Virtual Student Advisor using NLP and Automatic Appointment Scheduler and Feedback Analyser

N. Suvethan[1], K. Avenash[2]. M.A.Q. Huzaim[3], R. Mathusagar[4], M. P. A. W. Gamage[5], A. Imbulpitiya[6]

suvethann10@gmail.com[1], Avenash555@gmail.com[2], huzaim2006@gmail.com[3], mathusagar@gmail.com[4], anjalie.g@sliit.lk[5], asanthika@sliit.lk[6]

Faculty of Computing, Sri Lanka Institute of Information Technology, Sri Lanka

ABSTRACT

Virtual Student Advisor is a research project that mainly concerned on addressing a comprehensive solution to overcome the difficulties faced by the academic departments of any academic institutions. According to the context, the role of the Student Advisor is focused on helping students with problems related to their academic carrier at the University and also answer the general queries made by students related to the procedures conducted at the Institute. This happens to be a tedious and a very inefficient task for the Academics as they have to repeat the same answers for many students and also students coming to meet the academic in ad hoc manner without proper appointments makes both the student and the academic face lot of problems.

The Virtual Student Advisor system mainly consists of three components; A Natural language based inquiry management module, Priority wise automatic appointment scheduler and Feedback analyzer. The inquiry management module is responsible in handling user queries based on frequently asked questions. Users can get the answers for a query by entering the question in natural language. This is implemented as a mobile application as it is convenient for students to ask questions from any place at any time. The system will answer the queries and if it needs more explanation it will be directed to the relevant advisor. Auto appointment scheduler handles the student appointment requests and helps in managing the schedules. This will be accessed by both Academics to set their free time slots and the students to request an appointment according to urgency. Feedback analyzer handles the entire process of student feedback taken for each subject starting from preparing feedback forms till analyzing the collected information. The lecturer can prepare a feedback form using the question bank in the system and customize it further to fit the requirements and after conducting the feedback the system can process the data and provide reports on varies views of the data such as tables and pie charts.

The system uses Natural Language Processing to handle students' queries by tokenizing the sentences and extracting answers based on keywords and comparing synonyms WordNet lexical database. The appointment scheduler uses priority based Round Robin CPU Scheduling Algorithm to schedule the appointments according to urgency. The Virtual Student Advisor system consists of a mobile application for students and a web application for both students and lecturers to access the information and conducted their daily tasks related to academic administration effectively.

Key Words: Algorithm, Natural Language Processing, Text Classification

1. INTRODUCTION

In any kind of university, the academic department plays a vital role in the administrating the institution. Even though the technology has been influencing the working practices a lot, most of the day to day tasks handled by the department are still done using manual procedures. Due to this reason the responsibilities carried out by the Student Advisors are becoming more complicated.

Virtual Student Advisor is focused on providing a technical solution to overcome the difficulties faced in the domain. The system is comprised of a web application and a mobile application which is supported by a backend web service. The functional components of the system as follows:

- Inquiry Management module
- Automatic Appointment Scheduler

- Feedback Analyser

Inquiry Management Module is mainly targeted to handle queries of the students. It's one of the major problems faced by the academic advisors. This situation becomes really complicated for the staff when students visit them continuously with the same questions. When considering institutions with large student population the task creates a hectic situation.

On the other hand with the very less amount of Student Advisors available, meeting them during the working hours is very critical because of their busy schedule. Even though as a traditional practice the students are requested to get an appointment by contacting them through email before coming in person, many don't follow it. The communication between the students and academic advisors are distracted in many ways. Because of these issues the students really face difficulties in getting their inquiries and questions clarified at the right time. It also creates a messy situation in the office premises during the working hours when students roam around and looking for lecturers.

The problem faced by the academic department on this concern is unavoidable, because the inquiries of the students need to be assisted and solved on time. In the meantime difficulties faced by the staff and students must be considered equally.

