Comparison of QUARS with Various Question Answering System

Lovely Sharma
Research Scholar, Punjab Technical University, Kapurthala Road, Jalandhar, Punjab, India.

Vijay Dhir
St. Soldier Institute of Engineering & Technology, Jalandhar, Punjab, India.

Abstract – The search on web is very time consuming process because search engines returns the long list of lengthy documents. The User has to search the web links one by one manually. Sometimes, users get tired to search the relevant content from web links. Thus, there is need to develop QA system to provide the succinct answers to users in a short period of time. Therefore, QUARS is developed to provide the answers related to Indian Railway domain in the Hindi language. The results are satisfactory and compared with existing Restricted QA systems and various methods are developed by researchers to provide the accurate answers

Index Terms – Restricted Domain Question-Answering System, Key-word based approach, Partial semantic Analysis.

1. INTRODUCTION

Recently, Internet has grown very fast and creates a place in everyone's Pocket. Because of this, most of the users rely on the web for finding out the relevant information but search engine provides the lengthy documents, web links etc. and it is user's responsibility to find out the information from the web is equal to find a needle in a haystack which is very time consuming process.

Thus, there is need to develop Question answering dialogue system which particularly focus on providing succinct answers to arbitrary questions. This method is intuitive and save lots of time. In addition to this, QA dialogue system could be more helpful in Railway sector where people do not want to stand in a queue and want to access the information while sitting at home i.e. anywhere, anytime information.

Therefore, QUestion Answering Railway System (QUARS) is developed to provide information about Railway in the Hindi Language. It provides only the precise and succinct answers to users instead of lengthy documents. This system is implemented on a relatively restricted domain as mentioned below:

1. List of trains between two stations, but constraint is to input only those trains which have stoppage at Jalandhar.
2. User can ask about fare of that particular train as mentioned in step (1).

There are approximately 250 trains having stoppage at Jalandhar. System is considering only Mail/Express/Superfast, DMUs, Passenger trains, Shatabdi, Garibrath and Jan-Shatabdi.

Thus, the system consists of many steps (cf. Figure 1.) such as: Query Processing, Query Frame Analyser, Dialogue Manager, SQL Query Generation and Response Generation. In this section, we give the brief introduction to QUARS. In the rest of the Sections, we provide Literature Review of QA System (cf. Section 2), Design and Implementation (cf. Section 3), Evaluation and Results (cf. Section 4), Comparison with other Systems (cf. Section 5) and Conclusion (cf. Section 6).

Motivation

The motivation behind to implement QUARS is to motivate the Hindi Language and its users on the web and provide the accurate answers to users about Railways. People speak various forms of languages as estimated near about 4000. Among top 100 languages, Hindi occupies the fifth position in the world.

2. REVIEW OF LITERATURE SURVEY

The early systems are very restricted to their nature such as Green et al (1961) conducted a study on BASEBALL2, a program to provide answers about American league over one season. The classical work has been done by Eliza3. It works as chatter bots and it naturally communicates between human and computer. The best-known early program with restricted domain is the LUNAR system. This system provides the easy way to access, compare and assess the chemical analysis data of Lunar Rock4. Initial dialogue systems such as UC (Unix Consultant) 5 and The Berkeley UNIX Consultant project (UC) provide the reply of user's query about UNIX operating system6. SHRDLU 7 is perhaps the first AI System to accept the user query and send a reply to the user. Another well remembered system in this tradition is START1.

3. DESIGN AND IMPLEMENTATION OF QUARS

The various modules of QA architecture are Input decoder (cf. Section 3.1), Query Processing (cf. Section 3.2), Query Frame Analyzer (cf. Section 3.3), Dialogue Manager (DM) (cf. Section 3.4), Domain Specific Component (SQL Generation) (cf. Section 3.5) and Response Generation (cf. Section 3.6). These systems are described in subsequent sections. The foremost concern of QUARS is to design the Railway database and the Knowledge base is as follows:-

a) Railway Database: It provides the information about trains between two stations and fare of related trains. The tables are Fare, Framing, Route and Railways.

b) Knowledge base: It stores the knowledge related to language understanding. The related tables are Charcode, Identitification and Questions.

Actually, it is a long pipeline of various modules to intercommunicate with each other to provide quick and brief answers in a short period of time.

