International Journal of Management, IT & Engineering
Vol. 7 Issue 4, April 2017,
ISSN: 2249-0558 Impact Factor: 7.119
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

QUESTION ANSWERING SYSTEM WITH NATURAL LANGUAGE INTERFACE TO DATABASE

Dr. M.Humera Khanam*

S.Venkata Subbareddy*

Abstract:

Question Answering (QA) is an area of natural language processing research aimed at providing human users with a convenient and natural interface for accessing information. Nowadays, the need to develop accurate systems gains more importance due to available structured knowledge-bases and the continuous demand to access information rapidly and efficiently. The need to store data in an organized manner so that searching, retrieving and maintaining of data becomes easier. To efficiently operate these database, knowledge of Structures Query Language (SQL)becomes essential. But the usage of SQL restricts the access to databases from the users who don't have the knowledge of them. A need for interface comes into the picture to enable the access of these databases even to a non-expert users. This paper describes the design to develop Telugu language Question Answering system to database.

This paper describes about question answering system using Natural Language Interface to a database. Here we use the rule based algorithm for train the systems question classifier to achieve a high accuracy ratio.

Keywords —Natural Language Processing (NLP), Natural Language Interface To Database (NLIDB), Question Answering System(QAS), Structured Query Language(SQL).

* **Dept. of Computer Science and Engineering,SVU College of Engineering,Tirupati,India**

INTRODUCTION

Information plays a major role in our daily life. Database is the major source for data storage. SQL is the widely used database language to retrieve data from database. Hence everyone is not familiar with the usage of SQL. So that drawback makes the researchers to turned out to use natural language (NL), i.e. English, Telugu, Hindi, etc. One can express their ideas and emotions better by using natural language instead of artificial language like C, C++, and JAVA. NLIDB System is solution to this problem which is concerned with the interaction between human languages and the machine. This system allows any type of users mainly inexperienced (or) illiterate ones to retrieve data from database in a simple way. The Question Answering system about a regional database in Telugu has been described. This system uses rule based matching technique to convert the Natural Language Query in Telugu to SQL.

QA systems are complex systems that, given a question asked in natural language, can find an answer to this question, in a corpus or in the Web, and justify it by quoting their source(s). From the user's point of view, they can be considered as an improvement over traditional search engines such as Google or AltaVista because they provide a more direct and precise access to the desired information. The counterpart is that finding the correct answer to a question requires much more analysis and processing than a typical search engine.

APPROACHES IN QUESTION ANSWERING SYSTEM

Various approaches are used for Question Answering system are Rule based approaches, Machine Learning techniques or Statistical approaches. Both methods can be combined to yield best results.

a. Rule based approach

Rule based approach requires hand written rules which requires knowledge on specific language. In this approach rules are used to identify exact question what was given query. QA system uses gazetteer to classify tokens. In this approach some language based rules and other heuristic are used to classify words. It needs rich and expressive rules and gives good results.

b. Machine learning based approach

Machine learning techniques uses large amount of annotated data to train the model. Several Machine Learning techniques include Conditional random fields and Support Vector Machines. This approach explores the study and algorithms that can learn from and make predictions on data. This approach is used to build a model from example inputs in order to make predictions or decisions.

c. Hybrid approach

In Hybrid approach both rule based approach and machine learning approach is used to improve accuracy of a model. Some times more than one machine learning approaches are used in a model in order to improve accuracy. For example Support Vector Machine model can be used to design a model.

DESIGN CHALLENGES

A group of researchers wrote a detailed roadmap of research in question answering, identifying the issues and challenges in building a Q&A systems.

1. Question classes: Different types of questions require different strategies to find an appropriate answer.

2. Question Processing: There are various ways(Interrogative, assertive) to present a question with the same information request. This creates a problem of being understood as two different questions. A semantic model would recognize similar questions, regardless of how they are presented.

3. Context and Q&A: Questions are usually asked within a specific context and answers accordingly.To resolve ambiguities in question, context can be used by the Q&A systems.

4. Data sources for Q&A: It must be known beforehand, what knowledge sources are available and are relevant to the question. If the knowledge base / data sources, doesn't contain the answer to a question, no matter how well programmed the system is, a correct result is difficult to obtain.

5. Answer Extraction*:* Answer extraction depends upon the complexity of the question, on the answer type provided by question processing, on the actual data where the answer is searched, on the search method and also on the question focus and context.

