An E-Learning Teaching Tool Based On Natural Language Processing

A.G.R.F Shereen

Faculty of Management & Commerce, South Eastern University Of Sri-Lanka

Abstract: The Growth of Technology has influenced on E-Learning through Natural Language Processing. It has completely changed the traditional way of learning with the rapid growth in the development of information retrieval, which makes an expressive and efficient way of quick and effective reach to the students over vast information through Natural Language Processing. This research represents to a Question Answering teaching tool which is limited for a particular area; number of relevant answers for the question is stored in the database by experts. Students have a facility to type the question in a given space and the system automatically generates the templates, and all the user-questions are taken as Natural Language queries as input. Here each and every template representing a question are pre stored in a database with its answer, just when the best matched template for the question is found, the corresponding answer will be returned to the end user. The Template matching technique has been used to match the most suitable answer. Major development starts with Frequently Asked Question with in-built as automated question answering process. This research analyze as how to understand a natural language question and how to represent the semantic content of an instructional text question as well as how to measure the relevance between a question and the answer for it in content –based retrieval. The main contribution is to give more suitable answer and its improved versions of answers for the question which typed and it is understood by natural language processing techniques. This research resolves in real environment, whenever they have a doubt in the relevant area.

Keywords: Template Matching, Natural Language Processing (NLP), Question Answering System, FAQ.

I.   INTRODUCTION

As one of the key driving forces in the 21st Century, Information Technology is changing the fundamental ways of learning. Nowadays students give more attention on technical aspects of learning more than moving their fingers on book by searching row by row. Developing an approach for using computers to answer user questions is a major interesting problem with the rapidly increased use of computers. This mechanism facilitates users to ask questions in natural language and give a concise and accurate answer. Understanding user questions in Natural Languages requires natural language processing.

This technique is the computerized approach of analyzing text based on both a set of theories and a set of technologies. It gives a pressure to be able to ask questions and obtain answers, using Natural Language (NL) expressions by automatically generated templates, rather than the keyword based retrieval mechanisms. The system can be better satisfy the needs of users as they will provide an accurate, quicker, convenient and effective way of giving answers to user question. Here this system mainly focus on the non-native English speakers, and achieved the target by using template matching technique, disemvoweling, and synonym matching mechanism. The Natural Language Processing technique developed for question retrieval does not analyze user queries. Instead analysis is applied for the database.

Using the knowledge base all the entries is stored for a limited closed area, so this facilitates the system more portable because it can be changed according to the problem domain. As well as its more and more evolving, if question answering ability improves as more entries are added.

There are different ways of answering the system, commonly we used the search engine to retrieve the answers, which covers a vast area which is not specified, and they are done by Indexed Based keyword matching technique. Other one is a closed specific area answering method, which gives most suitable answers by using Natural Language Processing (NLP).

This Research is organized as follows. In Section **II**, It describe about Background of the research. Section **III**. Literature survey. In Section **IV**, describe the System Architecture and Section **V** Template matching method. In Section **VI**, Testing in Section **VII** Conclude the research.

II.   BACKGROUND OF THE RESEARCH

This is an e-learning toolkit which helps a student to understand and clear the doubt of a particular defined area (Here this research describes of Student Experimental Processor (SEP)) and its Instruction Set Architecture to how to deal it with Assembly language. Student type the question relevant to this architecture and doubts related to it.

Through NLP techniques it understands the question and finally output the answer and its improved versions too. Here in contrast to Information Retrieval, which provides a list of potentially relevant documents in response to a user's query, question-answering provides the user with either just the text of the answer itself or answer- providing passages.

Question understanding evaluations are in efficient manner should be more accurate understandable to the exact point of view. Here it recognized the value of natural language processing (NLP) systems for general language extraction tasks, in order to measure performance and portability of the different NLP Techniques. Each question was associated with a specified task that consists of extracting particular information from where student type the question.

Though there are several techniques used in the Natural Language Processing here it used the efficient method of pattern matching algorithm in this research.

III.   LITERATURE SURVEY

In this Literature survey analyzed as similar systems of this research but not found the exact one. There are some systems that have performed to some extends of this system. Those systems are briefly described below.

