Keyword based Question Answering System in Natural Language Interface to Database

Jyothsna Cherapanamjeri, Lavanya Lingareddy, Himabindu. K

Abstract— **Natural language interface to database system is one of the area of Natural language processing. A recent development in Natural Language Processing is Question Answering System. Question Answering accepts queries formulated in natural language and respond to the user with a specific answer in Natural Language Sentence. Keyword based question answering system in NLIDB for crop statistic information has been described. Different components had been developed in this system for processing the query and translating it into formal SQL query. Based on the results from the SQL query, a natural language answer is generated.**

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

I. 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 [4]. 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/illiterate ones to retrieve data from database in a simple way [2][5]. The Question Answering system about crops like crops price, production, yield in Telugu has been described. This system uses keyword based matching technique to convert the Natural Language Query in Telugu to SQL [3].

II. RELATED WORK

Different techniques are used to develop NLIDB system such as [6]

1. Pattern Matching system,
2. Syntax based system,
3. Semantic based system
4. Intermediate language representation.

Among these pattern matching was most widely used techniques. The main advantages of this approach is its simplicity, no elaborate parsing and interpretation modules are needed. Also pattern matching systems gives some related responses even if the input is ambiguous [1].

In syntax based and semantic grammar systems, the parse tree is generated for user's queries. The parse tree is directly mapped to database query language. The drawback with these approaches is difficult to map the parse tree to query language [7].

In order to overcome this drawback, intermediate representation languages were proposed. In this first, sentence is mapped to logic query language followed by the translation of the logical query into general database query [2]. We proposed keyword based technique for Telugu language interface (TLI) system [3].

III. PROPOSED SYSTEM

. We proposed keyword based technique for Telugu language interface (TLI) system. In this system, we will map all keywords in the user query to the database. All the 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. The architecture for keyword matching system is expressed in detail in section IV.

IV. SYSTEM ARCHITECTURE

- At first, user gives Telugu language query which is converted into WX-Notation using code converters, 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

Manuscript received December, 2014.

***C. Jyothsna Cherapanamjeri**, Computer Science and Engineering , Yogananda Institute of Technology and Science,., Tirupathi, India,9989008689.*

***Lavanya Lingareddy**, Computer Science and Engineering , Yogananda Institute of Technology and Science,., Tirupathi, India.*

***Hima Bindu K**, , Computer Science and Engineering , Yogananda Institute of Technology and Science,, Tirupathi, India*

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.

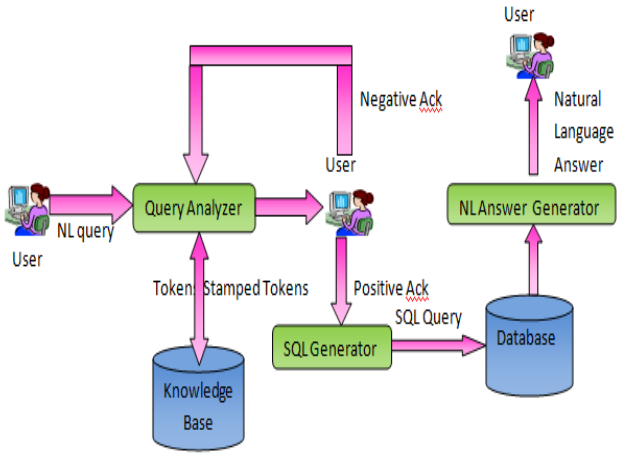

Figure1: System Architecture

V. ALGORITHM

Keyword matching systems to natural language interface provides the user to give 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 crop statistics 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: ***Keyword Matching***

Step 1: Input the query in Telugu language

Step 2: Telugu language query is converted into WX notation

Step3: Tokenize the query

Step 4: *if token==keywords (table_name)*

4.1 :Equivalent predefined NL query is generated to user

else

goto step 6

Step 5: *if user_ack==true* //Positive acknowledgement

5.1: SQL query is generated and executed

5.2: Corresponding NL answer is retrieved from database

else

Alternative NL Query is generated until user gives positive acknowledgement

If user_ack==true

goto Step 5

else

goto step 6

Step 6: invalid input query, try again.

The following example depicts the conversion of Telugu language query to its SQL query.

EXAMPLE:

కడప నందు రాగి ధర తెలుపుము

kadapa naMxu rAgi Xara weVlupumu (WX- Notation)

Place Name: **కడప** (kadapa)

Crop Name: **రాగి** (rAgi)

Keyword: **ధర**(Xara [Price])

The [price] query frame is selected

```
SELECT price
FROM crop_info
WHERE Crop_Name='rAgi'.
```

VI. RESULTS

Telugu language query regarding crop statistics information as shown in the above example is given as input. Then the processing of algorithm is performed when ఫలితము సాధించు button is pressed. The system starts working by utilizing the semantic information in the knowledge base. Hence it implements the SQL query for query frame and the result is displayed to the user in natural language. The figure 2 shows the screen shot for inputting the Telugu query such as "2000 లో కడప నందు రాగి ధర తెలుపుము"

and the figure 3 shows the screen shot for confirmation of user query such as

"2000 సంవత్సరము రబి, ఖరీఫ్ సీజను లో కడప నందు రాగి ధర తెలుపుము ఇది సరిగా ఉంటే అవును అనే చదరముపై నొక్కండి లేదా కాదు అనే చదరముపై నొక్కండి".

