Journal of Engineering and Applied Sciences 12 (Special Issue 5): 7074-7076, 2017
ISSN: 1816-949X

Information Extraction from Relational Database Using Speech Recognition

Kajal A. Jain and Avinash J. Agrawal
Ramdeobaba College of Engineering and Management, Nagpur, India

Abstract: As we know, today information plays an important role in human life. And in the world of computer, internet and technology this information got the form of database. All the information or data are stored in database in organized table format. By using SQL queries people are able to access data from these databases. People who are able to access these database are developer or those who known to query language but casual user or non-technical are not. The study describe the methodology that user will give audio command in natural language which will be converted into text, i.e. in natural language and from the generated text the SQL query will be formed. Result of this methodology will be the information from database which was demanded by user.

Key words: HMM (Hidden Markov Model), NLIDB (Natural Language Interface for Database), NLP (Natural Language Processing), SQL (Structured Query Language), information, database

NTRODUCTION

In this era, the main and biggest source of information is database and to access these databases only required is SQL query. But writing query command is not that much possible for non-developer or non-programmer as much as for a developer or programmer. Thus, creating interaction between non technical person and database without burden of queries, SQLs leads to natural language interfaces for databases. Thus in another word, NLIDB is an interface between user and database.

The main purpose of natural language query processing is for an English sentence to be interpreted by the computer and appropriate action taken. The application that will be possible when computer would be able to process natural language, translating language accurately and extracting information from data source depending on the user's request.

In this study, we studied the various approaches proposed by the many of researches and these studies helped to develop more efficient framework for accessing database by natural language processing along with audio input. Ontology based information retrieval technique which is guided with predefined knowledge based rules.

Literature review: Basically, NLIDB is not new area for research. Many of researchers are working on it since long time.

LUNAR (Hendrix *et al*., 1978) involved a system that answered questions about rock samples brought back from the moon. Two databases were used, the chemical analyses and the literature references. The program used an augmented transition network and wood's procedural semantics. The system was informally demonstrated at the Second Annual Lunar Science Conference in 1917.

LIFER/LADDER (Hendrix *et al*., 1978) was one of the first good databases NLP system. It was designed as a natural language interface to a database of information about US Navy ships. As described by Hendrix in his study, system used semantic grammar to parse questions and query a distributed database. The LIFER/LADDER system could only support simple one-table queries or multiple table queries with easy join conditions.

Question-answering system (Dang and Thi, 2009) proposed a method to build a specific question and answering system which is integrated with search system for e-Books in the library. One can use simple English question for searching the library with the information about the needed e-books, such as title, author, publisher, etc.

It has been observed in research study that by using shallow semantic analysis many of NLIDB system are designed. Using database with NLP in application for area like colleges, industries, banks, etc., require detail analysis. Success rate of detailed analysis is greater in comparison with shallow analysis. Therefore a method is

Corresponding Author: Kajal A. Jain, Ramdeobaba College of Engineering and Management, Nagpur, India

proposed to use ontology to represent domain knowledge and language modeling to represent language knowledge (Agrawal and Kakde, 2013).

The main objective of NLIDB is to accep the query sentence and try to understand it by applying lexicon, syntactic and semantic analysis and then convert it into SQL (Androutsopoulos *et al*., 1995). Natural language interface for database deals with structured text which has been parsed also its entities and attributes have been identified before.

Parsing will identify the type of dependencies. Parse tree will generate in technical from (Popescu *et al*., 2004). The user's query gets parsed by the syntax-based system and then the direct mapping is done between resulted from parsed tree and an expression in some database query language.

MATERIALS AND METHODS

Problem statement: Database acts as main source of information and organizes the data in a model that supports processes requiring information. More number of employee will required if one is willing to obtain information on basis of SQL commands. As NLP is a process by which user can enter query in natural language will be converted into SQL query (Fig. 1).

Any non-technical or ordinary person is not expected to know SQL commands or SQL language. Thus, this ideology will help to generate same, so will not require extra technical employee for this.

In simple term, asking question to database in natural language is very convenient and easy method of data access, especially for casual user who does not understand complicated database query language such as SQL. This model will proposes the architecture for translating English query into SQL.

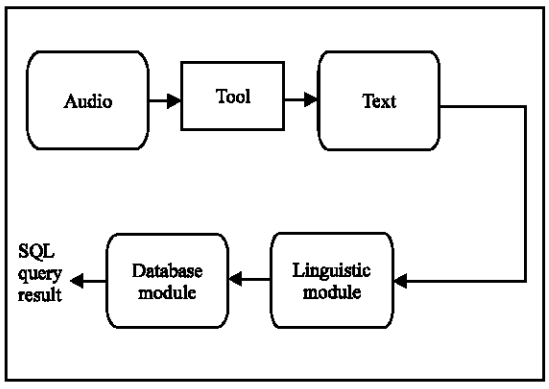

Fig. 1: Problem definition

RESULTS AND DISCUSSION

Proposed framework: One of primary concept in any natural language processing system is sematic analysis. It analyses the meaning and represent meaning in proper format. One of the methods in semantic analysis is faster but with less accuracy is shallow semantic analysis and another method is not much flexible but more accurate is deep semantic analysis (Sangeeth and Rejimoan, 2015).