PROPOSED SYSTEM

We proposed Rule Based technique for Telugu language interface (TLI) system. In this system, we will map all keywords in the user query to the database. If the keyword matches, then the corresponding SQL query is generated and required answer will be retrieved from the database. The main advantage of the system is if the input is ambiguous, the system will manage to give reasonable output based on the keywords in the query.

At first, user gives Telugu language query which is then divided into a set of tokens by using whitespace as delimiter in query analyzer.. Each token is then searched in the knowledgebase, if a token is found in knowledgebase, its information is stored in memory as <key,values> pair. Otherwise it is simply discarded assuming that it does not provide any useful information in deciding the query frame. A natural language query equaling the user requested query is generated from the <key,values> stored in memory and a conformation is requested from the user asking whether the generated question is same as the one which user is expecting to be. If the user gives negative acknowledgement, then alternative natural language query is generated until the user gives positive acknowledgement or there are alternative queries that can be generated. If there are no alternative queries then the system aborts the user request and informs the user to ask the same question in a different manner so that there may be a possibility that user can get the answer. If the user gives positive conformation then the system can decide on the query frame and can transform the given natural language query into a set of SQL queries. These SQL queries are executed over the database and the retrieved data set is transformed into Telugu natural language sentences using a template based approach and is forwarded to the user as the answer.

In this model, we define some rules to create SQL queries for corresponding Telugu query. For example,

తిరుమలఎక్కడఉంది(tirumala ekkaDa undi?)

This query is tokenized.

తిరుమల(tirumala) |ఎక్కడ (ekkaDa)|ఉంది(undi?)

T1 T2 T3

Each token is compared with keywords in lookup table.

Rule to create SQL query:

If T_i is found in lookup table

SELECT [COLUMN NAME] FROM [TABLE NAME] WHERE [COLUMN NAME]=T_i

ARCHITECTURE

Question Answering System, The below fig shows the whole process of the Question Answering System with Natural Language Interface to Database. In this user can write the query in the form of telugu language, after that the query spilt into tokens ofter that those are matched with knowledge base or lookup table what I was created. If match to lookup table or knowledge base it will generate SQL query. That SQL query retrieve the exact answer for the given question.

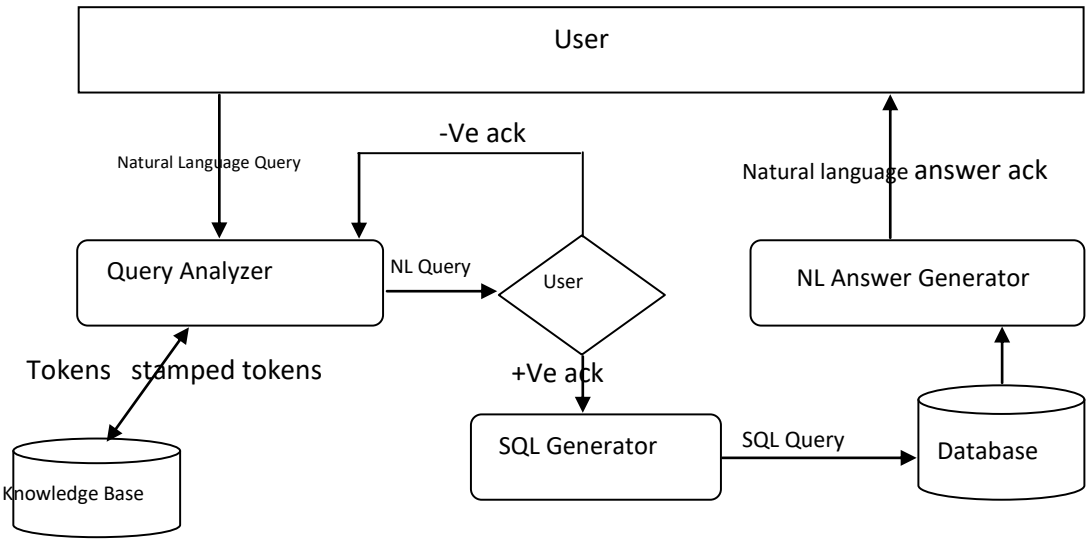

Fig: Architecture for QA System

To identify question is the major task in the QA System. Because one question ask in many ways, but each question has same answer. So, identify of question is the important task in this Question Answering System interface to database. This paper discussing about natural language as Telugu. Telugu language not has a proper capitalization and more ambiguity compare to English language. Here identify the questions based on the keywords in that sentence or question. Following example shows the how control ambiguity in Telugu sentence.

