

## SQL Query Generation Using Human Language Processing

Sarvada Phadtare  
VOGCOE  
Mohili-Aghai, Thane  
*sarwada07@gmail.com*

Kadambari Kadulkar  
VOGCOE  
Mohili-Aghai, Thane  
*kady11395@gmail.com*

Manali Mhatre  
VOGCOE  
Mohili-Aghai, Thane  
*manalimhatre@gmail.com*

Akshay Misal  
VOGCOE  
Mohili-Aghai, Thane  
*Akshaylove21@gmail.com*

Govardhan Jadhav  
VOGCOE  
Mohili-Aghai, Thane  
*govardhanj30@gmail.com*

**Abstract:** This paper proposes a method of querying with the databases by means of a natural language interface. This is not issue in the area of database management is to provide a high level interface for nontechnical users. Normal users are not aware with the formal language like SQL. Then the problem is how they interact with the database system. A normal user may find him/her self-handicapped to deal with the database system. The paper presents an interface module that converts user's query given in natural language into a corresponding SQL command. Asking questions to databases in natural language like English is a very convenient and easy method of data access from database system, especially for normal users who do not understand complicated database query languages such as SQL. This paper proposes the architecture for translating English Query into SQL.

**Keywords:** Databases, Database Management System (DBMS), Structured Query Language (SQL), Natural Language Interface for used. Databases (NLIDB), Natural Language Processing (NLP)

\*\*\*\*\*

### I. INTRODUCTION

Database applications assume an essential part in today's business framework. A large portion of the organizations require these sorts of utilizations by utilizing the SQL dialect. Natural language processing (NLP) is getting to be a standout amongst the most dynamic systems utilized as a part of Human-PC Interaction. It is a branch of AI which is utilized for Information Retrieval, Machine Translation and Language Analysis. The primary objective of NLP is to empower correspondence amongst human and PCs without remembrance of complex Commands and techniques. At the end of the day, NLP is the procedures that can make the PC to comprehend the common dialects utilized by people. Today's prerequisite of business framework is to removing information from a Data Base Management System, for example, MS Access, Oracle and others. A man without learning of SQL may get himself/herself impeded in comparing with the information.

### II. LITERATURE REVIEW

**LUNAR (1972)** is a system that answers questions about samples of rocks brought back from the

moon. In this System the two Databases are used, the chemical analyses and the literature references. This system uses an Augmented Transition Network (ATN) parser and Wood's Procedural Semantics. **RENDEZVOUS (1977)** in this system, users was able to access databases via natural language. In this Codd's system, special emphasis on query paraphrasing and in engaging users in clarification dialogs when there is difficulty in parsing user input. **LIFER/LADDER (1978)** was one of the first good database NLP systems. It was designed as a natural language interface to a database of information about US Navy ships. **CHAT-80 (1980)** is one of the best known NLIDB of the early eighties. CHAT-80 is developed in Prolog language. In which English text is transferred into prolog expressions, which were evaluated against the Prolog database. The code of CHAT-80 was circulated widely and formed the basis of several other experimental NLIDB's. **ASK (1983)** allowed end-users to teach the system new words and concepts at any point during the interaction. ASK was actually a complete information management system providing its own built-in database and ability to interact with multiple external databases electronic mail program and other computer applications. All the applications connected to ASK were accessible to the end-user through natural

language requested

### III. WORKING ON SQL

This printed material uses NLP for interfacing with the Database utilizing normal dialect. In this work just English dialect is utilized as a mean for giving inputs. In this framework we consider a database ORACLE and a default table is utilized which is appropriately standardized. A framework is produced that takes out the issue of ordinary client to associate with database with inflexible dialect SQL. The clients can get to data's by issuing question in basic English dialect. The framework is created in JAVA dialect. The technique embraced for setting this up programming can be grouped through after steps.

**Morphological Analysis:** Individual words are analyzed into their components word tokens such as punctuation is separated from the words.

**Syntactic Analysis:** Linear sequences of words are transformed into structures that show how the words relate to each other.

**Semantic Analysis:** The structures created by the syntactic analyzer are assigned meanings.

**Discourse integration:** The meaning of an individual sentence may depend on the sentences that precede it and may influence the meanings of the sentences that follow it.

**Pragmatic Analysis:** Structure representing what was said is reinterpreted to determine what was actually meant.

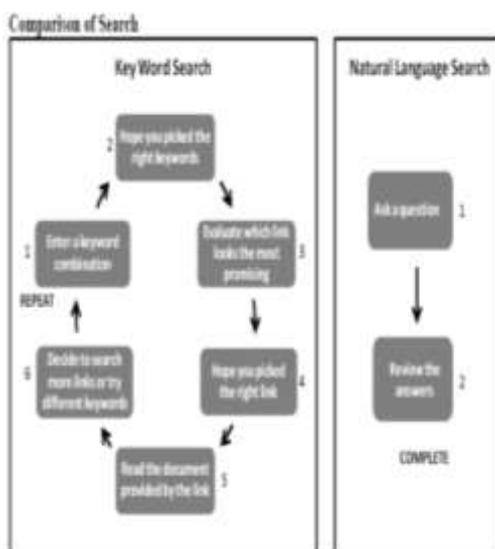


Fig.1 The methodologies used for solving the problem

Firstly, an input query is passed in natural language like English. e.g. English query: What is the Saravjeet's RollNumber.