When intension is to analyses the meaning of any natural language question from perspective of database concept natural language interface to database is required. Thus, database query is expected from natural language question. Pre-processing is required for starting natural language processing. When audio speech is converted to text, this text is natural language question. Pre-processing is done on this text which involved Tokenization, POS tagging and Chunking.

If given a string or sentence, tokenization is the tasks to chopping it up into pieces, called tokens, also remove the characters present in string such as punctuation. The process of assigning one of the part of speech to given words is called parts of speech tagging. POS tagger is program that does this job. Shallow parsing is another term used for chunking, it identifies the part of speech and short phrases. Related words are groups into meaningful sentence.

Another concept included here is Ontology. Ontology is nothing but to represent the knowledge by a set of concept within a domain and relationships between those concepts. Every information system has its own ontology.

The system comprises of 5 bean classes namely Tables, Columns,Values, Foreign keys and Query which are used to save the data about the tables and the key words used to identify them (Enikuomehin and Okwufulueze, 2012). The DB helper class allow us to handle all the comparison and interfacing required with the database (Fig. 2).

System design

Speech to text: Speech recognition involves receiving speech through a device's microphone which is then checked by speech recognition service again a list of grammar. When a word or phrase is successfully recognized, it is returned as result as a text string and further actions can be initiated as a result.

Pre-processing: Pre-processing will remove stop words, stem words and analyze key words. For example: if input text is "who is head of computer", then keyword will be 'head' and 'computer'.

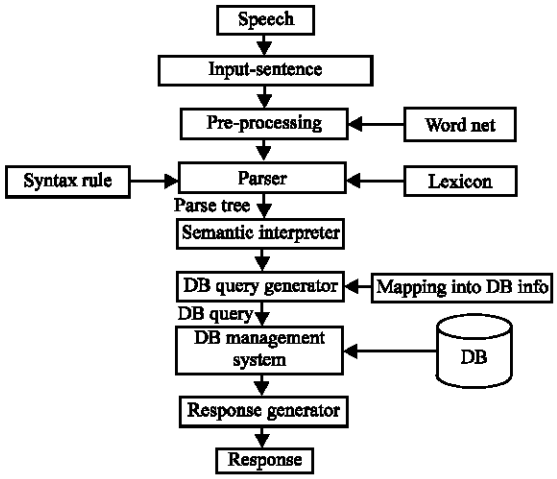

Fig. 2: Proposed framework

Parsing: Parsing will identify type of dependencies. Parse tree will be generate in technical form. The user's query gets parsed by syntax-based system and then direct mapping is done between resulted parsed tree and an expression in some database query language.

Ontology: Actual mapping of database phraseology or database language with terms of domain ontology. With above example target is 'head', i.e., predicate, 'who' and 'is' are subject and 'computer' is object.

Database query generator: With successful grouping of all words in statement, the target word and object will be trigger to SQL query generation.

Database management system: It will consist of normalized table with appropriate information.

CONCLUSION

As per the working approach in present, database plays cardinal role in every domain. Thus, this application can be very convenient to use. Another vital specification is that, no technical query language knowledge is required.

RECOMMENDATION

In future, research to be done is formation of query with join operator and other clauses in SQL.

REFERENCES

Agrawal, A.J. and O.G. Kakde, 2013. Semantic analysis of natural language queries using domain ontology for information access from database. Intl. J. Intell. Syst. Appl., 5: 81-90.

Androutsopoulos, I., G.D. Ritchie and P. Thanisch, 1995. Natural language interfaces to databases: An introduction. Nat. Lang. Eng., 1: 29-81.

Dang, N.T. and T.T.D. Thi, 2009. Natural language question answering model applied to document retrieval system. World Acad. Sci. Eng. Technol., 51: 36-39.

Enikuomehin, A.O. and D.O. Okwufulueze, 2012. An algorithm for solving natural language query execution problems on relational databases. Intl. J. Adv. Comput. Sci. Appl., 3: 169-175.

Hendrix, G.G., E.D. Sacerdoti, D. Sagalowicz and J. Slocum, 1978. Developing a natural language interface to complex data. ACM. Trans. Database Syst., 3: 105-147.

Popescu, A.M., A. Armanasu, O. Etzioni, D. Ko and A. Yates, 2004. Modern natural language interfaces to databases: Composing statistical parsing with semantic tractability. Proceedings of the 20th International Conference on Computational Linguistics, Aug. 23-27, Geneva, Switzerland, pp: 1-7.

Sangeeth, N. and R. Rejimoan, 2015. An intelligent system for information extraction from relational database using HMM. Proceedings of the International Conference on Soft Computing Techniques and Implementations (ICSCTI), October 8-10, 2015, IEEE, Faridabad, India, ISBN:978-1-4673-6792-9,pp: 14-17-10. 1109/ICSCTI. 2015.7489594.