Another important task handled by the academic department is collecting and monitoring the feedback about the lecturers and subjects. Manually analyzing the feedback is a tedious task which involves a massive human effort. Starting from preparing a feedback form to analyzing the collected information involves many numbers of phases. The basic steps of feedback handling process are as follows,

- Preparing feedback forms
- Printing and distribution of forms to students
- Collecting manually filled forms
- Summarizing results
- Analyzing the information
- Making decisions depending on the Information

It's very difficult to keep track of the progress and past records with the currently prevailing manual method.

The Virtual Student Advisor research project is targeted in providing an enhanced and effective way to carry out the day to day tasks of academic departments of any university.

This system includes three main components to cater the solutions for different identified needs in the context of academic departments. Managing inquiries based on frequently asked questions is a key feature of the system. The students can post a question and get the system answered for it based on frequently asked questions. The Student Advisors can add new questions to the repository and manage them. The system also includes a priority wise automatic appointment scheduling feature. Students can request a timeslot to meet a Student Advisor through the system. Once the request is made, the system processes it and assigns a timeslot by considering the priority of the request and the availability of the staff. Also the system provides with additional functionalities to manage the timeslots and appointments.

Feedback analyzer component is mainly focused on managing the feedback process of institution. The Student Advisor can create customized feedback forms through the system depending on the requirements and publish it to be accessed by the students. Students can simply provide their feedback on lecturer/module through the system itself. The feedback analyzer helps in summarizing the results and presenting them in an understandable manner.

2. METHODOLOGY

Several technological approaches have been used in order to develop the question answering system. The Inquiry Management module mainly uses the concepts of Natural Language Processing in its internal operations.

1. Inquiry Management Module

Inquiry management module of the virtual Student Advisor system is concerned on the successful establishment of the question answering system based on frequently asked questions. Fig. I shows how retrieving an answer for an entered question happens in two phases; Question Processing and Question Comparing.

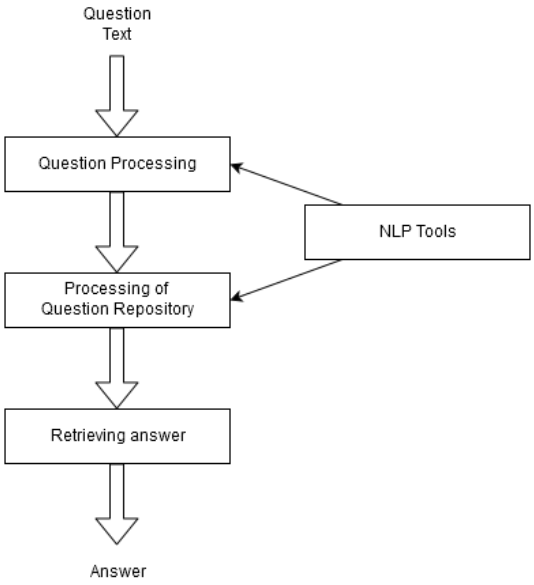

Fig I: Flow of Question Answering

Question Processing

Question Processing involves procedures to extract subjective information from question text entered by the user [7]. The main goal of question processing is to produce structural information of the entered text against the language grammar. Syntactic and semantic analysis techniques of Natural Language Processing have been used in order to capture the keywords from the question text.

- Syntactic analysis: process of analyzing a string of symbols, in natural language, conforming to the rules of a formal grammar.
- Semantic analysis: process of relating syntactic structures, from the levels of phrases, clauses, sentences and paragraphs to the level of the writing as a whole, to their language-independent meanings.

The question text has been parsed using lexicalized parser and the keywords of the question are extracted from it [1]. The noun phrases in a question will be detected through parsing and the phrases matching to the abbreviations in the user maintained dictionary will be replaced by the actual terminology [4].

Question Comparing

Once the keywords are extracted from the question sentence, it need to be compared with the frequently asked questions in the repository in order to find the answer [2] as showed in Fig. II. The Text Classification approach of natural language processing has been used in the process of finding the best matching record. The main objective of the text classification is to use machine learning, learn classifiers and perform category assignments automatically.

