International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 2, Issue 11, November 2015

Artificially Intelligent Student Information Network

[1]Ashwin Suresh Babu, [2]Krishna Pokkuluri
[1] Bachelor of Engineering in Computer Science, [2] Bachelor of Engineering in Computer Science
[1]awesomeash2014@gmail.com, [2]kriskrishna270@gmail.com

***Abstract*— This paper provides the ways to implement an application to act as an artificially intelligent user-friendly agent to mine the available information feedback from students and faculties to present it appropriately. It will also perform the appropriate probability and numerical calculations for setting up question papers, assigning salary, fee structure and many more. Further it will also provide a network which connects the college community without the need of network providers or Internet. The user interaction is either through touch or voice input and the output is viewing or downloading the related information or establishing an easy communication between the college fraternities.**

***Index Terms*—Artificial Intelligence, Natural Language Processing, Data Warehousing, Data Mining, Wireless Mesh Networks**

I. INTRODUCTION

The 21st century has brought about many innovations which are mostly in automation to make the lives of people easier and better to live. People look up on technology to artificially sense their needs for reducing their burden and the time taken to complete their jobs. People expect the machine which is an artificial intelligent equipment to sense his / her needs. We would also want to relate relevant day to day information and feedback to get a different perspective to the available information which is useful for interpretation, analysis and evaluation. Also an inexpensive means to communicate with people in nearby places without the need for connection to the internet is the need of the art. The artificially intelligent student information network would make all of the above a possibility under a single cloud.

We have already seen how the usual college management works. The attendance, the marks, the syllabus, the time table and all are drawn up or recorded manually. This actually takes some time and is also not perfect. We may not be the perfect judge as to which a tough subject is or who the best teacher to handle crisis is. Further students and faculties would want up to date information in their hands to decide on anything such as the study pattern, the important topics to be covered and all that. All the information related to placements are not authentic or incomplete in some cases in certain websites. We may need to communicate with a person within the college (say 100 meters) wirelessly without incurring any additional expense or overhead. So we are bringing out an artificially intelligent application which heavily uses data mining to present the best knowledgeable information about the best teacher for each subject, the rating for the teacher and student likewise, performance of teachers and likewise for students, time table framing, identified tough subjects, all previous year question papers and then the list of important questions. And further information related to placement is also available. The application facilitates a communication medium using wireless mesh network. The aforementioned information is presented dynamically and not by the administrator by the Artificial Intelligent agent using Data Mining.

The Artificial Intelligent Agent for the application can be implemented using the Natural Language Processing for voice based input and set of other codes written for arithmetic and probability calculation using the Python Programming Language. In case of a voice input, there are 3 stages which is the syntactic, semantic and pragmatic analysis. This will basically act as a decision support system. For implementing the Data Warehouse for our application, we make use of MySQL and for analysis and interpretation we make use of Python Database API to connect to MySQL database. We simply require a radio transmitter with 802.11g standard with support over 100 meters and a 2.4GHz radio spectrum.

II. LITERATURE SURVEY

An extensive study was made on different existing applications which inspired us to come up with this application for the college. Siri in Apple IOS allows you to use your voice to send messages, schedule meetings, place phone calls, and more. But the functionalities of less than

what everyone would imagine. For example, we can't ask Siri to compute a salary waiver for each faculty or draw up time tables which we regularly do manually in college as a result. Enterprise Resource Tools such as Fedena Pro Support are in their native state and requires human skill to extract vital information from available patterns. The brain child of this innovation has to be the WebFOCUS Business Intelligence platform. It does higher end analytics and presents graphical comparative and vital information. But while this focusses only on analytics and interpretation, our application focusses on the higher level college activities such as periodic calculation of salary of teachers based on college performance and rating and also with the help of this application, we can set up question papers with different range of difficulties. During the recent Chennai floods, Fire chat served as the medium for communication and passing vital information. In our application, we are extending these features to a voice call within the college premises with all the devices supporting the application as a node in the wireless mesh network. The book on the Natural Language Processing with Python by Steven Bird, Ewan Klein and Edward Loper tells you how Python Programming language is simple and powerful for including excellent functionality for processing linguistic data and presenting it appropriately to the user. We can construct a powerful data analysis tool using MySQL and Mining tool using Python.