Q1 అలిపిరివెకడఉంది

(alipiri yekada undi?)

Ans: అలిపిరితిరుపతిలోఉంది.

(alipiri tirupati lO undi)

Q2 కడపగురించితెలపకండి

(Kadapa gurinchi telapankandi)

Ans: it doesn't show anything because we ask not showing about kadapa.

Q3.తిరుపతిగురించితెలపండి?

(tirupati gurinchi telapanDi?)

Ans: It showing description about tirupati, what was store in database about tirupati

తిరుపతిచిత్తూరుజిల్లాలోనిఒకనగరంఆంధ్రప్రదేశ్రాష్ట్రంలోఉంది .

ఇదిఆంధ్రప్రదేశ్లోతొమ్మిదవఅత్యధికజనాభాకలిగిననగరం .

ఇదిఒకమున్సిపల్కార్పొరేషన్మరియుప్రధానకార్యాలయాన్నితిరుపతిలోకలిగి

ఉంది .

తిరుపతిపవిత్రమైనఒకటిగాపరిగణించబడుతుందిహిందూమతంఎందుకంటేయాత్రికులస

ైట్లుతిరుమలవెంకటేశ్వరఆలయంఇతరచారిత్రకఆలయాలుపాటు , మరియు

Holy Place	District	Location
అలిపిరి	చిత్తూర్	తిరుపతి
చంద్రగిరి	చిత్తూర్	తిరుపతి
ఇస్కాన్	చిత్తూర్	తిరుపతి
కాణిపాకం	చిత్తూర్	చిత్తూర్
కాళహస్తి	చిత్తూర్	తిరుపతి
తిరుమల	చిత్తూర్	తిరుపతి

Holy Place	Location	Distance
అలిపిరి	తిరుపతి	5
చంద్రగిరి	తిరుపతి	20
ఇస్కాన్	తిరుపతి	1
తిరుమల	తిరుపతి	27
కాళహస్తి	తిరుపతి	24
కాణిపాకం	తిరుపతి	45

Holy Place	Location	Distance
కాణిపాకం	చిత్తూర్	24
లక్ష్మి నరసింహ స్వామి దేవాలయం	చిత్తూర్	20
గుర్రంకొండ	చిత్తూర్	16
తిరుమల	చిత్తూర్	75
కాళహస్తి	చిత్తూర్	82
చంద్రగిరి	చిత్తూర్	40

"ఆంధ్రప్రదేశ్ఆధ్యాత్మికరాజధాని "వలెసూచిస్తారు

Fig: Sample tables in the database

The about figure shows sample tables in the database, In this database contain multiple tables. These tables distinguish with their keyword of the lookup table. Whenever token match with lookup table it directly go to particular table name as the token in the database then it will be executed based on the query as in the form of telugu sentence in the Question Answering System.

ALGORITHM

Question Answering Systems provides Natural Language Interface to the user to ask query in Telugu. The Telugu language query is divided into tokens and then each token is matched with keyword tables in the database. If match found resultant natural language answer is generated to user. Otherwise user is asked to give question in different manner. In this paper regional holy places database is used as case study to test our model. To achieve this task following algorithm is developed to generate the SQL query from a given query in Telugu.

Algorithm:

Step 1: Create database for regional holy places.

Step 2: Create user interface to interact with user in Telugu language.

Step 3: Create a table with keywords related to regional holy places database.

Step 4: Input the query in Telugu language.

Step 5: Tokenize the query.

Step 6: Search the table for each token.

Step 7: If any token is match with keywords in table, SQL query is generated for given Telugu query.

Step 8: Execute query and get results

Step 9: Change the results into meaningful telugu sentences and display it to the user.

Step 10: If tokens are not found in table, ask user to input another query.