First of all, Dialogue System accepts the Query posed by user in Hindi language using Hindi Transliteration[1].The next stage is Query processing module which analyse the real text and search for the keyword in the query. These accepted Keywords will match against the database, Query frame analyser fills the slots, appropriate query is generated and answer is provided to the User in a simplest manner.

The diagram is of QUARS as shown below in Figure 1:

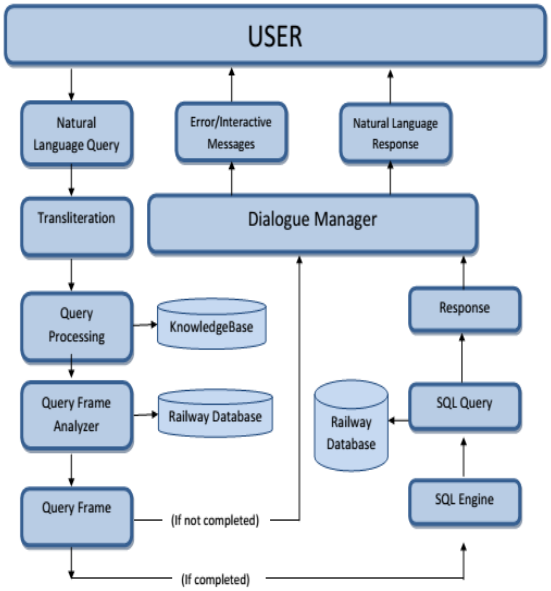

Figure 1.Detailed architecture of QUARS

3.1 Natural Language Query/Input Decoder

The first step receives the query from the user and Transliteration transliterates English into Hindi language.

3.2 Query Processing

The processing of Natural language is a most complex task to be built. The linguistic processing undergoes three levels according to QA architecture such as: (i) Tokenization (cf. Section 3.2.1) (ii) Normalization (cf. Section 3.2.2) and (iii) Keyword Selection Module (cf. Section 3.2.3) as shown in the following Figure 2.

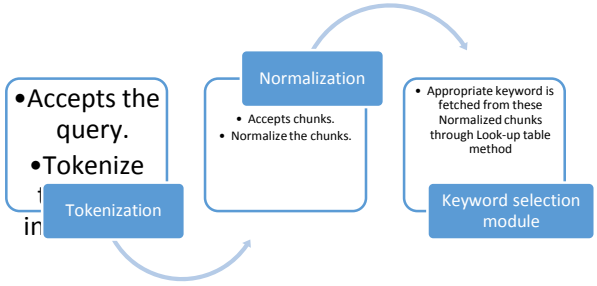

Figure 2. Query Processing Module

3.2.1 Tokenization

Tokenization involves text that needs to be segmented into linguistic units [16].

3.2.2 Normalization

The Non Standard Words (NSW) might be various dialects, also the impact of native languages, variation in English words and abbreviations leads to variation of the similar words [17]. All these phonetic variations or NSW representations must typically be normalized.

3.2.3 Keyword Selection Module

This module very much influences the accuracy of the system. In the previous stage (cf. Section 3.2.2), chunks are normalized and before the acceptance of these chunks, each chunk is compared with Knowledgebase through look-up table method until a word gets matched and that particular matched word is transformed into keyword. This keyword will help the system to find the purpose of the Natural language query.

3.3 Query Frame Analyser

The precision of the system is greatly influenced by the selection of the Query Frame. At this stage, keywords are recognized and accepted by the system such as system

identifies किराया (Fare) keyword as explained in the previous section (cf. section 3.2.3) and this keyword will be mapped against Railway Database again through look-up table method to find the appropriate Query Frame[18]. Each Natural Language Query has an individual Query frame.

On the basis of keywords the appropriate Query Frame is chosen and each Query Frame has slot-filling format. At this stage, two cases arise. A). Slot is complete. B). Slot is not complete.

If the slot is complete then corresponding SQL query will be triggered on the Railway database and succinct answer will provide to the User in a simplest manner and if slot-filling is incomplete, then system will enter into dialogue with the User with the help of Dialogue Manager and system is supposed to ask the queries from the user in accordance to fill the slots.

