

Question Answering System for Election Database in Telugu Language

Smt.M.Humera Khanam

Department of Computer Science and Engineering
Sri Venkateswara University College Of Engineering
Tirupati, India
e-mail:humera_svec@yahoo.co.in

MS.V.Bhargavi

Department of Computer Science and Engineering
Sri Venkateswara University College Of Engineering
Tirupati, India
e-mail:vemalabhargavi@gmail.com

Abstract—A Question answering (QA) system for Election information in Telugu has been described. The main objective of QA system is to handle dialogues between user and machine. The goal is to develop an automatic system capable to give the political information in Telugu Language. User can type questions in Telugu; system will generate exact answers in Telugu so that this system is very useful to native language speakers. Same question can be questioned in different ways but the meaning and answer is same in Telugu. This automatic system will give the best answer even though user pose questions different manner. The political QA system is helpful to the peoples who are preparing for civil exams, school children's. Political knowledge is the basic necessary requirement to every citizen. The proposed Election QA system needs an interface between user and database i.e., Telugu Language Interface to Database has been developed. It can be called as Natural Language interface to Database (NLIDB). By generating SQL Queries, answer can be collected from single Table or multiple tables. The accurate answers will be very useful and time saving.

Keywords- *Question answering, natural language interface, database, NLIDB, SQL Statements*

I. INTRODUCTION

The process of question answering system is a technique in information extraction and information retrieval. Most of the question answering systems are the process of retrieving appropriate answers for the user queries typed in natural language. Language is a medium of communication used by humans to express their views, ideas and emotions. Humans can able to learn new concepts and express their views is so natural but it is difficult to find how to process this language. Natural Language processing is a data driven empirical science. Natural language processing systems are built by training language independent and generic machine learning algorithms on large scale language data. Natural language processing (NLP) is an interpretation of language between human and machine. Natural language processing is so difficult because human language is complex and ambiguous. Language understanding needs contextual and general language apart from linguistic knowledge. Representing all this knowledge computationally is the challenging topic in NLP.

Question answering (QA) system is an automated system capable to answer user questions in short and exact manner. Most of the QA systems are of two types 1. Open domain 2. Restricted domain. When compared with restricted domain open QA system can answer any type of questions and it is not restricted to any particular topic. Restricted Domain QA has a long history, beginning with systems working over databases (e.g., BASEBALL (Green et al., 1961), and LUNAR (woods et al., 1972)).

The proposed question answering system for election information is restricted domain and it is monolingual language. i.e., both source and destination language is same (Telugu). Telugu is a source language for communication in AP and it belongs to Dravidian family. An automatic election QA system provides a user interface for users who can type

their questions regarding election information. The system need to extract appropriate information from the database for this it needs an interface between user and database. Lots of information is stored in database using tools like MS Access, Oracle and Others. Pupil who doesn't have an enough knowledge regarding Structured Query Language will face difficulty in handling and extracting useful information from the database. The proposed automated system is useful for non-technical and as well as technical users. In the proposed Election QA system user types questions in Telugu and system understands and generate answers in Telugu. Answers are extracted from the database so that, the system needs an interface i.e., Telugu Language Interface to Database. Using this system, user can query the database without by typing natural language questions and can access the requested information from the database. We call it as Natural Language Interface to Database (NLIDB). Pattern matching, syntactic, semantic these are the techniques in NLIDB. The current proposed Election QA system uses pattern matching technique and can access the data from single table or multiple tables using some operations like nested queries, conjunction and can also collect the data from related documents.

II. SYSTEM ARCHITECTURE

In this pattern matching technique, the user enters the input as a query in his/her natural Language i.e., Telugu. Now, query statement is broken in to tokens, collects the common keywords and then use knowledge base. The structured query format is selected based on tokens and the keywords in the input query statement useful to identify the minimum requirements that it can match with our database in order to have an accurate results. Each structured Query format is combined with a SQL generation process by using tokens. SQL statement is executed and extracts the result from the database. The resultant natural language answer is submitted to the user interface.

