# Kisán Váṇi: Android based Agricultural Helpdesk System for Kannada Language

Roshan Fernandes
Asst. Professor
Dept of CSE
NMAMIT, Nitte

Shruthi M. N.
Dept of CSE
NMAMIT, Nitte

Sharmada B.
Dept of CSE
NMAMIT, Nitte

## ABSTRACT
The evolution of World Wide Web has such an impact on human being is that people not only access internet on desktop but also on mobile in their day to day life. The inventions of mobile phones have affected human life through their varieties of applications they provide. They are useful in times of emergencies and make it easier to access information on the go. Also, most mobile phone applications are very user friendly such that even the common man has the ability to use it. There are people in our country who cannot read or write English and sometimes even their own mother tongue. These people find it difficult to navigate the Internet and access information that is readily available. This applies particularly to the farmers. Farmers in general, form the backbone of our country but we will be concentrating on the farmers of Karnataka. This paper gives a description of existing systems and implementation details of Android application that can be easily used by farmers to solve their day-to-day problem queries related to agriculture and farming. There is no typing or reading required. All the farmer has to do is press a button, speak his problem and he will get the solution in Kannada.

## General Terms
Speech recognition, Natural Language Processing, Algorithm, Agriculture, Web mining, Web Services.

## Keywords
XML, SOAP, WSDL, UDDI, Discovery, HTML, Speech synthesis.

## 1. INTRODUCTION
### 1.1 Web Services
Web service is a method of combining many web based applications using XML SOAP, WSDL and UDDI standards over an Internet protocol. Web service composition gathers information from existing web services on specific topic so as to satisfy the requirement of complex applications [1]. Composition can be done in two ways: Syntactic and Semantic. Syntactic web composition cannot compose web services together which is heterogeneous. In order to solve this problem, semantic web was proposed. [1] proposes an approach of web service composition which divides every composition process into two basic modules, the composition generation and execution of web service composition. A survey on web services showed that a standard agricultural web service doesn't exist. Therefore, we use web mining instead.

### 1.2 Web Mining
Extracting previously unknown information from large quantities of data such as text, audio, video, etc, and cleaning and modifying them according to the user need is known as data mining. When such a method of data mining applied to a web data, we call it as web data mining or wed mining. Web mining is nothing but discovery and analysis of useful information from the World Wide Web[2]. According to [3][4] categorizes web mining into three different categories:

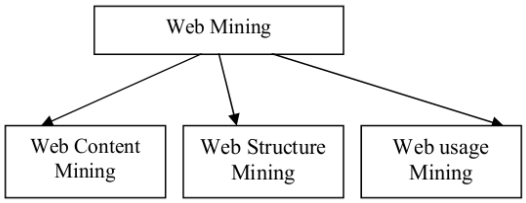



**Figure 1: Web mining classification**

1. The purpose of web content mining is to analyze the data available in the different web documents. This may include all types of data such as text, images, audio files, etc.
2. The purpose of Web structure mining is to study on link information. It analyzes the way in which different web documents are linked together.
3. The objective of Web usage mining is to study about the transaction data from log files. Usage mining is sometimes referred to as 'log mining', because it involves mining the logs of web server which has been written during user interaction with the web server.

In this paper, we are using web content mining. Web Content Mining is mining data from the content of web pages [5], where data consists of text, graphics, tables, data blocks and data records. Web Content Mining uses the concept of data mining and process of knowledge
discovery. Due to the large number of documents available on the Web, it is difficult to provide relevant information. A survey of many web pages showed that main content is mostly in the <p> tag. So, in our implementation we extracted the content from <p> tag using regular expression.

## 2. EXISTING TOOL
The development of a Hindi speech recognition module of VideoKheti[6], a mobile video search application for farmers is taken for our literature survey. Here, the Salaam Method [7] is used to to get a high quality speech engine in English to develop a mobile speech based application video search for farmers in India. Predefined Hindi words are stored in the database in the form of voice samples. The input voice is

compared against these stored samples and based on the result, related videos are shown.

# 3. ARCHITECTURE

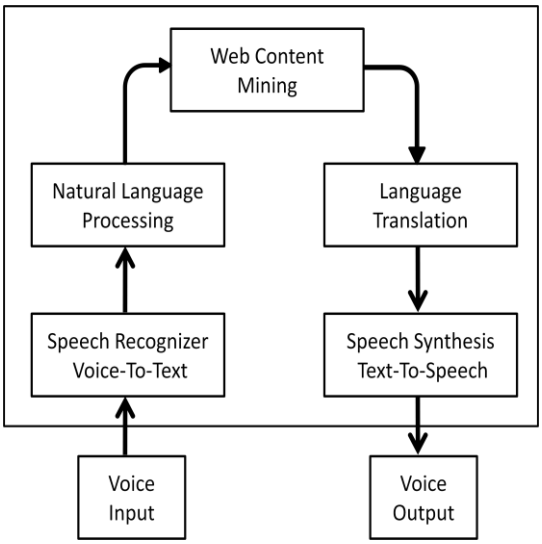



**Figure 2: Architecture of the proposed system**

## 3.1 Speech Recognizer

For this component, the user gives Kannada voice input using a microphone which is coupled to the system.

## 3.2 Natural Language Processing

The user speech input will be in the form of a sentence. From this sentence, a keyword related to agriculture is extracted by comparing the captured .wav file with the pre-stored .wav files that are present in a database.

## 3.3 Web Mining

The keyword obtained from NLP, is used to mine the web wherein many agriculture based web pages are searched for relevant information. This content is extracted from the HTML source code of web pages.