3.4 Dialogue Manager

Sometimes, Slot-filling remains unfinished due to insufficient information given by the User. So, system will seek for more information. At this moment, DM enters into dialogue with the user and sends interactive messages to the user to know about the missing information as explained in Table 2. Thus, Dialogue Manager (DM) facilitates the interaction between the user and the system[19]. The flow of the dialogue is controlled by the DM by deciding how the user and the system will respond to each other.

3.5 SQL Generation

In this module, SQL Query generation is based on the completion of Query frame as explained in Section 4.3. After the completion of Query frame the appropriate SQL Query is generated.

3.6 Answer Generation

As discussed in the previous section, SQL Query is generated, triggered on the Database and appropriate response is redirected to the DM and display to the user. According to the foregoing example, the reply for the query is 120 Rupees.

4 EVALUATION AND RESULTS

Restricted domain systems tend to be situated not only within a specific domain, but also within a certain user community and within a specific task domain. The certain user community who tested the QUARS are based on knowledge of computer, knowledge of Railway, knowledge of the Hindi Language and knowledge of the English Language.

The actual evaluation of success of any QA system is based on the satisfaction level of the User. As much as the User will satisfy, the success rate of QA system will raise too. A generic evaluation is neither sufficient nor suitable for a restricted domain system, but widely used parameters are Precision, Recall. The various evaluation parameters of QUARS are as follows:

4.1 Evaluation of Transliteration

Google Transliteration IME offers several features focused on an improved user experience, including offline support, word completion, personalized choices, easy-to-use keyboard, quick search and several cool customization options. As a user types, a suggestion menu is displayed with alternatives and word completions. For example, as we type "**जालंधर**" we will see five options as "**जालंधर, जलंधर, जालन्धर, जुलंधर, जल्न्दहर**"from we can select the correct one.

So, the accuracy of the Google Transliteration IME is quite high because of the features provided by Google Corporation as mentioned above. The evaluation result is shown in Table1 below:

Total users	70
Total sets of Queries	120
Total Queries (Each set of dialogue consisted of around 1 to 2 NL Queries)	**34**2
Total words	5381
Wrong words	4
Correct Words	5377
Accuracy of Transliteration	99.9%

Table 1. Evaluation of Transliteration

Various users enter the 5381 words out of which 5377 words are correct where rest of 4 words are not correctly entered by user because of various reasons such as various types of users, abbreviations of English Language, variations in English words, variations in Hindi words etc. So, the accuracy of this module is 99.9**%** but the accuracy of **T**ransliteration module does not affect the accuracy of whole system because NSW representation is converted into Normalized form.

4.2 Evaluation of Keyword Selection Module

The most critical and burdensome module of the system is to select the keyword from the User Query. The system is based on key based approach and performs partial semantic and syntactic analysis of the Query but sometime task is petty ambiguous while finding out the keyword from the User query because of variations of the words etc.

So, all the variations of the word are already stored in charcode Table to provide the accurate answer to the User and system is completely capable to recognize these words. The correctness

of this module is dependent upon the identification of keyword from the User Query. Thus, the various people tested the QA system as shown in Table 2:

Total users	70
Total sets of Queries	120
Total Queries (Each set of dialogue consisted of around 1 to 2 NL Queries)	342
Queries containing Keywords	120
Correct responses	118
Accuracy	(118/200)*100=98.3%

Table 2. Evaluation of Keyword Selection Module

The rest of 1.7 % accuracy is less because of **में (**Main) as Hindi word.

4.3 System performance

QUARS is evaluated by the various people. I demonstrate the constraints of queries and considering two measures for evaluating the system with and without dialogue management. Those are Precision and Recall.

Out of 342 queries, system has generated output for 340 queries. For the remaining 2 queries, system was unable to generate the exact output because of Hindi word **में (**Main).

Precision = (Number of correct outputs generated by the system / Number of outputs generated by the system) * 100.

Precision= (338/340)*100 = 99.4%

Out of 342 outputs, 338 are identified as correct output. For the remaining 4 queries, system was unable to generate the exact output because of restricted domain and Hindi word **में (**Main).

Recall = (Number of correct outputs generated by the system / Number of Natural Language Queries given to the system) * 100.