The Election QA system Architecture is shown in figure 1. User interface is the GUI visible to the user and can type questions regarding election information. The input query is broken down in to tokens and collects the keywords from the given input query. Using predefined pattern matching templates, tokens and keywords are compared with each templates. If templates are matched with user input query the associated SQL statement is executed and the result from the database is arranged in a structured answer format and finally, Natural Language answer is submitted to user interface.

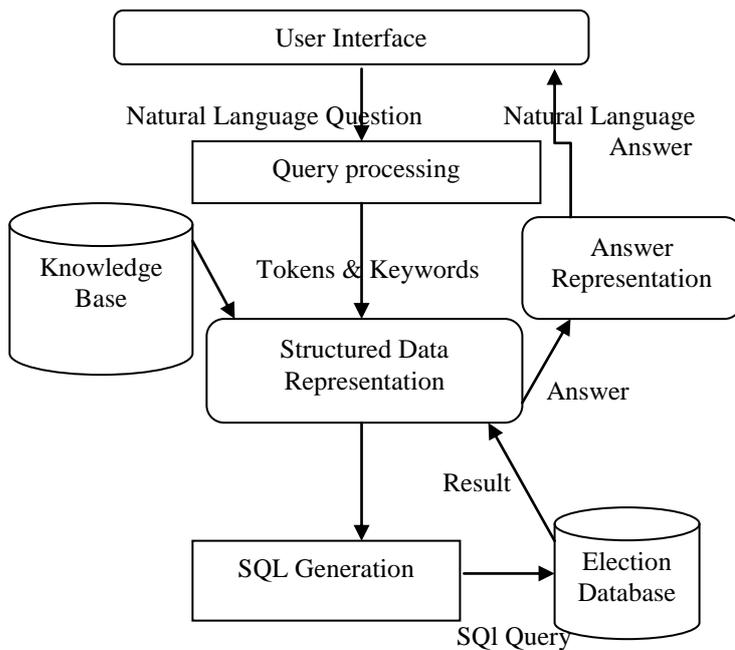


Fig1: Election QA System Architecture

III. RELATED WORK

NLP researches have been working on Question Answering system since 1970's with the systems like, BASEBALL Green et al., 1971), it provides answers to the questions about the American Baseball. Peoples have questions and they need answers not documents. Automatic Question answering will definitely be a significant advance in the state of art information retrieval technology. Early QA SANVY is best natural language processing system uses pattern matching technique. The main advantage of pattern matching approach is simplicity.

In Syntax based systems, the input query is processed syntactically and generates parse tree. The resultant parse tree is mapped to query language to retrieve information from the database. This type of approach is used to domain specific database systems. LUNAR [1] is best example for this technique. The Geologists use the LUNAR system to ask questions about moon rocks.

LADDER system is good example for semantic approach. SANVY, LUNAR, LADDER all these QA system uses English as source and destination Language.

In the year 2006, dialogue based question answering system for Telugu language [2] has been developed. In this system

user can ask questions through speech and get the relevant answer through sounds. Here, Speech is the source and destination language.

A Hindi Question Answering System “Prashnottar” [3]. In this system, input given by the user is HINDI language then system translates Hindi to WX format and searches on database. Finally, the resultant answer is translated from WX notation to Hindi language.

IV. DESIGN OF ELECTION INFORMATION SYSTEM

The key role of election system lies in designing election database and knowledge base.

Database means storing information in such a way that the information can be retrieved. Consider Relational database, information is stored in tables in the form of rows and columns. In election information system, database contains fields like constituency name, elected person, votes, party, symbol, state. Parliament, assembly are the relations in the election database. The political information like elected persons of MLA's, MP's, ministers and other information is stored in election database in the form of rows and columns. Data independence is the advantage of Relational model.

Election question answering system maintains knowledge base it is quite obvious in restricted domain. It greatly improves disambiguation and parsing.

The Election QA system provides a user interface which is visible to user who can type their input query in Telugu. Whenever pupil types their input query the program will generate the tokens from the given input query and collects the keywords by applying stemming. The program will have predefined templates generated by the programmer, if user input query matched with the templates then the corresponding SQL query will execute and returns the result to the user interface. For example: if the user input queries will be,

“తిరుపతి నేయోకవర్గానికి ఎమ్మెల్యే ఎవరు?” [tirupathi neyokavarganiki mla evaru?]

