

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology



ISSN 2320-088X
IMPACT FACTOR: 6.017

IJCSMC, Vol. 6, Issue. 6, June 2017, pg.195 – 199

Common Database Interface with NLP

Yawar Abass Mir¹, Zahid Zahoor Koul², Syed Irfan³

¹Student, Computer Science Engineering, SSM College of Engineering & Technology, Kashmir, India
miryaware479@gmail.com

²Student, Computer Science Engineering, SSM College of Engineering & Technology, Kashmir, India
koulzahid@gmail.com

³Assistant Professor, Computer Science Engineering, SSM College of Engineering & Technology, Kashmir, India
syedirfan.ssm@gmail.com

Abstract: *In our daily life we always have to deal with information which plays an important role in our lives. One of the major sources of information is database. We all use database technology which is having the deep impact on today's growing use of computer and internet. However, database system is not easy to understand by every user because the user needs to be technically good in the database technology. Hence people with no knowledge of database technology may find it difficult to access. Therefore, there is need to find out the new technique and methods that will help user to provide an environment where a non-expert user will be able to write and execute the database queries with ease and efficiency by using Natural Language Processing, without the proper knowledge of SQL (Structured Query Language), which is the basic language to interact with the database server. Therefore this idea of using natural language instead of query triggered, the development of a different type of processing system called Natural Language processing (NLP) came into existence. NLP allows common users to only retrieve the information from database using a natural language (English) and to the database administrator all the interfaces are available which will help administrator to manage, modify and update its database with ease. Further we develop this enhanced engine as web based application with common interface to all database servers so as to make availability of interface globally, and preventing extra burden of installation of database servers on users. This building of the interface which is common to all database servers with the Natural Language Processing is called CDINLP hence increases reliability. This CDINLP is built with an extra functionality which enables the user to interact with the database server by only selection of the requirement on the interface and this functionality is termed as Query builder.*

1. INTRODUCTION

Common Database Interface is the new developed system which increases the reliability for the user to access different types of database servers via common interface. Again this complex developed system has been embedded with an important subsystem which increases the more reliability of the system and this subsystem is called Natural Language Processing. Again these different types of database servers can be remotely accessed by using IP addresses of the respective nodes.

Natural language processing is a field of artificial intelligence where we Retrieve Information, Translate this information for Machine and Analyse Language. The target of fetching/retrieving data from database server by natural language processor is to build data access simpler for the non-expert people. As natural language is the simplest path to learn and use the system, it has proved to be the toughest for a computer to become expert. To access any database server/engine a user must know the query

languages that database server understands or for which the server is designed. Those users who are expert in these languages can access data or information from any database server/engine. So to access the data from database server/engine, a graphical interface is used which needs basic training for using these systems. There are many people who know English but are not expert to write queries in database servers. This interface helps an end user to query the system in natural languages like English and can see the results.

Further we develop this enhanced engine as web based application with the functionality of graphical selection of the fields on the interface that implicitly generates the query for the user and will give the result according to the selection done by the user. Also this enhanced feature improves our system qualitatively and is applicable for all types of the database servers for which the common interface is build.

2. OBJECTIVE

Developing a system for the users who know only English language with common interface to all database servers. User can access database using English language and to get the result in the same language in the same way as the database servers does. All other objectives are given bellow:

- To develop common database interface to all relational database servers which would support Natural Language Processing.
- To create GUI where the user can input the query in English Language and also can interact with the system by graphical selection of the fields.
- To build a system that process the English query and fetch the data from multiple databases, tables & columns by using joins implicitly.
- To build a system that can handle quires like select, update, delete on relational databases of multiple database servers.
- To execute queries with the help of graphical selection of fields. The subsystem that enables this feature is called the query builder.
- To execute queries that can arrange data of tables in the form ascending order and descending order.

3. PROPOSED SYSTEM

3.1 Natural Language Interface for Database

Common Database Interface with Natural Language Processing (CDINLP) has an interesting area called the development of a natural language interface to database systems. In the previous few decades many such interface systems have been developed. Using these interface systems, users can interact with database server/engine in a more easy and simple way. Due to this reason, the application of CDINLP is globally used today. In the past time Natural Language Processing has been a very interesting area of research. The aim of Natural language processing with common database interface is to provide such type interface where user can interact with database more conveniently using the natural language (English). At the same time the graphical selection of fields (query builder) provides the enhanced way for the user to interact with the database servers who know only graphical interaction with the system. Hence the CDINLP is the system which can convert the query from native language (English) to SQL.

3.2 Sub components of CDINLP

There are two sub components of CDINLP

- Linguist components
- Database components

Linguistic Component

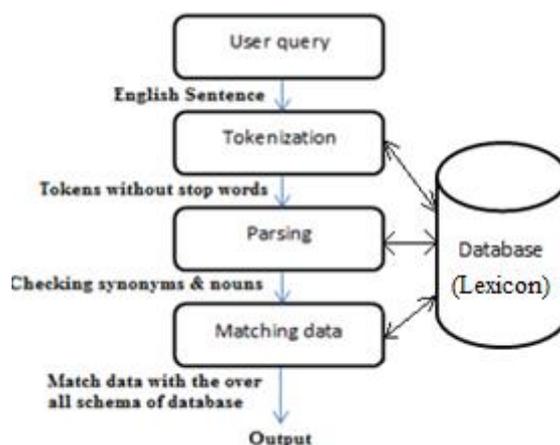
This component translates the natural language into a formal query and generates a natural language response based on the results from the database search.