Recall= (338/342)*100 = 98.8%

Number of outputs generated	340
Number of correct output generated	338
Number of natural language queries given to the system	342
Precision	(338/340)*100 = 99.4%
Recall	(338/342)*100 = 98.8%

Table 3. System Performance Result

5 COMPARATIVE STUDY WITH VARIOUS QA SYSTEMS OF QUARS

The comparisons of various techniques are based on technique, result and limitations. The below is the detailed comparison between various techniques:

a). Based on Technique:

The selected QA systems for comparison have developed various methods whether it is their own or in-built and also provides satisfactory results with some merits and demerits. Researchers always choose appropriate method to provide accurate results to the User. The following is the comparison table between various methods as follows:

S. No.	Paper Title	Method	Ref. No.
	QUARS	Keyword based approach	
1	A Practical QA System in Restricted Domains	QA system with IE engine	[20]
2	Prashnottar : A Hindi question Answering system	Hindi Shallow Parser using Query logic language (QLL)	[21]
3	Web Based Application for Hindi Question Answering System	Question Classification	[22]

4	Algorithm for Punjabi Question Answering System	Develop step algorithm and question classification	[23]
5	Intelligent query processing in Malayalam	Key-based approach, Morphological analysis and shallow parser	[24]
6	Tamil Speech Recognizer Using Hidden Markov Model for Question Answering System of Railways	Hidden Markov Model	[25]

Chung et al. and Gupta et al. implemented the QA system with the help of Information Extraction (IE) and Information retrieval (IR) respectively while Sahu et al. and Sukumar et al. follows the shallow parser and morphological analysis on the query processing. The another system is purposed by Stalin et al. which is based on Question classification and vignesh et al. uses the Hidden Markov Model (HMM) to analyse the Query posed by the User.

The good results of the implemented QA system are always based on the adopted method by the Researcher. In comparison with QUARS, the above mentioned all the methods also provide the satisfactory results but not the better results. The QUARS implements the keyword based approach using the Look-up table method and partial semantic analysis of the query posted by the User in Hindi language for the domain of Indian railways.

The technique implemented on QUARS is much easier and simple as compare to above mentioned methods. The result provided by QUARS is also better than the result provided by these methods. As we already discussed that proficiency of the system is always dependent on the adopted technique for the QA system.

b). Based on Result:

The good result of every QA system is based on the long pipeline of various modules of the QA system. The output of the each module is input for the next module. The various methods as discussed provides the satisfactory results but these results must be improved by overcoming the limitations of the every QA system. The following table shows the result chart of all these methods as below:

S. No.	Paper Title	Results	Ref. No.
	QUARS	Precision =99.1% Recall = 98.4%	
1.	A Practical QA System in Restricted Domains	Precision =90.9% Recall = 75.0%	[20]
2.	Prashnottar : A Hindi question Answering system	Accuracy=68%	[21]
3.	Web Based Application for Hindi Question Answering System	Inconclusive request	[22]
4.	Algorithm for Punjabi Question Answering System	Average accuracy=73%	[23]
5.	Intelligent query processing in Malayalam	Precision=87.50% Recall=80% (without dialogue management)	[24]
6.	Tamil Speech Recognizer Using Hidden Markov Model for Question Answering System of Railways	Accuracy=85%	[25]

Table 5. Comparison based on Result

In comparison with QUARS, the accuracy of these systems is very low because of their own limitations. The QUARS provide the Precision and Recall rate as 99.8% and 98.4% respectively. The result is much higher than other systems because of adopted technique as keyword based approach. We

can conclude that keyword based approach is better than another methods because it does not support question classification, syntactic and semantic analysis, morphological analysis and if we follow these methods then task of Query processing becomes more complex as a result these techniques generates low results.

c). Based on Limitations:

Every QA system has their own limitations. Chung et al. and Sukumar et al. have similar limitations such as insufficient data of domain in Database (DB).

Chung et al. also has some more limitations i.e. error due to Query Frame Analysis and another error was caused by the flaw of keyword based Query Frame Decision approach while QUARS achieve the 100% accuracy to extract the correct keyword from the user's query and further QUARS generates the correct query frame through look up table method.

Sahu et al., Stalin et al. and Gupta et al. has identical problems i.e. all of these techniques follows question classification due to this some type of questions are not able to identify by these QA systems. In addition to this, another system implemented by vignesh et al. has Word Error Rate (WER) for some restricted words.