“తిరుపతి ఎమ్మెల్యే ఎవరు?”

“తిరుపతి నియోజకవర్గానికి ఎమ్మెల్యే ఎన్నికయిన అభ్యర్థి ఎవరు?”

“తిరుపతి నియోజకవర్గానికి టిడిపి అభ్యర్థిగా ఎన్నికయిన వ్యక్తి ఎవరు?”

The above four questions have similar meaning i.e., “who is tirupati mla?” But the question framing is different. Here, తిరుపతి represents the constituency, ఎవరు represents the Who,

ఎన్నికయిన, గెలిచినా represents the elected person. All these predefined knowledge is stored in the knowledge base. The corresponding SQL statement for the above input queries will be

```
SELECT name FROM assembly WHERE constituency="తిరుపతి";
```

We store some of Telugu inflections like [వమ్మలగా (mla ga),ఎన్నికయన(ennikayana),నియోజకవర్గానికి(neyojakavarganiki)] గా,యన,కి all these are morphological analysis of input query used to identify root words. After identification, system analyzes the whole input statement and gets the tokens such as, constituency, elected person, party, and some keywords. For ex: The input query,

చంద్రగిరి నుండి ఎన్నికయన వ్యక్తి ఎవరు? (chandragiri nundi ennikayana vykthi evaru?) Who is chandragiri mla? The input statement is parsed through spaces and applying stemming rules used to identify keywords which are helpful to frame correct query format. Here, in this example: చంద్రగిరి

[constituency name], వ్యక్తి [person], ఎవరు [who]. The above input query represents the name of the person who elected as MLA form chandragiri constituency. Each structured query format is associated with SQL generation procedure. SQL is the language designed to use the relational databases. SQL statement for the above input query is

SELECT name FROM assembly WHERE constituency="చంద్రగిరి";

The above statement is executed and returns the natural language answer to the user.

[చంద్రగిరి నుండి ఎన్నికయన వ్యక్తి బాస్కరరెడ్డి]. Here, the result is collected from single table. The automatic election question answering system provides answers for factoid questions like (ఎవరు, ఎక్కడ,) for example: వెంకటరమణ ఎక్కడ నుండి ఎమ్మెల్యేగా ఎన్నికయారు?

సతనపల్లి నేయోజకవరం ఎమ్మెల్యే ఎవరు?

In some cases, we need to use multiple tables to get an answer. Nested queries are used to retrieve answer from multiple tables. చిత్తూర్ జిల్లా ఎంపి మరియు ఎమ్మెల్యే ఎవరు?

The corresponding SQL query is,
SELECT name FROM assembly a, parliament WHERE a.constituency = p.constituency and constituency= "చిత్తూర్";

The above SQL statement will give answer to the user using two tables such as, assembly, parliament using condition specified in WHERE clause.

If the input query,
తిరుపతి నుండి ఎన్నికయన టిడిపి అభ్యర్థి ఎవరు? The corresponding SQL statement is,
SELECT name FROM assembly WHERE constituency="తిరుపతి" AND party="టిడిపి";

The Election QA system will returns the result,
తిరుపతి నుండి ఎన్నికయన టిడిపి అభ్యర్థి ఎం.వెంకటరమణ.

Suppose if user requires information which is stored in multiple columns.For ex:

తిరుపతి నుండి ఎన్నికయన ఎమ్మెల్యే ఎవరు అతను ఎ పార్టీ అభ్యర్థి?

In this input query, user needs information about elected person and party using SELECT clause it is so simple to get the result stored in multiple columns just by specifying their

field names after SELECT clause. The SQL statement for the above input query is
SELECT name, party FROM assembly WHERE constituency="తిరుపతి";

The natural language answer generated by the system for the above query is

తిరుపతి ఎమ్మెల్యే వెంకటరమణ అతను టిడిపి అభ్యర్థి.