Universal Indexing Frame:

Bruke and Kass developed a video indexing schemes based on "Universal Indexing Frame" to retrieve video clips presentation in a cased-based teaching environment. This index frame contained slots such as "Anomaly", "Theme", "Goal", "Plan" which explicitly indicated the points of the interest or anomalies in a video story.

Transportation Command personnel in planning for operation:

Johnson et al developed a Trans-ASK system that contained 21 hours of video detailing the experience of United States Transportation Command personnel in planning for operations such as Desert Shield and Desert Storm. The videos were segmented into a collection of video clips in which experts told "war stories" of their actual experiences.

 Natural Language approach to content –Based Video Indexing and Retrieval for Interactive E-Learning:

Instructional videos are extensively used in e- Learning (Web based distance learning) Here an interactive multimedia-based e-learning environment that enables users to interact with it to obtain knowledge in the form of logically segmented video clips. They propose a natural language approach to content – based video indexing and retrieval to identify appropriate video clips that can address users needs. The method Integrates natural language processing, named entity extraction, frame- based indexing and information retrieval techniques to explore knowledge - on- demand in a video – based interactive e-Learning environment.

Question and Answering (Q&A) system research received considerable attention from the research community through Text Retrieval Conference Q&A track since 1999. The original aim of the track is to systematically evaluate both academic and commercial Q&A systems. Maybury has discussed the characteristics of Q&A systems and resources needed to develop and evaluate such systems. Main approaches in Q&A systems could be found in which template based approach discussed in detail. Although, most Q&A systems are based on Web environments, SMS has also been used as an environment in contexts such as in learning and agriculture.

IV.   SYSTEM ARCHITECTURE

This section describes the architecture of this system. The complete system is sub modules into four modules. (A) Introduction module, (B) Templates module, (C) Question Answering module and (D) Download module. Each module is described in detail in the following subsections.

A. Introduction Module:

This contains a brief introduction of the Instruction Set Architecture and some detail explanation of the overall flow of implementation and how to use it. Hence it includes all kind of architectures and operations.

B. Templates Module:

To expand it and to obtain flexibility, whenever system overcomes a new question the administrator feels free to add a particular question and automatically the system generates the template, as well as answers, and to update the answer whenever it's required, but it should follow the standard syntax when developing the templates.

C. Question Answering Module:

Using generic methods answer can be generated. When it does the statistical techniques it checks with

- Word resemblance
- Distance similarity
- Sentence length similarity
- Word order similarity

This assigns values to reflect the similarity. The sum of these values is used for comparison. This approach works rather poorly for short questions and when the query and FAQ use different wording to carry the same meaning, here when we use the Template matching approach, the templates are matched against the question asked by users. The success of the question answering thus depends a lot on the quality of these templates.

D. Download Module:

This facilitates users to download the related ISA Architectures and the other useful documents.

Non- Functional Requirements:

- Accurate

By introducing system is more accurate for question answering according to User request. The administrator can add new questions and can update it more accurate.

- Availability

The application should be available 95% percent of the time for user access. The hours of use should relate to the working hours.

- Efficiency

This system is an efficient way to reduce waste time of Hardware failures.

- Reliability

Backups of the system details get daily to recover the Hardware, Software and Power Failures. Only authorized person can delete records from system. Therefore it is assured of the reliability of the system.