BLOCK DIAGRAM

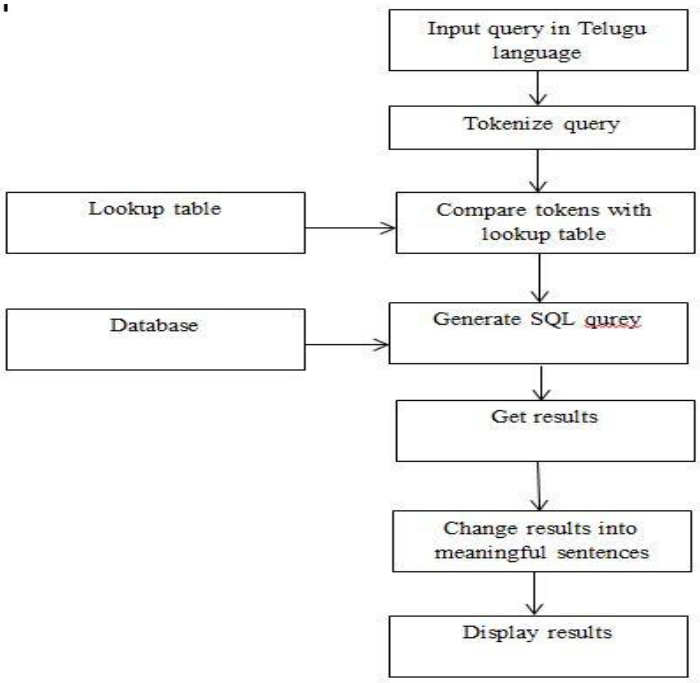

Fig: Block diagram for QA System

RESULTS

Experiments are conducted by taking the Regional holy place database as a case study.