The QUARS has also limitation of some restricted words but it also overcome the various limitations of above mentioned techniques such as it has sufficient data related to domain in DB and does not support question classification ,so provides better results.

S. No.	Paper Title	Limitations	Ref. No.
	QUARS	Sometime keyword not found	
1.	A Practical QA System in Restricted Domains	1. Insufficient data of weather-domain ontology 2. Error due to query frame analysis 3. Low recall rate due to invalid date and topic.	[20]
2.	Prashnottar : A Hindi question Answering system	Accuracy of Question type 'Where' is low because the answer type of this question is Proper noun i.e. Location.	[21]
3.	Web Based Application for Hindi Question Answering System	Less semantic implementation of answer extraction	[22]
4.	Algorithm for Punjabi Question Answering System	Do not support which, how much, how many type of questions.	[23]
5.	Intelligent query processing in Malayalam	Very restricted in nature	[24]
6.	Tamil Speech Recognizer Using Hidden Markov Model for Question Answering System of Railways	WER for some restricted words	[25]

Table 6. Comparison based on Limitations

6. CONCLUSION

QUARS system is a keyword based system, employed for an Indian Railways System in Hindi language for Northern India. The system is extremely involved in Query Processing Module. To implement this, the detailed analysis of input query has been done with the help of Tokenization (cf. Section 3.2.1), Normalization (cf. Section 3.2.2) and Keyword Selection Module (cf. Section 3.2.3). The generated keywords (cf. Section 3.2.3) help us to select appropriate Query frame and in the next step, SQL Query is generated with the help of these Query frames. According to above comparison, QUARS provides the better results with the keyword based approach as compare to other methods which support in-built methods or their own techniques.

Sr. No.	Title	Accuracy	QA type	Model/Technique Used for Query processing	Limitations	References
	QUARS	Precision =99.1% Recall = 99.1%	Restricted	Keyword based approach	Sometime keyword not found	
1.	A Practical QA System in Restricted Domains	Precision =90.9% Recall = 75.0%	IR/IE	QA system with IE engine	1. Insufficient data of weather-domain ontology 2. Error due to query frame analysis 3. Low recall rate due to invalid date and topic.	[20]
2.	Prashnottar : A Hindi question Answering system	Accuracy=68%	Restricted	Hindi Shallow Parser using Query logic language (QLL)	Accuracy of Question type 'Where' is low because the answer type of this question is Proper noun i.e. Location.	[21]
3.	Web Based Application for Hindi Question Answering System	Inconclusive request	Restricted	Question Classification	Less semantic implementation of answer extraction	[22]
4.	Algorithm for Punjabi Question Answering System	Average accuracy=73%	Restricted	Develop step algorithm and question classification	Do not support which, how much, how many type of questions.	[23]
5.	Intelligent query processing in Malayalam	Precision=87.50% Recall=80% (without dialogue management)	Restricted	Key-based approach, Morphological analysis and shallow parser	Very restricted in nature	[24]
6.	Tamil Speech Recognizer Using Hidden Markov Model for Question Answering System of Railways	Accuracy=85%	Restricted	Hidden Markov Model	WER for some restricted words	[25]

Table 7. Comparison of QUARS with other Methods

REFERENCES

[1] Winograd T. *Understanding natural language*. Cognitive psychology, 1972, 3(1), pp. 1-191.

[2] Green BF, Wolf AK, Chomsky C, Laughery K. *Baseball: an automatic question-answerer*. In Proceedings of IRE-AIEE-ACM '61 (Western) Papers presented at the May 1961, western joint IRE-AIEE-ACM computer conference, 1963, pp. 219-24.

[3] Weizenbaum J, *ELIZA—a computer program for the study of natural language communication between man and machine*, Communications of the ACM, 1966, 9(1), pp. 36-45.

[4] Woods WA. *Progress in natural language understanding -an application to lunar geology*. In proceedings of the national computer conference and exposition, 1973, pp. 441- 50.

[5] Chin DN. *Knowledge Structures in UC, the UNIX* Consultant*. In Proceedings of the Twenty-first American for Computational Linguistics, 1983, pp. 159-63.

[6] Wilensky R, Chin DN, Luria M, Martin J, Mayfield J, Wu D. *The Berkeley UNIX consultant project*. Journal of Computational Linguistics, 1988, 14(4), pp. 35-84.