- Security

To increase the security of the system passwords are kept with the Administrator, and will act as the administrator for the system. This system is an efficient way to reduce waste time between program schedules
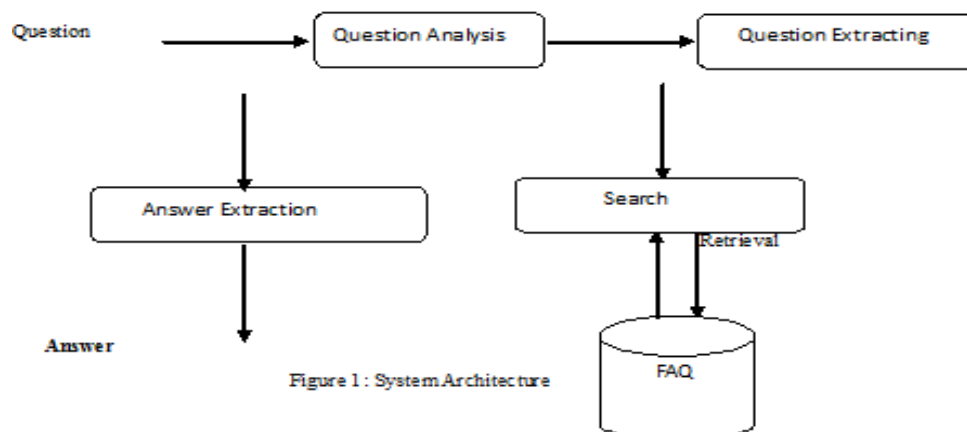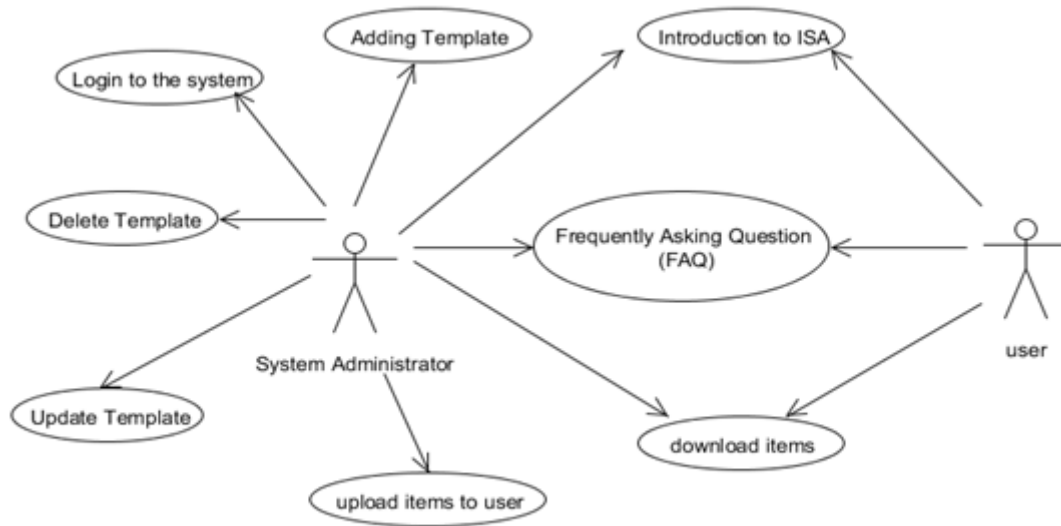

Figure 1 : System Architecture

Figure 1: Use-case diagram of the system

Table 1: Use case for Login to system

1	Use Case Name	Login to system
2	Use Case ID	ICS-01
3	Primary Business Actor	System Administrator
4	Other Participating Actors	System user
5	Brief Description	Authorized person can login to system.
6	Pre Conditions	Check authentication
7	Basic Flow	By entering correct user name and password admin can enter into the system.
8	Post Conditions	The system displays the main page.

Table 2 : Use case for Adding templates

1	Use Case Name	Adding templates
2	Use Case ID	ICS-02
3	Primary Business Actor	System Administrator
4	Brief Description	Authorized person can add templates.
5	Pre Conditions	Check authentication
6	Basic Flow	Login to system as administrator and click "Template" link.
7	Post Conditions	The system display page to enter to the add template interface and after confirm add template to the system.

Table 3 : Use case for delete templates

1	Use Case Name	Delete templates
2	Use Case ID	ICS-03
3	Primary Business Actor	System Administrator
4	Other Participating Actors	…………………….
5	Brief Description	Authorized person can delete templates.
6	Pre Conditions	Check authentication
7	Basic Flow	Login to system as administrator and click "Template" link.
8	Post Conditions	The system display page to enter to the delete template interface and after confirm delete template to the system.

Table 4 : Use case for update template

1	Use Case Name	Update templates
2	Use Case ID	ICS-04
3	Primary Business Actor	System Administrator
4	Other Participating Actors	…………………….
5	Brief Description	Authorized person can update templates.
6	Pre Conditions	Check authentication
7	Basic Flow	Login to system as administrator and click "Template" link.
8	Post Conditions	The system display page to enter to the update template interface and after confirm update template to the system.