III. MATERIALS AND METHODS

A.Materials Required and Feasible Environment

The application is made utilizing the following: a workable computer with operating system that supports the Python programming language, a Python platform with version 3.4.2 which includes the support for PyBrain library, support for MySQL database description, manipulation, control and transaction, Database API to support MySQL connection to Python code and a radio transmitter with support for 802.11g standard. This application does not suffer from any major interference from outside environment and the establishment of wireless mesh network ensures that there is almost zero interference even when the signal is in contact with some physical medium.

B.Program Design

The Natural Language toolkit, or more commonly NLTK, is a suite of libraries and programs for symbolic and statistical natural language processing (NLP) for the Python programming language. It segments the sentences and tags these according to the part of speech. The program can be designed in the language that is suitable to both the tablet device and the Personal Computers and which properly incorporates the Natural Language Processing Toolkit. The Bluetooth specification and characteristics are similar to the existing Bluetooth Technology. There are three major aspects of any natural language understanding theory: The syntax describes the form of the language. It is usually specified by a grammar. The semantics provides the meaning of the utterances or sentences of the language. The pragmatic component explains how the utterances relate to the world. To understand language, an agent should consider more than the sentence; it has to take into account the context of the sentence, the state of the world, the goals of the speaker and the listener, special conventions, and the like [1]. Machine learning is the science of getting computers to act without being explicitly programmed. Machine learning is a scientific discipline that explores the construction and study of algorithms that can be learned from data. Such algorithms operate by building a model based on inputs and using that to make predictions or decisions, rather than following only explicitly programmed instructions. This plays a very important role in including different algorithms for implementation. All machine learning algorithms (the ones that build the models) basically consist of the following three things.

1. A set of possible models to look thorough.
2. A way to test whether a model is good.
3. A clever way to find a really good model with only a few test with which any function can be included.

A useful data type built into Python is the dictionary. Unlike sequences, which are indexed by a range of numbers, dictionaries are indexed by keys, which can be any immutable type; strings and numbers can always be keys. Tuples can be used as keys if they contain only strings, numbers, or tuples; if a tuple contains any mutable object either directly or indirectly, it cannot be used as a key. A pair of braces creates an empty dictionary: {}. Placing a comma-separated list of key: value pairs within the braces adds initial key: value pairs to the dictionary; this is also the way dictionaries are written on output. The main operations on a dictionary are storing a value with some key and extracting the value given the key. It is also possible to

delete a key: value pair with DEL. If you store using a key that is already in use, the old value associated with that key is forgotten. It is an error to extract a value using a non-existent key. The keys() method of a dictionary object returns a list of all the keys used in the dictionary, in arbitrary order (if you want it sorted, just apply the sorted() function to it). To check whether a single key is in the dictionary, use the "in" keyword.

Enterprise Data warehouse contains all information about subjects. The Data Mart is the departmental subset that focuses on selected subjects. The aforementioned artificial intelligence mechanism forms the Decision Support System (DSS). We perform drill downs and then maintain meta-data for our warehouse. The data about the fee structure, marks and salary forms the operational data. The user guidance support forms the informational data. We make use of the Structured Query Language to add the primitive data and the feedback which is in short the operational data. The informational data is added using Python snippets as mentioned above. We use MySQL Database Management system as our Relational Database Management System for the warehouse.

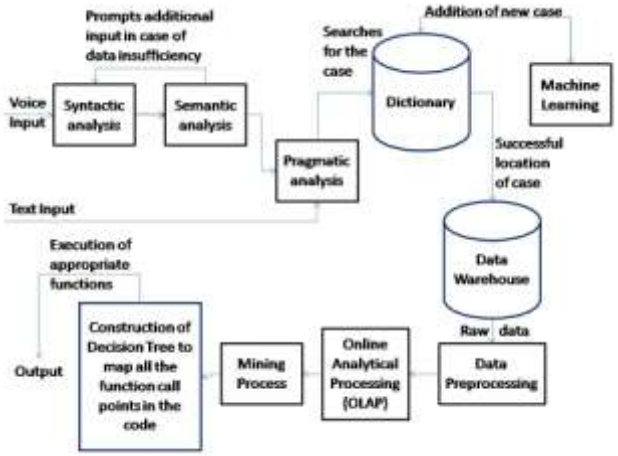
Data mining refers to extracting or mining knowledge from large amounts of data. The tasks involved in Data Mining: characterization, discrimination, association, classification, prediction and clustering analysis can be implemented in the Python programming Language. Main objective is the identification of a particular pattern to aid interpretation. This pattern may be either subjective or objective. We include functionalities to perform preprocessing of data to present the final data, the decision tree and also the probability prediction using Bayesian network which is a representation of variables or data and their conditional dependencies. Outliers can be implemented for safety of privacy [2].

