.

INTRODUCTION

Internet provides high source of data but English is the dominant language. Though the people get information from search engine it may or may not be useful and relevant. They spend much of time to get the relevant answer. The key idea of this work is to find and represent the exact answer to the user's query. The difficulty of finding and validating the right answer makes a QAS complex than the general information retrieval task.

DOI: 10.4018/978-1-6684-6001-6.ch010

The system processes Tamil questions and provides answers in Tamil to the user in the agriculture domain. Uzhavan mobile application for agriculture domain in Tamil provides information regarding agriculture schemes and subsidy. But it does not give concise answers to the farmer queries. To address such problems Tamil QAS system gives exact answers to user queries.

Table 1 shows the different type of questions in Tamil. In this work, the user queries in the agriculture domain is processed by the proposed model to get answer in Tamil. The key idea of this system is to find agriculture domain based question answering system in Tamil language. Stop words in Tamil are collected manually and the Tamil words are POS tagged. Locality sensitive hashing is used to retrieve relevant documents. Finally the answers are ranked using Jaccard similarity.

Table 2 shows the different types of QASs. QAS are classified based on application domain, analysis done and techniques used. Based on application domain it is classified as open and closed domain. Based on analysis QAS is further classified as semantic QAS, syntactic QAS and morphological QAS. Based on technique used it is classified as web based and ontology based.

Table 1. Types of Question

Question Type	Description	Sample Question (In Tamil)
Factoid Question	Description and definition type of questions	வறண்ட லத்திற்கு எவ்வாறான பழப்பயிர்களை தரேந்தெடுக்க வணேடும்?
Listing Question	Types and list of questions	நற்பயிரதைத் தாக்கும் பஞ்சிகள்?
Affirmative Question	Yes or No questions	கிழங்கு சாகுபடிக்கு 20 ச.மீ இடவளெளி இருப்பது சரியா?

Table 2. Types of Question Answering System

Types of QAS	Description	Example
Open Domain QAS	Users can ask questions in any domain and get relevant answers from the system.	IBM Watson
Closed Domain QAS	Users can ask question in one particular domain such as medical, science domain.	QAPD (physics domain)
Ontology Based QAS	The system makes use of ontology such as DBpedia Ontology and SparQL query to answer users questions as a knowledge source.	AquaLog and DeepQA IBM Watson System
Web Based QAS	The system exploits web resources such as Google, Wikipedia as knowledge repository to answers the users.	WEBQA
Semantic Analysis based QAS	The system conceptually analyzes the user questions and provides exact answers.	QUERIX and PANTO
Syntactic Analysis based QAS	Keyword based analysis is done by the system to answer queries.	ASKME
Morphological Analysis based QAS	The Morphological analyzer breaks the word (பயிர்கள்) into root word (பயிர்) and associated morpheme features (கள்).	TamilQAS

8 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/tamil-question-answering-system-using-machine-learning/314142

Related Content

Agreement Technologies for Conflict Resolution

Vicente Julian, Victor Sanchez-Anguix, Stella Herasand Carlos Carrascosa (2020). *Natural Language Processing: Concepts, Methodologies, Tools, and Applications* (pp. 464-484).

www.irma-international.org/chapter/agreement-technologies-for-conflict-resolution/239951

Deep Learning for Sentiment Analysis: An Overview and Perspectives

Vincent Karasand Björn W. Schuller (2021). *Natural Language Processing for Global and Local Business* (pp. 97-132).

www.irma-international.org/chapter/deep-learning-for-sentiment-analysis/259785

Logs Analysis of Adapted Pedagogical Scenarios Generated by a Simulation Serious Game Architecture

Sophie Callies, Mathieu Gravel, Eric Beaudryand Josianne Basque (2020). *Natural Language Processing: Concepts, Methodologies, Tools, and Applications* (pp. 1178-1198).

www.irma-international.org/chapter/logs-analysis-of-adapted-pedagogical-scenarios-generated-by-a-simulation-serious-game-architecture/239985

Preservation of Cultural Heritage in an Ethnic Minority Using Internet of Things and Smart Karaoke

Alberto Ochoa, Roberto Contreras-Masse, Jose M. Mejiaand Diego Oliva (2021). *Handbook of Research on Natural Language Processing and Smart Service Systems* (pp. 180-195).

www.irma-international.org/chapter/preservation-of-cultural-heritage-in-an-ethnic-minority-using-internet-of-things-and-smart-karaoke/263101

Smart Grid and Demand Side Management: Application of Metaheuristic and Artificial Intelligence Algorithms

Alper Ozpinarand Eralp Ozil (2020). *Natural Language Processing: Concepts, Methodologies, Tools, and Applications* (pp. 139-159).

www.irma-international.org/chapter/smart-grid-and-demand-side-management/239934