The process of text classification mainly begins with transforming documents, which is typically set of strings, into a representation suitable for the learning algorithm and the classification task [6]. The process basically ends up with attribute value representation of a given piece of text. Each distinct word corresponds to a feature, with the number of times the word occurs in the document as its value. Stop words are ignored when training.

The outputs of the question processing phase have been directed to the comparison stage. The information from the entered text has been tokenized and for each word token, the language synonyms have been appended in order to prepare the document. WordNet lexical database has been used to extract synonyms of each word [5]. Then the resultant document has been classified using Vector Classification approach to generate the respective term – vector map and the system is trained to classify any document against it [3][8]. Then the list of frequently asked questions from the database will also go through the same process and finally produces the map of frequently asked question record with the respective probabilistic score of relatedness.

The probabilistic score has been calculated depending on the Cosine Similarity between any two documents. Ideally more the score closer to 1 it is expected as more semantically related.

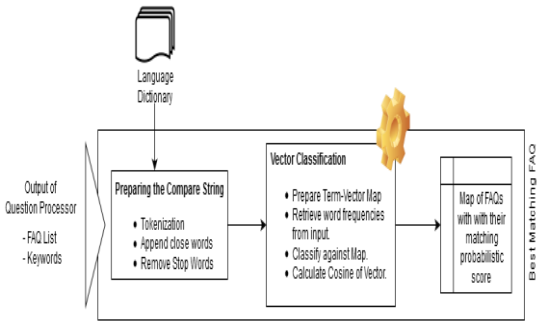

Fig II: Flow of Question Answering

2. Auto Appointment Scheduler

As shown in Fig .III the automatic appointment scheduler is responsible for scheduling the appointment with the academic advisors by considering the priority and the urgency of the request. The scheduling must be done without any much user involvement once the advisor adds all the timeslots to the system.

Based on research the priority based Round Robin CPU Scheduling Algorithm has been

selected to implement the auto appointment scheduling between students and advisors. It has the least waiting time comparing to other scheduling algorithms. Priority based Round Robin CPU Scheduling Algorithm is a combination of Priority scheduling algorithm and Round Robin Scheduling algorithm.

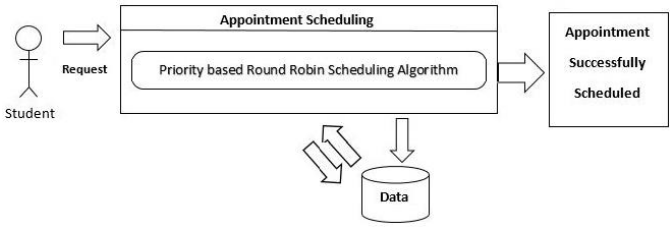

Fig III: *Scheduling Process*

Once the student made a request, the system checks for free timeslots, if there is a free slot then it will be assigned for the student depending on his/her priority. The scheduling based on FIFO (First In First Out) for the requests with same priority.

The system also records requests even though there are no free timeslots available. When a free timeslot is added, the existing requests will be scheduled with respective to their priorities.

3. Feedback Analyser

Feedback analyser module is mainly focused on automating the entire feedback collection process. This module is an alternative to the handling of feedback process in manual. The system allows the user to create customized feedback forms. The forms are stored in the system so that it can be reused depending on the need. The forms can contain different types of questions which can have choice based answers or text based answers.

Once the student submits the form, the sentimental analysis technique has been used internally for the questions with text based responses in the feedback form. When an answer for a text based question is submitted, it then needs to be analysed to understand and extract the entities and opinions related to those entities from it. Different levels of analysis techniques involved are,

- Document level (The task at this level is to classify whether a whole opinion document expresses a positive or negative sentiment)
- Sentence level (The task at this level goes to the sentences and determines whether each sentence expressed a positive, negative, or neutral opinion.)
- Entity level(a structured summary of opinions about entities and their aspects can be produced, which turns unstructured text to structured data)

The approach of Document Level analyzing has been used to analyze the answer for a given question. In here opinion mining will be done in document level, So that each question will be considered as an entity and assumed each answer focuses on single object or entity and the extracted opinions are belongs to that single entity.