Then the Morphological analysis are identified each word like what, is, the, Saravjeet's RollNumber. Individual words are analyzed into their components, and it separated noun and adjective in the sentences. Morphological analysis must pull apart the word "Saravjeet's" into the proper noun "Saravjeet" and the possessive suffix "'s". A limited data dictionary is also used to store all related words about the system. After this, Syntactic rules checks the grammatical mistakes of a sentence and Semantic analysis must map individual words into appropriate objects in the knowledge base or database and the meanings of the individual words combine with each other and find out the meaning of simple English query. Example: Meaning of query: RollNumber of student with name Saravjeet.

Then the translator will change the above sentence with SQL query and with the help of SQL query, we will able to find out the results.

SQL Query: Select RollNumber from student where student name = "Saravjeet".

### IV. RESULTS

The user has to first login and then connect to the database. Once the database is setup, a user can input his/her query in simple English language. The system can be used by following these steps:-

- i. Firstly user has to login and enter to the home page of the system.
- ii. Type a query in English language on the text box given in the home page.
- iii. Click on "Execute" button, it will display the results in new page.
- iv. Click on the SQL button, it will display SQL query in new page.
- v. The system will ask the user for the expected meaning. In case of ambiguities the user has to select the desired query.
- vi. Change Password Wizard to be used by Database Administrator to change the password.
- vii. Add New Topic Wizard to be used to add a new table and the columns of the Table in to the Database.



Fig.4.1 login details

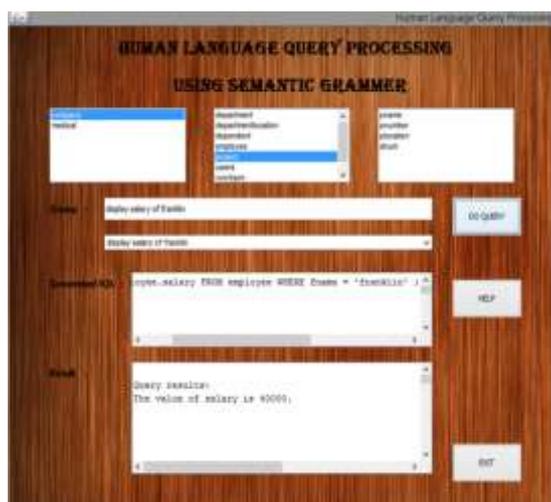


Fig.4.2 query processing using sematic grammar

## V. CONCLUSION

In this work a system is developed that is able to execute both DDL and DML queries, input by the user in his/her natural language (English). The system is developed in JAVA Programming language and various tools of java are used to build the system. An Oracle database is used to store the information's. Input given by the user is not required in the form of questions (who- form like what, who, where, etc). A limited Data Dictionary is used where all possible words related to a particular system are included. The Data Dictionary of the system must be regularly updated with words that are specific to the particular system. Ambiguity among the words will be taken care of while processing the natural language. The results show that our software is correct and handles the SQL Queries without any problem.

## VI. ACKNOWLEDGEMENT

We would like to thank Prof. AmrutaMhatre for helping and guiding us in the proper way, without their guidance it was not possible to successfully complete our paper work.

## REFERENCES

- [1] W.A. Woods, R.M. Kaplan, and B.N. Webber, "The Lunar Sciences Natural Language Information System", Final Report, vol. BBN Report 2378, Bolt Beranek and Newman Inc., Cambridge, Massachusetts, 1972.
- [2] Hendrix, G.G., Sacerdoti, E.D., Sagalowicz, D., Slocum, J. "Developing a natural language interface to complex data", ACM Transactions on database systems, 3(2), pp. 105-147, 1978.
- [3] Androutsopoulos, G.D. Ritchie, and P. Thanisch, "Natural Language Interfaces to Databases – An Introduction", Journal of Natural Language Engineering 1 Part 1, 29–81, 1995.
- [4] M. Templeton and J. Burger, "Problems in Natural Language Interface to DBMS with Examples from EUFID", In Proceedings of the 1st Conference on Applied Natural Language Processing, Santa Monica, California, pp 3–16, 1983.
- [5] C.D. Hafner, "Interaction of Knowledge Sources in a Portable Natural Language Interface", In Proceedings of the 22nd Annual Meeting of ACL, Stanford
- [6] W.A. Woods, R.M. Kaplan, and B.N. Webber, "The Lunar Sciences Natural Language Information System", Final Report, vol. BBN Report 2378, Bolt Beranek and Newman Inc., Cambridge, Massachusetts, 1972.
- [7] Hendrix, G.G., Sacerdoti, E.D., Sagalowicz, D., Slocum, J. "Developing a natural language interface to complex data", ACM Transactions on database systems, 3(2), pp. 105-147, 1978.
- [8] Androutsopoulos, G.D. Ritchie, and P. Thanisch, "Natural Language Interfaces to Databases – An Introduction", Journal of Natural Language Engineering 1 Part 1, 29–81, 1995.
- [9] Elamsri and Navathe, Fundamentals of Database Systems 5<sup>th</sup> Edition, Addison – Wesley, 2007[Chapter 3 and 4].
- [10] Herbert Schildt, The Complete Reference Java, TMH, Fourth Edition, 2002, [570-610].
- [11] Jim Keogh, J2EE – The Complete Reference, Tata McGraw Hill, 2007.
- [12] Ian Sommerville, Software Engineering, Pearson Education Ltd, 2001,[304-316].
- [13] F.Siasardjahantighi, M.Norouzifard, S.H.Davarpanah, M.H.Shenassa "Using Natural Language Processing in Order to Create SQL Queries"- ©2008 IEEE.