If user needs information about how many constituencies are there in Andhra Pradesh? Here, the special operator * is used. For ex: The Input query

ఆంధ్రప్రదేశ్ లో ఎన్ని నియోజకవర్గములు ఉన్నాయి? The SQL statement for this query is SELECT count (*) FROM assembly WHERE state="ఆంధ్రప్రదేశ్"; Here, COUNT function is used to return the number. Finally, the natural language answer generated by the system is

ఆంధ్రప్రదేశ్ లో 175 నియోజకవర్గములు ఉన్నాయి

V. EXPERIMENT AND RESULTS

Experiments are conducted to the automatic election QA system by checking with different models of queries as shown in below table.

Natural Language Question	SQL Statement	Natural Language Answer
ఎం.వెంకటరమణ ఎ పార్టీ అభ్యర్థి?	SELECT party FROM assembly WHERE name="ఎం.వెంకటరమణ";	ఎం.వెంకటరమణ టిడిపి అభ్యర్థి.
పూతలపట్టు ఎమ్మెల్యే ఎవరు అతను ఎ పార్టీ అభ్యర్థి ఎ జిల్లా కిందకి వస్తుంది?	SELECT name, party, district FROM assembly WHERE constituency="పూతలపట్టు";	పూతలపట్టు ఎమ్మెల్యే రవి అతను కాంగ్రెస్ పార్టీ అభ్యర్థి చిత్తూర్ జిల్లా కిందకి వస్తుంది.
తెలంగాణలో ఎన్ని అసెంబ్లీ నియోజకవర్గములు ఉన్నాయి?	SELECT count(*) FROM assembly WHERE state="తెలంగాణ";	తెలంగాణ లో 119 అసెంబ్లీ నియోజకవర్గములు ఉన్నాయి.

Election QA system, provides the user interface which is available to the users who can type their questions in Telugu and then click OK button. System process and generates Natural language answer to user interface. The snapshot for user interface as shown in below fig 2. The snapshot for answer generation as shown in below fig 3.



Fig2: snapshot for user interface.

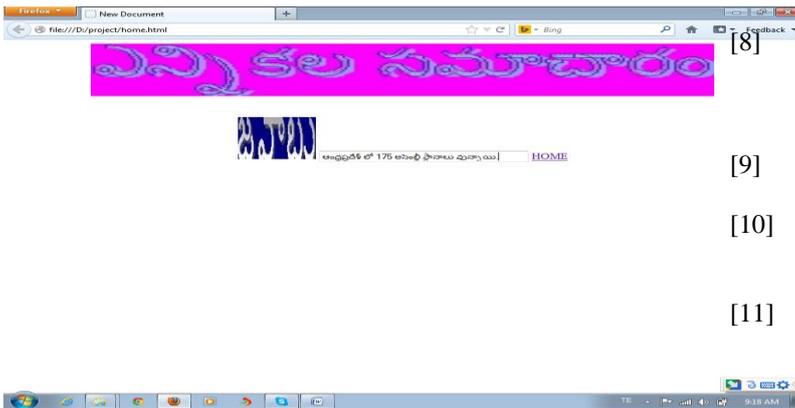


Fig 3: snapshot for answer generation.

After, answer is returned to the user window. If user has one more question to query the system Home hyper link is available. When user click on Home link the corresponding home page will open where user can type their question.

VI. CONCLUSION

The proposed automated Election Question Answering system for Telugu Language performs pattern matching technique uses election database, accepts input in Telugu and generate SQL query using tokens and keywords from the input query and natural language answer is submitted to the user window. The system performs operation on single table, multiple tables using functions AND, Nested Queries. The proposed system will retrieve answers to the factoid questions, description type questions regarding elections in Telugu Language. The future extension of this system is to implement a Telugu QA system using Syntax and semantic approaches.

REFERENCES

- [1] W.A. woods R.M. Kaplan and B.N Webber “ The LUNAR Sciences Natural Language Information Systems” Cambridge 1972.
- [2] Rami Reddy Nandi Reddy, Sivaji Bandyo padhyay “Dialogue based Question Answering System in Telugu” published at EACL 2006 Workshop on Multilingual Question Answering - MLQA06.
- [3] jaspreet kaur, vishal gupta “comparative analysis of Question Answering system in Indian Languages” published at International Journal of Advanced Research in Computer science and software engineering, may 2014.