Table 5 : Use case for uploading

1	Use Case Name	Upload Items
2	Use Case ID	ICS-05
3	Primary Business Actor	System Administrator
4	Brief Description	Authorized person can upload items
5	Pre Conditions	Check authentication
6	Basic Flow	Login to system as administrator and click "Download" link.
7	Post Conditions	The system display page to enter to the Download interface and after confirm upload to the system.

Table 6 : Use case for introduction module

1	Use Case Name	Introduction to ISA
2	Use Case ID	ICS-06
3	Primary Business Actor	User
4	Other Participating Actors	System Administrator
5	Brief Description	Users can go through for get a clear understanding of this ISA and admin can insert all notes.
6	Pre Conditions	Check authentication
7	Basic Flow	Login to system as user and click "Introduction" link.
8	Post Conditions	The system display page to enter to the Introduction interface

Table 7 : Use case for FAQ

1	Use Case Name	FAQ access
2	Use Case ID	ICS-07
3	Primary Business Actor	User
4	Other Participating Actors	System Administrator
5	Brief Description	Users can ask for an FAQ by typing the question.
6	Pre Conditions	Check authentication
7	Basic Flow	Login to system as user and click "FAQ" link.
8	Post Conditions	The system display page to enter to the FAQ interface

Table 8 : Use case for downloading

1	Use Case Name	Downloading items
2	Use Case ID	ICS-08
3	Primary Business Actor	User
4	Other Participating Actors	System Administrator
5	Brief Description	Users can download relevant materials through this interface.
6	Pre Conditions	Check authentication
7	Basic Flow	Login to system as user and click "Download" link.
8	Post Conditions	The system display page to enter to the Download interface

V. TEMPLATE MATCHING

This section, discuss the templates used and their syntax. This system is based on automatically generating templates for each Frequently Asked Question. Query interface is used to retrieve the question posted by the user .Questions those are stored in a database coupled with the answers. Next the query analyzer phrase the question into subject, verb, object ect. It is also used to improve the performance of the QA system. The templates are matched against the questions asked by users to find the best matched template. The success of the question answering thus depends a lot on the quality of these templates. Questions classification used to identify type of the question after that type of the answer will be specified.

The syntax of the templates is defined so that a single template could match many different variants of the same question. A question might be asked in different ways due to one or more of the following reasons: different tenses; singular/plural forms; usage of synonyms; the order of using words; and the use of optional words. In Table I, we tabulate the syntax used for the templates of the questions. Using the above syntax arbitrary complex templates can be constructed. Also phrases can be nested within each other, and synonym list could also contain phrases that have the same meaning as a single word.

Table 9 : Syntax used in the template

No.	Syntax	Description
1	Asterisk (*)	This symbol should follow at the end of a group of characters means that additional characters could follow. Used to handle stemming Eg: go*- going, goes,gone
2	Forward slash(/)	When words are separated by / either one of the words must match with user question.
3	Dollar ($)	A $ at the beginning of a terms specifies checking with the synonym list.
4	Square parenthesis ([ ])	Words group with [] denotes phases.
5	Hash (#)	Used only square parenthesis, terms separated by hash, should appear in the designated order.
6	Colon (:)	Used to separate terms, a question must contain all terms of a template in order to match.
7	Semi colon (;)	Used only within square parenthesis, Terms separated by a; should directly follow each other .
8	Synonym List	It's important to list synonyms for each term, since users will often ask the query using different terminology than the person who produces the FAQ.This list is referred when a '$' sign appears in a template. Eg: $ do = describe* make* evaluate*

Architecture of a QA system:

1. Query interface is used to retrieve the question posted by the user

2. Next the query analyzer phrase the question into subject, verb, object ect. It is also used to improve the performance of the QA system.

3. Questions classification used to identify type of the question after that type of the answer will be specified.

4. Query reformulation plays a major role in QA system because this component is used to find the correct answer to user question.

5. Based on Co-occurrence words and semantic relations existing in database ontology, answer type, and keywords which extracting in question processing module, system filter candidate answers collection.