[7] Ward N, *SHRDLU* , In Encyclopaedia of Cognitive Science, Nature Publishing Group, 2002.

[8] Riloff E, Thelen M, *A Rule-based Question Answering System for Reading comprehension tests.* In Proceedings of the 2000 ANLP/NAACL Workshop on Reading comprehension tests as evaluation for computer-based language understanding systems, 2000, Volume 6, pp. 13-19. Association for Computational Linguistics.

[9] Denecke M and Tsukada H. *Instance-Based Generation for Interactive Restricted-Domain Question-Answering Systems.* In Natural Language Processing–*IJCNLP*, Springer Berlin Heidelberg, 2005, pp. 486-97.

[10] Sravanthi MC, Prathyusha K ,Mamidi R. *A Dialogue System for Telugu, a Resource-Poor Language.* Computational Linguistics and Intelligent Text Processing, Springer International Publishing: Switerzerland, 2015, pp. 364-74.

[11] Vignesh G, Ganesh SS. *Tamil Speech Recognizer Using Hidden Markov Model for Question Answering System of Railways.* Advances in Intelligent Systems and Computing. Springer –Verlag: Berl\in Heidelberg, 2015, pp. 855- 62.

[12] Dubey P. *Testing and Results of Hindi-Dogri Machine Translation System.* Indian Journal of Science and Technology, 2015, 8(27), pp. 1-8.

[13] Puri R, Bedi RPS, Goyal V. *Punjabi Stemmer Using Punjabi WordNet Database Punjabi.* Indian Journal of Science and Technology, 2015, 8(27), pp. 1-5.

[14] Kulkarni A, Srivatsa BR, Baji C, *Hindi-Kannada Named Entity Transliteration:*
Issues and Possible Solutions. Indian Journal of Science and Technology, 2015, 8(27), pp. 1-5.

[15] Belinkov Y, Cedeno A, Mubarak H. *Answer Selection in Arabic Community Question Answering: A Feature-Rich Approach.* In Proceedings of the Second Workshop on Arabic Natural Language Processing, Beijing, China, 2015, pp. 183-90.

[16] Webster JJ, Kit C. *Tokenization as the initial phase in NLP*. In Proceedings of the 14th conference on Computational linguistics, Nantes, France, 1992, vol. 4, pp. 1106-10.

[17] Sproat R, Black AW, Chen S., Kumar S, Ostendorf M, Richards C. *Normalization of non-standard words.* Computer Speech and Language. 2001, 15(3), pp. 287–33.

[18] Sinha RMK. *On design of a question-answering interface for hindi in a restricted domain.* In International Conference on Artificial Intelligence, Las Vegas, 2006, pp. *319-24*.

[19] Reddy RRN, Bandyopadhyay S. *Dialogue based Question Answering System in Telugu.* In Proceedings of the Workshop on Multilingual Question Answering, 2006, pp. 53-60.

[20] Chung H, Song Y, Han K, Yoon D, Lee J, Rim H. *A Practical QA System in Restricted Domains*, Workshop on Question Answering in Restricted Domains, 42ndAnnual Meeting of the Association for Computational Linguistics (ACL-2004), Barcelona, Spain, , 2004, pp 39–45.

[21] Sahu S, Vashnik N and Roy D. *Prashnottar: A Hindi Question Answering System*, International Journal of Computer Science and Information Technology, 2012,4(2), pp.149-158.

[22] Stalin S, Pandey R and Barskar R, *Web based Application for Hindi Question Answering System*, International Journal of Electronics and Computer Science Engineering, 2012,vol. 2, pp. 72-78.

[23] Gupta P, and Gupta V, *Algorithm for Punjabi Question Answering System*, International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE), 2013, vol.3, pp. 902-909.

[24] SA Raji, Anto PB. *Intelligent Query Processing In Malayalam*, International Journal on Computational Sciences & Applications (IJCSA),2013, 3(2).

[25] Vignesh G and Ganesh SS, *Tamil Speech Recognizer Using Hidden Markov Model for Question Answering System of Railways*, Artificial Intelligence and Evolutionary Algorithms in Engineering Systems, 2014, vol. 325, pp. 855-862.