- [4] D. Ramesh, suresh kumar “Telugu Language Interface to Database” published at International Journal of Advanced research in computer and communication Engineering, july 2013.
- [5] poonam gupta, vishal gupta “A survey of existing question answering techniques for Indian Languages” published at Journal of Emerging Technologies in Web Intelligence, may 2014.
- [6] S. QUARTERONI, S. MANANDHAR “ Designing an Interactive Open domain question answering system, 2007 at Cambridge university.
- [7] David Elworthy “ Question Answering using a large scale NLP, Microsoft Research Limited, St. George House, 1 Guildhall Street, Cambridge CB2 3NH, Uk.
- [8] Akshar bharthi, vineet chaithanya, Rajeev sangal “Natural Language Processing A paninian perspective”, department of computer science and engineering at Indian Institute of Technology Kanpur.
- [9] Poonam Gupta, Vishal Gupta “A survey of Text Question Answering Techniques.
- [10] David L.waltz “An English Question Answering System for a Relational Database at University of Illinois at urban-champaign.
- [11] Hoojung Chung, Young-In Song, Kyoung-Soo Han, Do-Sang Yoon, Joo-Young Lee, Hae-Chang Rim and Soo-Hong Kim. 2004. A Practical QA System in Restricted Domains.
- [12] Androutsopoulos I, Ritchie G. D, and Thanisch P. 1995. Natural Language Interfaces to Databases.

Appendix: Examples of Election QA System

Example 1:

User: చిత్తూర్ జిల్లాలో ఎన్ని అసెంబ్లీ నియోజకవర్గములు ఉన్నాయి ?

(Chittoor jillalo enni assembly neyajakavargamulu vunayi [How many constituencies are there in chittoor district])? The corresponding SQL statement generated by the system is, SELECT count (*) FROM assembly WHERE district=” చిత్తూర్”;

System: చిత్తూర్ జిల్లాలో 14 అసెంబ్లీ నియోజకవర్గములు ఉన్నాయి?

Example 2:

User: రాయచోట్ నియోజకవర్గం నుంచి అసెంబ్లీ సభ్యునిగా పోటీ

చేసిన వై.ఎస్.ఆర్.సిపి అభ్యర్థి పేరు ఏంటి ? (rayachott neyajakavargam nunchi assembly sabhuniga poti chesina Y.S.R.CP abhyarthi peru enti)? The corresponding SQL query generated by the system is,

SELECT name FROM assembly WHERE constituency=” రాయచోట్”AND party=” వై.ఎస్.ఆర్.సిపి”;

System: రాయచోట్ నియోజకవర్గం నుంచి అసెంబ్లీ సభ్యునిగా పోటీ చేసిన వై.ఎస్.ఆర్.సిపి అభ్యర్థి పేరు శ్రీకాంత్ రెడ్డి.

Example 3:

User: చిత్తూర్ జిలలోని పదనాలుగు నియోజకవర్గంలో టి.డి.పి ఎన్ని

నియోజకవర్గములు గెలుచుకుంది? (chittoor jilaloni padanalugu

neyojakavargamlo Tl.Dl.Pl enni neyojakavargamulu

geluchukundi?) The corresponding SQL query generated by

the system is, SELECT count (*) FROM assembly WHERE

district=" చిత్తూర్" AND party=" టి.డి.పి ";

System: చిత్తూర్ జిలలోని పదనాలుగు నియోజకవర్గంలో టి.డి.పి 6

నియోజకవర్గములు గెలుచుకుంది.

Example 4:

User: సతనపల్లి నియోజకవర్గం అసెంబ్లీ స్థానాని ఏ పార్టీ

గెలుచుకుంది?

(sathhanapalli neyojakavargam assembly sthananiki e party

gelichindi?) the corresponding SQL statement generated by

the system is, SELECT party FROM assembly WHERE

constituency=" సతనపల్లి":

System: సతనపల్లి నియోజకవర్గం అసెంబ్లీ స్థానాని టి.డి.పి పార్టీ

గెలుచుకుంది.