Database component

It performs Database Management functions. A lexicon database is used to map the words of the natural input onto the language understandable by database. This lexicon is used by all the three phases of the CDINLP. A tokenisation phase takes the formal response as its input, and inspects the parse tree in order to generate adequate natural language response. Natural language database systems make use of knowledge about syntactic and the actual database server in order to properly relate natural language input to the language structured for database. Linguistic component of the system is the area where the syntactic knowledge resides in the syntax analyzer whereas knowledge about the actual database resides to some extent in the semantic data model used. Queries written in interface in natural language are translated into a statement of formal query language. Once the statement unambiguously formed, the query is processed by the database management system in order to produce the required data for the user. These data then passed back to the user in the way that actual query passes in the tubular form or in the way that database server passes.

4. ARCHITECTURE OF THE SYSTEM

Architecture of Common database interface with Natural Language Processing is explained and given below. This architecture is known as CDINLP i.e. Common Database Interface with Natural Language Interface. There are three important phases i.e. Tokenization, Parsing & Matching data.



In the tokenization the tokens of user English query is generated and elimination of stop words are done.

Parsing is the phase where the remaining tokens changed to their synonyms which database can understand. Again the parsing involves one of the phase where foreign words can be identified and these foreign words are called as Nouns. Once the nouns are identified then the last phase come into play and this phase is called Matching data phase.

Matching data phase is the phase where we match the foreign strings or words identified. The query is generated as base on three conditions:

Condition 1: - The query will be generated as

‘Select * from foreign_word’

If matching of foreign_word was traced as the data table.

Condition 2: - The query will be generated as

‘Select foreign_word from tablename’

If matching of foreign_word was traced as tuple of table.

Condition 3: - The query will be generated as

‘Select * from tablename where column name =foreign_word’

If matching of foreign_word was traced as data of any table.

5. METHODOLOGY

To achieve the above mentioned objectives some methodology has been followed and it is given below:

- Create employee database which will store information about employee.
- Identify the nature of queries i.e. select, update, delete, create, order by queries and query builder (Graphical selection of fields or tuples of table) using common database interface.
- Proper mapping of tokens with database keywords should be done by extracting table, columns information from input English sentences.
- Generate SQL query by mapping input query with the stored values in the database tables.
- Successfully execute the query in the form of English.

6. ADVANTAGES

- Common to all database servers/engines.
- It relieves the problem of learning syntax. Hence no need to learn the database languages like SQL or Oracle.
- Remote access to any server via web application by using IP address.
- Physically installation of database servers is optional because user can access the database servers installed on other machines also by using IP address.
- Enables the user to interact with the system who only knows the graphical manipulation.

7. CONCLUSION

This Paper is on Common Database Interface for Relational Database using NLP that accepts the query in simple English Language which is easy to read and understand for the people who do not have any knowledge of SQL language or other query language. It is very much useful for non-technical person to retrieve data from database and get knowledge from it. With the help of this common database Interface for Relational Database using NLP we are world-wise performing all the operation such as select, update, delete, create and moreover the query building by query builder subsystem which implicitly generates the query for the user who is running the CDINLP with graphical selection of the fields and provides result to the user as expected.

The idea is to map a sentence into a logical query language first, and then further translate this logical query language into a general database query language, such as SQL. The Common Database Interface for Relational Database using NLP first converts the simple English language question into an intermediate sensible query. The intermediate logical query expresses the user's question in terms of high caliber world concepts, which are independent of the database structure. The logical query is then translated to an expression in the database's query language, and evaluated against the database.

REFERENCES

- [1] Mrs. Neelu Nihalani, Dr. Sanjau Silkari and Dr. Mahesh Motawani, "Natural Language interface for Database: A brief Review", IJCSI International Journal of Computer Science Issues, vol. 8,no. 2, pp. 600-608, March 2011.
- [2] Androutopoulos, G.D. Ritchie, and P. Thanisch, Natural Language Interfaces to Databases - AnIntroduction, Journal of Natural Language Engineering 1 Part 1 (1995), 29-81
- [3] Rajendra Akerkar and Pritti Sajja, "Natural Language Interface: Question Answering System" in Knowledge based System, Johnes and Bartlett Learning, pp. 333-344, Aug. 2009.
- [4] B. Sujata, S. Viswanadha Raju and Humera Shaziya, "A Servey of Natural Language Interface to Database Management System" International Journal of Science and Advance Technology", vol.2, no. 6,june 2012 .
- [5] Mohit dua, Sandeep Kumar, Zorawar Singh Virak, English Language Graphical User interface to Database Management System" 12th International Conference of Machine Learning and Application 2013.
- [6] IEEE paper Template in A4(v1)-IJERT from www.ijert.org/browse/volume-2-2013/marc-2013 edition.

BIOGRAPHIES



Yawar Abass Mir, is pursuing B.E Degree from SSM College of Engineering & Technology in Computer Science Engineering from University of Kashmir, J&K, India. His field of interest is ASP.Net & SQL.



Zahid Zahoor Koul, is pursuing B.E Degree from SSM College of Engineering & Technology in Computer Science Engineering from University of Kashmir, J&K, India. His field of interest is ASP.Net & SQL.



Syed Irfan, is the Assistant Professor at SSM College of Engineering & Technology in The Department of Computer Science Engineering. I have completed my M.Tech & Currently I am pursuing Phd. His area of interest is Databases, Cloud Computing & Software Engineering.