Since the user is allowed to give answer in natural language, it is really complicated for the system to directly identify all the opinion words. The system uses the dictionary based approach to identify the set of opinion words which have list if synonyms and antonyms to compile opinion words.

The most important asset for analysis are the sentimental words which are commonly used to express positive or negative opinion. The two types of sentimental words are, Base type (e.g. good, bad, beautiful) and Comparative type (e.g. better, worst).

The system uses the Dictionary based approach to identify the set of opinion words which have list of synonyms and antonyms to compile opinion words.

First of all the set of sentimental words with known positive and negative orientation is collected and then a sentimental strength or score will be is assigned each word using a probabilistic method. At the beginning, each positive seed word is given the score of +1, each negative seed is given the score of -1, and all other words are given the score of 0. The scores are revised during the propagation process. When the propagation stops after a number of iterations, the final scores after a logarithmic scaling are assigned to words as their degrees of being positive or negative.

3. RESULTS AND EVIDENCE

Managing requests and queries of a large amount of students in an educational institute is not an easy task. Virtual Student Advisor comes in to play to handle few major problems in this given context. The system provides a web and a mobile interface for users. Based on their role on the system, they have to use the respective interface. In order to function without any hassle, the system should meet an 80% of accuracy and a prominent threshold value in terms of efficiency.

The Inquiry Management Module consists of three main components as the Service Component, Web Component and the Mobile Component. Web and Mobile Components provides an interface for

the user to interact with the system while all the analysing and the querying techniques are implemented in the service component. Once the user interacts with the system, the input will be processed within the Service Component and the output will be delivered in JSON format as below.

Example Question:
How many consecutive attempts I can sit for repeat exams?


```
{
  "status":true,
  "answer":"{"numID":32,
        "varQuestion":"How many consecutive attempts I can sit for repeat exams?",
        "varAnswer":"Students are allowed to sit for a paper which they have failed for 3 consecutive attempts.",
        "numCategoryID":0,
        "varKeyPhrase":["sit","repeat","exams"]}"
}
```


The output consists of the Question, Answer, Key Phrases and the Category. Key Phrases represents the primary key words used to match the question with the existing knowledge base. The system produces a term vector map using the key words and the cosine value of each vector will be compared with the existing term vector values of the questions. If there are many matching questions, the system will pick the question containing the highest value. Table I shows a sample output of the values.

Table I: Comparison values for sample output

Question No	Cosine Value
32	1.0000002
33	0.8900034
43	0.5000440

The system will get Question Number 32 as the answer and returns it with the JSON output value.

As shown in Fig. IV in the priority based Appointment Scheduler Module, all the free time slots will be saved in the database. Depending on the availability and the priority of the request the respective users will be given a chance to have an appointment with the lecturer