We implement the communication feature using the wireless Mesh topology which is a communication network in a mesh topology made up of radio nodes and this network would not require network provider and Internet service [3]

C. Methodology

The following working model is explained keeping in mind its implementation using the Python programming language. This application supports both speech and text input and they both would meet at the Pragmatic Analysis and thereafter the functionality is common.

Fig. 1 The working model of Artificially Intelligent Student Information Analysis and Extraction



The input is given either in the form of text or voice. In case of text, you will be put through a series of selection and choosing interfaces until you get the desired result. If it voice, then first the syntactic analysis is carried out to tokenize the sentences we speak. Then the semantic analysis, groups the necessary tokens to derive the real meaning it intends to put across. Finally the results of either text or voice goes to the Pragmatic Process module which decides what has to be done for the user request. The dictionary will check out if such user request can be handled. Else it adds the required functionality or data in the Warehouse automatically to handle such requests using machine learning module.

Then the control moves to the Data Warehouse containing a large number of databases or Data Marts. The tuples which we require is fetched from appropriate Data Marts. This raw data or tuples has to be preprocessed before analyzing the pattern between different relevant data. In this we perform cleaning, integration, transformation, reduction and finally discretion. Online Analytical processing supports execution of query tasks to get the relevant data from the tuples. The database manipulation and analysis is handled by the MySQL Relational Database Management System.

Now we utilize the database API of Python to connect the MySQL with the Python code. Now we are utilizing Python codes instead of Procedural and trigger statements of MySQL as Python includes easier and more functionalities. The Mining Process takes with the queried data as input and it involves several internal process to

obtain the required pattern. Data characterization involves summarization of the general characteristics or features of target data. In Data Discrimination, we perform comparison of the general features of target data with the general features of data from one or a set of contrasting classes. Then we establish a proper association relationship between different relevant datum considered. And then we classify the data based on numerical and probabilistic computations. Finally we cluster the data to include only those which we need to draw up a pattern.

With the help of that pattern we attempt to trace a decision tree which is a representation of the control flow across different functions to present the needed result.

Fig. 2 The Data Mining Process

The following is an illustrative working model of the wireless mesh communication network.

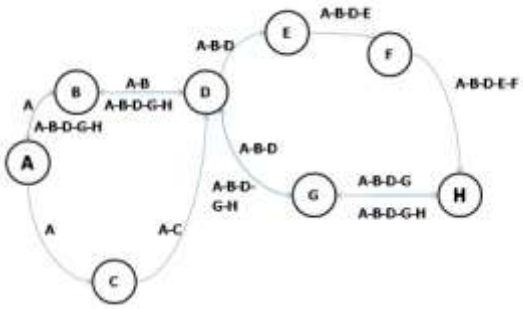


Fig. 3 The topology of Wireless Mesh Network

Suppose radio mesh node A wishes to communicate with node H which is within 100 meter radius, the slow motion of the process which is similar to Dynamic Topology is explained as follows [4]. The A node first locates its neighboring nodes and sends its delivery request. Both of its neighbors acknowledge the only request that they receive and since they are not the intended senders they pass on the information to their common neighbor which is node D. Now node D acknowledges the request which it receives first so as to support faster connection. This process repeats until the destination node H is reached. This H node sends an frame indicating that it is ready to place a call. This frame header has the address, which was attached to the header of each nodes at a shorter distance from the A node to the H node, for easy path trace back. And then the communication takes place.

D. Problems faced

There are no major problems in the implementation of the application. But efficiency of Natural Language Processing in real world implementation is not exactly 100 percent perfect. But it is just about accurate for the system to understand and generate an algorithm. The easily or mostly solvable problems include Spam Detection, tagging, named entity recognition. Further problems include Sentiment analysis, Co-reference resolution problem, and Word sense disambiguation problem, Parsing, Machine Translation and Information Translation. These problems are solvable if proper research is done and corresponding implementation is included. Some of the problems which are almost impossible to solve are summarization of input and implementation of a dialog system that prompts a related query to the input if the input is ambiguous. Our system simply reacts by saying something like “I do not understand you. Come again”. The world is new to the concept of Data Warehousing and Mining, so it would take some amount of time before we can understand how different patterns are related. Also we would not know firsthand how many databases are required as it is powered by artificially intelligent machine learning modules as opposed to the static databases added in the existing Enterprise Resource Tools. As a result, we can implement the Warehouse only by the Bottom-Up approach which is the incremental approach where in database can be incrementally added upon request. As in the case of the communication network, the problem is the range and also the primitive privacy policies for the mesh nodes in the network.