6. As a result some answers which are not related with the asked question will be eliminated.

The answers with higher priority show for validation, if the user accepts the suggested answer which system presented as an exact answer.
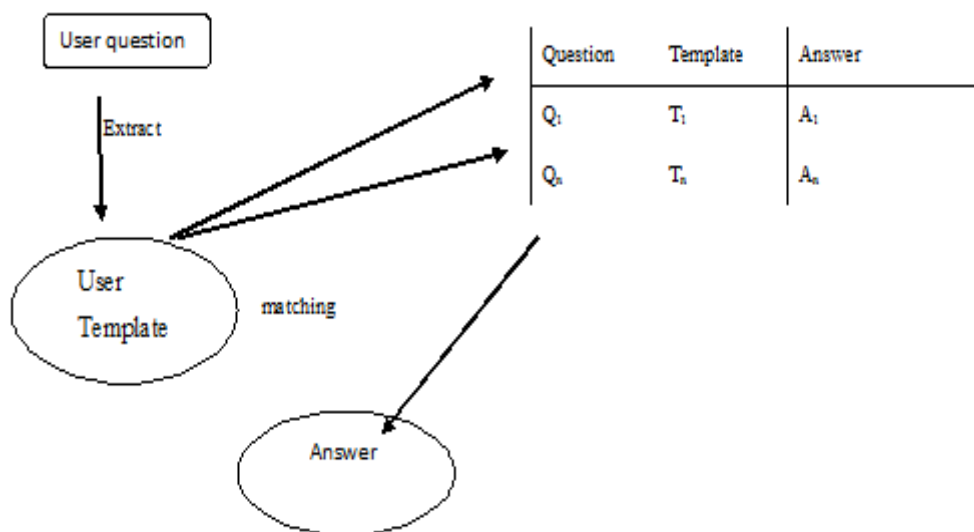

Figure 3: Answer retrieval process

In this answer retrieval process, questions are typed the system automatically generates the template, then the templates are matched with answers coupled in the Database. One question can be asked several ways, as well as one answer can gained by several questions.

TESTING:

Equation to approach the system:

$$\frac{\text{No of words matched with template}}{\text{No of meaningful words in a template}} * 100\ \%$$

Table 10: Summarized out put

	User Question	Percentage		Answer	
		Manually	programme	Expected	By programme
1	add two unknown values?	60%	80%	Temp1: A1	Temp1: A1
2	Add values immediate mode?	55%	75%	Temp1: A1	Temp1: A1
3	Add values immediate mode accumulator architecture?	75%	80%	Temp1: A1	Temp1: A1
4	Add values?	55%	59%	Temp1: A1	Temp1: A1
5	Add unknown values?	60%	65%	Temp1: A1	Temp1: A1
6	Add operations with modes?	50%	50%	Temp1: A1	Temp1: A1
7	Adding two unknown values with immediate mode using accumulator architecture	100%	93%	Temp1: A1	Temp1: A1
8	Add me and you?	25%	30%	Error or Default	Temp39: A39
9	Add A and B?	83%	80%	Temp2: A2	Temp2: A2
10	Subtract two unknown values with immediate mode using accumulator architecture	100%	97%	Temp31: B4	Temp31: B4

This Table consists with the probability of questions tested with the complete system.