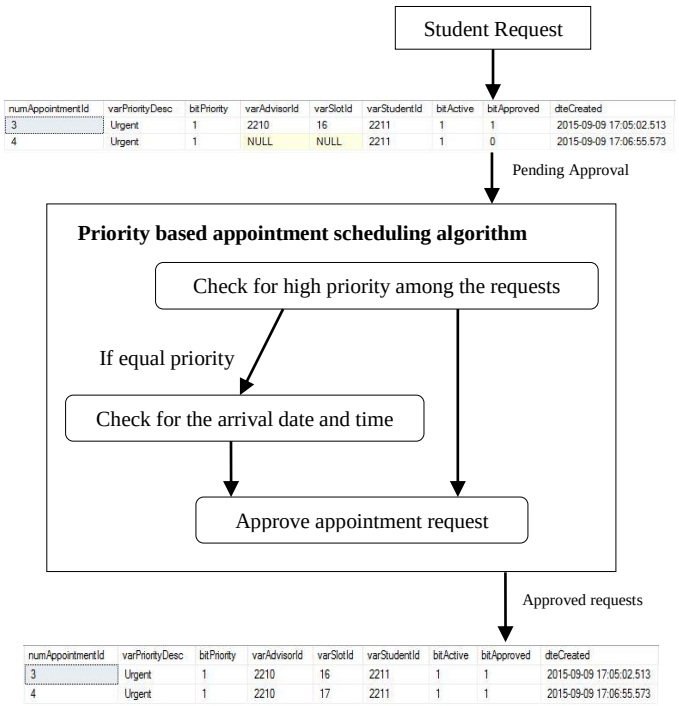

Fig IV: *Scheduling Process of a Request*

Feedback Analyser Module receives the student feedbacks in different ways. System presents the questions as a multiple/single choice and a text feedback will be collected as well. In the multiple choices module, the question represents a specific weight depending on the content. It's important that the user who prepares the question form should focus on this aspect. In the database the system will maintain the stats for each question as mentioned below in Table II.

	Type	Weight (1-5)	Answer #	Answer Weight (1-100)	Submitted Count	Score	Cum. Score
1	Multiple selection	3	1	100	30	28.5	28.5
			2	75	40		
			3	50	60		
			4	25	20		
			5	0	50		
2	Single Selection	5	1	100	50	56	42.25
			2	80	20		
			3	30	60		
			4	0	20		
3	Text	2	1	100	70 positive	18.66	30.45
4	Multiple selection	4	1	100	100	58.33	44.39
			2	50	80		
			3	0	60		

Table II: Sample score values for questions

The final score for each question will be calculated according to the formula (1) and (2) as stated.

Multiple selection type & Single selection

$$Question\ score = \frac{\sum_{k=1}^{n}(Answer\ weight * Submitted\ count)}{\sum_{k=1}^{n}(Submitted\ count)} * \frac{Question\ weight}{5} \quad (1)$$

Text

$$Question\ score = \frac{\sum_{k=1}^{n}(Answer\ weight * Submitted\ count)}{Total\ submitted\ feedback\ forms} * \frac{Question\ weight}{5} \quad (2)$$

Based on those score the user can generate different reports and graphs to examine the practical issues of the academic system.

4. CONCLUSION AND FUTURE WORK

Academic departments of any institution with large student population, face a vital challenge in managing their day to day operations. Mainly the department faces difficulties in handling of student queries, managing student-staff meetings and feedback analyzing process. Even though there are systems available to overcome these issues, they are not used as expected because of less effectiveness.

The Virtual Student Advisor system has a potential of overcoming the difficulties faced by the academic advisors in an effective manner.

For further enhancements the research work can be extended by adding techniques to optimize the answers much more. Organized information can be saved in a graph database which can be regularly updated and easy to query.

REFERENCES

[1] Raj, P. C. "Architecture of an Ontology-Based Domain-Specific Natural Language Question Answering System." *arXiv preprint arXiv:1311.3175* (2013).

[2] Burke, Robin D., et al. "Question answering from frequently asked question files: Experiences with the faq finder system." *AI magazine* 18.2 (1997): 57.

[3] Joachims, Thorsten. *Text categorization with support vector machines: Learning with many relevant features*. Springer Berlin Heidelberg, 1998.

[4] Kuo, Cheng-Ju, et al. "BIOADI: a machine learning approach to identifying abbreviations and definitions in biological literature." *BMC bioinformatics* 10.Suppl 15 (2009): S7.

[5] Li, Xiaobin, Stan Szpakowicz, and Stan Matwin. "A WordNet-based algorithm for word sense disambiguation." *IJCAI*. Vol. 95. 1995.

[6] Tong, Simon, and Daphne Koller. "Support vector machine active learning with applications to text classification." *The Journal of Machine Learning Research* 2 (2002): 45-66.

[7] C. Manning, 'Text-based Question Answering systems', 2015.

[8] R. Berwick, 'An Idiot's guide to Support Vector Machines (SVMs)', 2015.