## 3.4 Language Translation

The shift reduce parsing technique is used for translating English sentences into a grammatically correct Kannada sentences by reordering of English parse tree structure, generating and implementing phrase structure grammar(PSG) for Kannada sentences the [8].

## 3.5 Speech Synthesis (Text-to-Speech):

Then the Kannada sentences are converted to speech by using UNICODE of Kannada letter.

# 4. IMPLEMENTATION

The Kannada speech input is captured and stored in .wav format using Android APIs. These are then compared with stored .wav files using FFT (Fast Fourier Transform) and DSP (Digital Signal Processing) concepts, to get the agriculture related keyword in the form of text.
Keyword obtained is used to access the webpage information related to it. A list of URL of web pages is maintained. The required content is present in <p> tags of HTML code. These tags are extracted using regular expressions.

The extracted content is in English text which must be converted to Kannada text. This is done using English to Kannada Sentence Translation using POS Tagger by Reordering Parse Tree Structure [8]. The Phrase Structure Grammar used to construct the parse tree is as follows:

S → NP VP

NP → DET Propernoun | Propernoun | NP NP | NP NP NP | adj NP

VP → Verb | Verb NP | Verb NP PP | Verb PP

PP → preposition NP

Once the tree is constructed, the Kannada words are tagged to the corresponding English words and Kannada Grammar Rules [8] are added by parsing the tree to get meaningful Kannada sentence.
Kannada sentences are then translated to Kannada Speech by giving the Unicode of each letter to the TTS engine. TTS engine will readout the words in Kannada.


Algorithm:

Input: An agriculture related query in Kannada speech

Output: A solution to the given query in Kannada speech

1. Accept a word related to agriculture in the form of voice from the user.
2. Match given word to the set of keywords stored in the database.
3. If the keyword is present in the database,
4. Extract corresponding English keyword and store in a String variable
5. Else
6. Request for Valid keyword
7. Initialize web mining procedure by passing the keyword in the form of English text as an argument
8. Search for information relating to keyword in a predetermined set of urls stored.
9. Extract <p> tags and their content from HTML document
10. Get relevant content in the form of sentences.
11. English Sentences are then translated to Kannada using grammar productions for English and implementing Phrase Structure Grammar (PSG) for Kannada.

12. Unicodes of Kannada text is given as input for text-to-voice convertor.

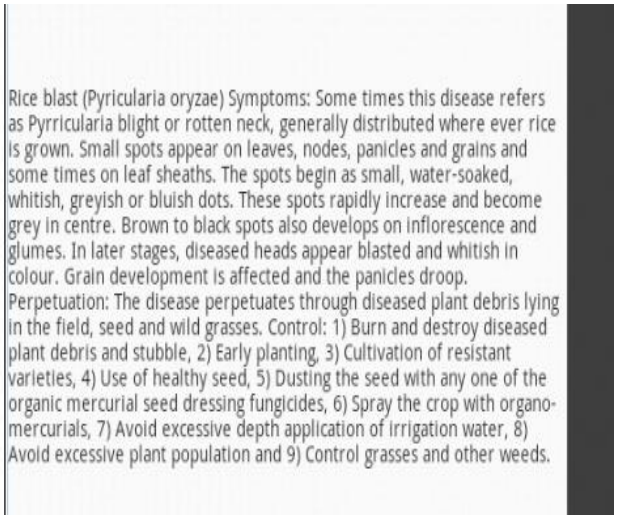
13. Corresponding Kannada speech is given as output.

## 5. RESULT

Since no standard web services related to agriculture were found, many web pages are used to extract the required information. Some of the urls used are:

- https://www.plantvillage.com
- http://www.agriinfo.in
- http://irri.org/our-work
- http://agritech.tnau.ac.in

-http://www.ca.uky.edu/agripedia




-www.agraskmanagementforum.org

The content of these pages are composed and converted to Kannada text that is then given as Kannada speech output. This can be easily understood by farmers.

## 6. FUTURE WORK

Different regions of Karnataka have different dialects of Kannada. This project can be improved to support all versions of Kannada or any other regional language.

## 7. REFERENCES

[1] GUO Song, "A New Approach for Web Service Composition Based on Semantic",

[2] R. Cooley , B. Mobasher, and J. Srivastava, "Web mining : Information and Pattern Discovery on the World Wide Web, IEEE 1997.

[3] C. S. Bhatia, Dr. Suresh Jain, "Semantic Web Mining: Using Ontology Learning and Grammatical Rule Inference Technique"

[4] S. Madria, S. Bhowmick, S. Sourav, W. K. Ng & E. M. Lim "Research Issues in Web Data Mining", Citeseer Beeta, 1999, P.4-12

[5] Xu, G.; Zhang, Y.; Li, L. (2011) *Web Mining and Social Networking*, Techniques and Applications, Australia: Springer

[6] Kalika Bali, Sunayana Sitaram, Sebastien Cuendet, Indrani Medhi, "A Hindi Speech Recognizer for an Agricultural Video Search Application", ACM, 2013.

[7] Chan, Hao Yee and Rosenfeld, R. Discriminative Pronunciation Learning for Speech Recognition for Resource Scarce Languages. Proc. ACM DEV 2012, Annual ACM Symposium on Computing for Development, March 2012,Atlanta, GA.

[8] Sharanbasappa Honnashetty, Mallamma Reddy & Dr. M Hanumanthappa, English To Kannada Sentence Translation Using POS Tagger by Reordering Parse Tree Structure, International Conference on Computer Science & Engineering (ICCSE), 17th March-2013, Pune, ISBN: 978-93-82208-74-7