IV. RESULTS

Suppose a department Head wants to set up the time-table for the new semester, here is how the process takes place in the application. The input would be "Please schedule the time-table". Due to the lack of information at the semantic phase, the application prompts the user to specify the branch and year of course. From the tokenization, the key words identified were "time-table", "CSE", "4th". Now referring to the dictionary, the appropriate case was located by the agent and the input was understandable. Now it fetches the Subject-Credit-syllabus table and the Subject-Result-year table from the Data Warehouse. After prior preprocessing, we extract the tuples needed for that particular branch and year from the first table and then we extract the previous year results of 4th year Computer Science and Engineering alone from the second table. In the mining process we obtain a pattern relating the difficulty of the subjects to the results and subsequently the allocation of number of classes. If some subjects have almost the same level of difficulty, then probabilistic calculations are made which decides which subject should be given the most number of classes after the comparison with other subjects. The data virtual tree is pruned to contain only the subject information and the related pattern. Finally a subject can be allocated at different times based on the conditional probability. Here is a decision tree snippet to illustrate class allocation.

Assigning difficulty,

```
IF  (result > 80) THEN difficulty = "easy";
ELSE IF (60 <  result <=80) THEN difficulty = "medium";
ELSE difficulty= "hard";
```

Same level of Difficulty issue,

```
IF difficulty= "hard" and probability="yes" THEN
num-class=5;
ELSE IF difficulty= "hard" and probability="no"
THEN num-class=4;
ELSE IF difficulty= "medium" and probability="yes"
THEN num-class=4;
ELSE IF difficulty= "medium" and probability="no"
THEN num-class=3;
ELSE num-class=3;
```

Suppose vital information is required related to the placement, we can collect the feedback of the seniors through the application, create a new database in the Warehouse and extract an pattern relating the placement and its trend. It will provide appropriate study materials for preparation, make the company profile evident and then making aware of the best companies in recruitment business by studying the table containing the company recruited year, base salary offered and the like.

Another instance would be the salary increment calculator of the faculties based on their rating and performance. Consider the students submitting the rating of their teachers periodically, then finally she will have a particular rating at the end of the semester and also her performance evaluated based on the pass percentage. The respective fields are fetched from the Warehouse and the average of the above 2 is calculated during the classification phase of the mining process. Then the following Decision Tree Snippet shows what to do:

Average Calculation,

```
current_average= ( rating+performance )/2 ;
```

salary increment calculation,

```
increment=( current_average-previous_average );
IF  increment>0 THEN salary+=increment*1000;
ELSE salary+= 0;
```

We can easily set a easy or a moderate or a tough question paper by utilizing the Bayesian Network mechanism in the Prediction phase. It is based on Conditional probability relating the probability of maximum or minimum occurrence of question with the probability of occurrence of other similar questions of the same discipline.

The teacher in need of the student can simply say "Place a call with Ashwin of 4th year CSE". If he is within 100 meters, say in laboratory, the call will reach the student without the need of Network provider or Internet pack.

V. CONCLUSION

This application is generic and can be extended to different areas such as food chains, tourism, defense, agriculture and the like. Future likely addition to the above application would be the sensors. For instance to automatically sense and record the attendance saving further time. This would lead to the creation of a powerful Internet of Things tool which is the need of the hour at the moment.

REFERENCES

[1] Steven Bird, Ewan Klein and Edward Loper, "Natural Language Processing with Python," O'REILLY.
[2] Alex Berson and Stephen J. Smith, " Data Warehousing, Data Mining & OLAP", Tata McGraw – Hill Edition, Tenth Reprint, 2007.
[3] Jiawei Han and Micheline Kamber, "Data Mining Concepts and Techniques", Elsevier, 2007.
[4] http://www.howstuffworks.com/how-wireless-mesh-networks-work.htm