Figure 4 depicts result generation after processing the input such as
"2000 సంవత్సరము రబి, ఖరీఫ్ సీజను లో కడప నందు రాగి ధర క్వింటాకు 449.00, 496.00 రూపాయలు".

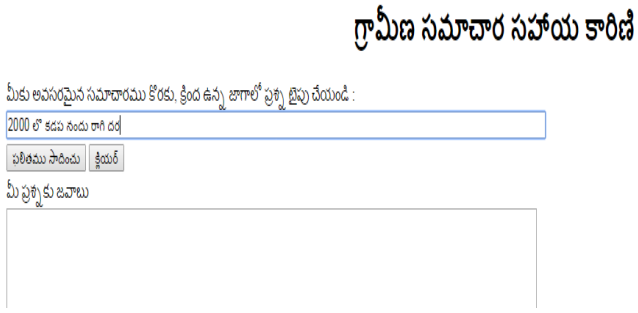

Figure 2: Screen shot for inputting the query

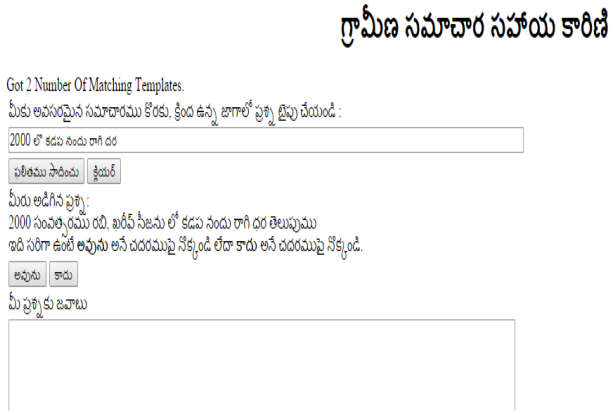

Figure 3: Screen shot for confirmation of user query

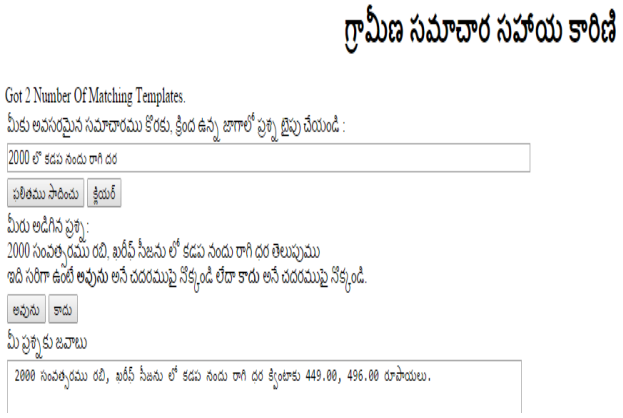

Figure 4: Screen shot for result generation

VII. CONCLUSION

The NLIDB systems developed so far are basically used for business purpose. Here we are using this NLIDB system for crop statistic information which is very much useful for the uneducated people who are from agriculture background. Our QA system follows keyword based matching approach . All words/tokens need not be knowledgebase. 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.

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 Lnguage 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.*Multilingual Question Answering*.,2006.

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

[5] Owda M, Zuhair. B, Crokett K "Conversation based Natural Language Interface to Relational Databases," *IEEE/WIC/ACM International Conference. 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)

Jyothsna Cherapanamjeri is presently working as Assistant Professor in Department of Computer Science and Engineering, Yoganandha Institute of Technology and Science , Tirupathi, India. She received master degree in Computer Science and Engineering in Sri Venkateswara University (SVU),Tirupathi. Her area of interest is Natural Language Processing.

Lavanya Lingareddy, is presently working as Assistant Professor in Department of Computer Science and Engineering, Yoganandha Institute of Technology and Science , Tirupathi, India.. She received master degree in Computer Science and Engineering in SVCE, Tirupathi, affiliated to JNTU A. Her area of interest secure computing and Natural Language Processing.

Hima Bindu K is presently working as Assistant Professor in Department of Computer Science and Engineering, Yoganandha Institute of Technology and Science, Tirupathi, India.. She received master degree in Computer Science in SVEC, A.Rangampet, affiliated to JNTUA. Her research work was published in international and national conferences and also in international journals. Her area of interest is software engineering, cloud computing, secure computing and Natural Language Processing.