Table 11: Percentage of key words matched

Percentage of keywords matched	No of answers	
	Correct	Incorrect
0-9%	4	2
10-29%	30	20
30-49%	25	15
50-69%	50	15
70-89%	50	9
90-100%	75	5
Total	234	66
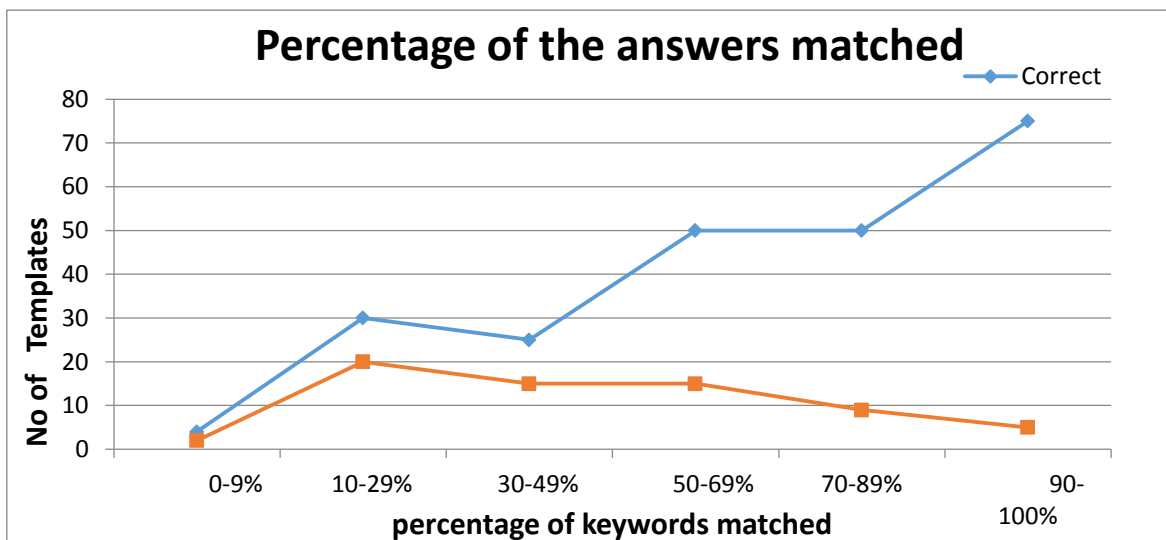
The results are with higher priority or probability is attempting the accurate answer which means more than 50%.

Below 50% Answers are rejected and below 30% probability is cannot identify the answer and its gives the Default answer.

VI.   CONCLUSION

In this research user type a question and its understanding process is done by natural language processing technique. If a question is typed incorrectly system should be able to classify the question and correction should be done automatically. The flow is done by following, Query interface is used to retrieve the question posted by the user, Next the query analyzer phrase the question into subject, verb, object etc. It is also used to improve the performance of the QA system. Questions classification used to identify type of the question after that type of the answer will be specified.

Query reformulation plays a major role in QA system because this component is used to find the correct answer to user question. Based on Co-occurrence words and semantic relations existing in database, answer type, and keywords which extracting in question processing module, system filter candidate answers collection.

As a result some answers which are not related with the asked question will be eliminated. The answers with higher priority show for validation, if the user accepts the suggested answer which system presented as an exact answer. Another facility whenever we overcome with a new question it automatically generates the template and answer also can be inserted.

One of the major challenges in these real time systems is understanding the question and accuracy, Flexibility and robustness and also most suitable for the task. When system is used in a dynamic background scenario or with the on time system where the answer appears on the screen, system fails to do its task perfectly, in such cases, complex searching algorithm is required rather than simple methods.

Page | 383

Effectively some more similar answers related to this type of question to needed. This proposed system deals with these challenges using technology in optimum level though these challenges could not be completely overcome.

REFERENCES

[1] Dongsonghang and Jay F. Nunamaker A Natural Language Approach to content –Based Video Indexing and Retrieval for   Interactive E-Learning. IEEE TRANSACTON MEDIA VOL. NO3 JUNE 2004

[2] S.R Balasundaram and B. Ramadoss "SMS for question Answering in the m-Learning   Scenario", Journal of Computer Science 3 (2): pp 119-121,2005

[3] A. Andrenucci and E. Sneiders, "Automated Question Answering": Review of the Main Approaches, in *International Conference on Information Technology and Applications.pp.514-519, 2005.*

[4] Fellbaum, Ch. (ed). 1998. WordNet: An Electronic Lexical Database. Cambridge: MIT Press.

[5] Wikipedia Contributors, "The SMS Language" [Online]. Available: http://en.wikipedia.org/wiki/SMS_language

[6] Text retrieval conference [Online]. Available: http://trec.nist.gov

[7] M.T. Maybury, 2002, "Toward a Question Answering Roadmap".

[8] S.R. Balasundaram and B. Ramadoss. "SMS for Question- Answering in the m-Learning Scenario", Journal of Computer Science 3 (2): pp. 119-121, 2007.

[9] M. Suktarachan, P. Rattnamanee and A. Kawtrakul, "The Development of a Question-Answering Services System for the Farmer through SMS: Query Analysis", in Proceedings of the 2009 Workshop on Knowledge and Reasoning for Answering Questions, pp. 3-10Suntec, Singapore, 2009.