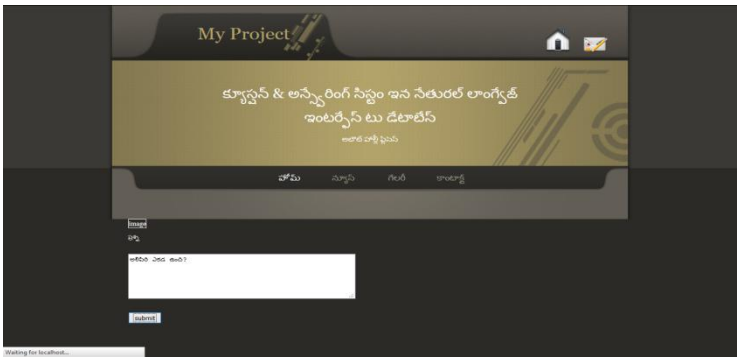

Fig: Question

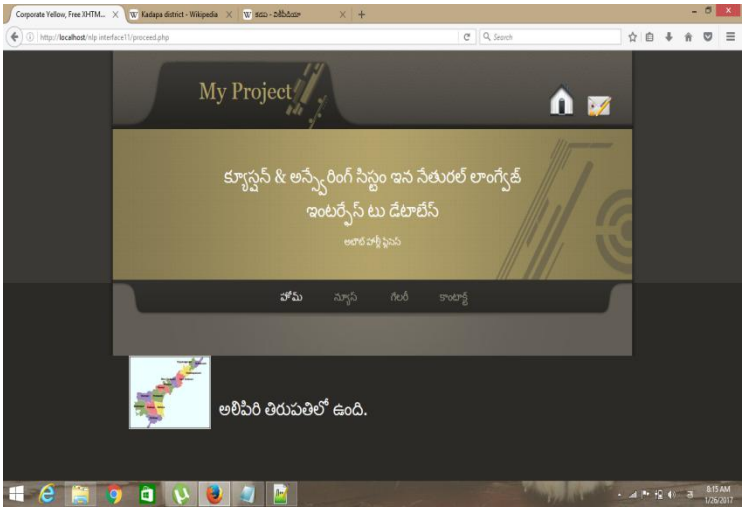

Fig: Answer

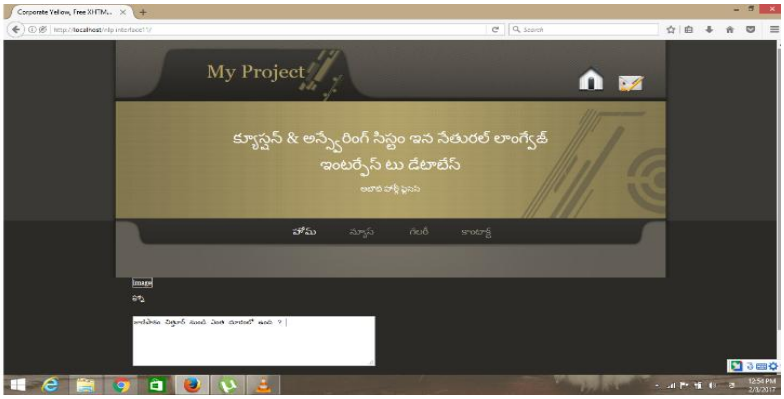

Fig: Question

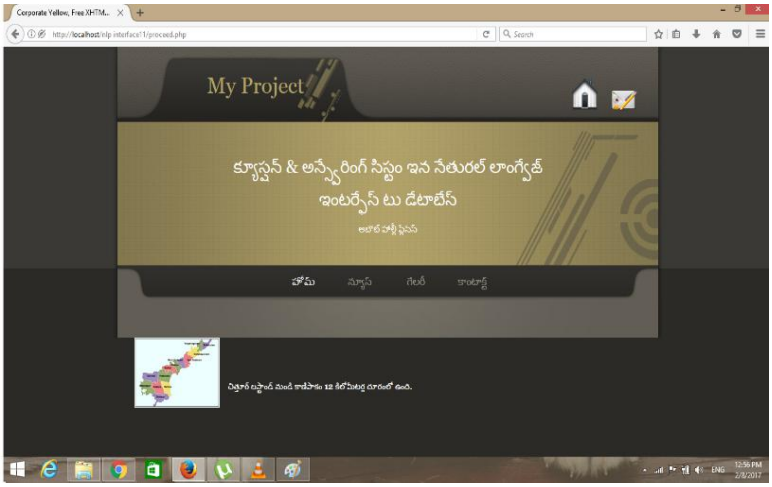

Fig: Answer

PERFORMANCE METRICS

There are several parameters that are used to analyze the performance of different Question Answering Systems. In this section we measure performance of the Question Answering System based on R and RU-Accuracy and Mean Reciprocal Rank techniques.

a)R-Accuracy and RU-Accuracy are used to measure Question Answering performance. A Question Answering system gives a list of ranked answer responses for every question, but R-accuracy and RU-accuracy only believes the correctness of the top 1 rank answer response on the list. An answer response is a pair that consists of an answer and its basis document.

b)Mean Reciprocal Rank MRR is used to measure the Question Answering performance based on all the top ranked answers, not presently top1 answer.

CONCLUSION

The NLIDB systems developed so far are basically used for business purpose. Here we are using this NLIDB system for holy place database which is very much useful for the uneducated people who are new to those places. Our QA system follows keyword based matching approach. All words/tokens need not be Knowledge base. The word which contain semantic information will be found in knowledgebase. Our system could achieve high successes rate if we restrict the coverage of questions. The future scope of the work could be done to improve the linguistic coverage of questions.

The future extension of this system is to implement a Telugu Interface system using Syntax and semantic approaches. Telugu is highly inflected with morphology, develop such a stemming algorithms without inflecting to the answers. Regional data is a huge information system in future can design systems which can handle complex user queries as well as aggregation functions in database. Machine Learning techniques are also using as classifiers for recognizing which type of question will given as an Input Query. This is also one of the possibility to adding aspect into future extension of this Question Answering System.

REFERENCES

[1] D. Ramesh, Suresh kumar Sanampudi, "Telugu Language Interface to Databases," *IJARCCE Trans. on Natural Language Processing*, vol.2, Vol. 2, Issue 7, July 2013.

[2] Ashish kumar, Kummar singh vaisha, "Natural Language Interface to Databases: Development Techniques," Elixir Computer Science and Engg Article.*May*.2013.

[3] N.Ramireddy, Sivaji Bandyopadhya, "Dialog based Question Answering system in Telugu," EACL Workshop on.*MultilingualQuestion Answering.*,2006.

[4] Li H,Shi Y, "A wordnet based natural language interface to relational databases," IEEE 2nd International conference. *on Computer andAutomation Engineering*,Feb 2010.

[5] Owda M, Zuhair. B, Crokett K "Conversation based Natural Language Interface to Relational Databases," *IEEE/WIC/ACM InternationalConference. On Intelligence and Intelligent agent Technology*, Nov 2007

[6] P. Resnik, "Access to Multiple Underlying Systems in JANUS", BBN report 7142, Bolt Beranek and Newman Inc., Cambridge, Massachusetts, (September 1989).

[7] AmandeepKaur, Parteek Bhatia, "Punjabi Language Interface to Database" communicated to International journal of computer science, WASET (World Academy of Science and Technology)

[8] Jyothsna Cherapanamjeri, Lavanya Lingareddy, Himabindu. K"Keyword based Question Answering System in Natural Language Interface to Database" *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3 Issue 12, December 2014*

[9] B. Thejesh " NLP Based Question –Answering System for Medical Information in Telugu Language" *International Journal of Computational Science, Mathematics and Engineering IJCSME-2-Volu*

me-Issue-12-December-2015.