ISSN : 2494-9150 Vol-01, Issue 12, MAR 2016.

Natural Language Processing Based SQL Engine

[1]Prof. S. N. Deshpande, [2]Harshada Shinde, [3]Shefali Mane, [4]Ankita Shiriskar
[1]Professor, [2,3,4]BE Student, [1,2,3,4]Computer Engg. Dept. SIGCOE, Koparkhairane, Mumbai, Maharashtra, India.
[2]shindeharshada894@gmail.com, [3]shefali.mane94@gmail.com, [4]ankitashiriskar18@gmail.com

Abstract - Natural Language Processing (NLP) is the technology that evaluates the relationships of words such as actions, entities, or events, comprised within unstructured text, meaning sentences within paragraphs found in a variety of text based documents. Question Answering Natural Language Processing Search is the Natural Language Processing technology that specifically solves the problem of finding answers to a question which can be asked by simply entering it into a search interface using *natural human language.* Using this technique manipulation of the texts is performed for knowledge extraction.

The main purpose of Natural Language Query Processing is for an English sentence to be interpreted by the computer and appropriate action taken. Asking questions to databases in natural language is a very convenient and easy method of data access, especially for casual users who do not understand complicated database query languages such as SQL.

In this paper data keywords are separated into tokens. Tokens synonymous names are stored into database. While parsing text, keywords are separated and based on similar word matching, results are stored into database. Each result is tagged and stored into the database from where the required output can be retrieved according to requirements.

Keywords:* *database, DBMS,. token, NLP, SQL, UML.

I. INTRODUCTION

Natural language processing is becoming one of the most active areas in Human-computer Interaction. The goal of NLP is to enable communication between people and computers without resorting to memorization of complex commands and procedures. In other words, NLP is a technique, which can make the computer understand the languages naturally used by humans.

While natural language may be the easiest symbol system for people to learn and use, it has proved to be the hardest for a computer to master. Despite the challenges, natural language processing is widely regarded as a promising and critically important endeavor in the field of computer research.

Unlike keyword search in Google or Yahoo, Natural Language Processing Question Answering Search specifically allows users to ask questions in their natural language and then retrieves the most relevant answers within seconds. The standard search process requires the execution of multiple keyword combinations that then force the searcher to click on links, which are visited too frequently, if he doesn't find any satisfactory answer, then the process of searching continues until the user finds something or gives up. In Natural Language Processing Search, there is no extra work and no need to search multiple links, resulting in immense time savings. Entering a question is simple for the user even though the technology behind the scenes is highly complex.

In this project, the system will accept user's question as an input. The program will check whether the question is valid or not using Reed-Kellogg syntax function. This function will generate tokens by performing the division of the question clause. The token from the question clause is compared with clauses already stored in the dictionary. Then the algorithm scans the clauses & tries to find utterances most similar to pattern by comparing syntax and semantics.

If both syntax & semantics match, the algorithm starts building the syntax fragment common for both utterances. More syntax nodes have been matched, higher is the matching score. As a result the best answer is shown.

If no utterances are found matching with the pattern then system will display "No answer found". While parsing the input, if the input is not a valid question, the system will display message "Cannot parse question".

II. LITERATURE SURVEY

A. Review Of NLP

Anyone who has used a search engine to perform market, consulting, or financial research, can tell you the pain of spending hours looking for the answer to a seemingly simple question. Add up all the questions a researcher must ask and the hours really rack up[1,2].
Just how big is the search problem? Furthermore an Accenture study found that 50% of information retrieved in search by middle managers is useless."

Money wasted if can't find appropriate result
According to International Data Group the average Knowledge worker makes $60,000 per year out of which $14,000 is spent on search. Knowledge workers spend 24% of their time on search. Here is a quote from Network World, "A company that employs 1,000 information workers can expect more than $5 Million in annual salary costs to go down the drain because of the time wasted looking for information and not finding it, IDC research found last year.

Prior knowledge of complex DBMS languages required
When any person requires data from the database, for which he should possess a prior knowledge about complex database languages like SQL, Oracle, etc. In order to acquire that kind of knowledge, he is forced to learn those complex database languages.

When NLP comes into the picture
Now imagine you are on the phone with your firm's senior risk managers (your boss's boss's boss) and you are asked a question that you don't know the answer to? Imagine if you could type a short question into a search box and come up with an answer in time to provide an intelligent and correct response to the question? That is the power of natural language processing, you type in a question in "natural language" and be provided with an instant result containing the answer that saves the day.

B. Origin Of NLP

In 1950, Alan Turing published his famous article "Computing Machinery and Intelligence" which proposed what is now called the Turing test as a criterion of intelligence. This criterion depends on the ability of a computer program to impersonate a human in a real-time written conversation with a human judge, sufficiently well that the judge is unable to distinguish reliably - on the basis of the conversational content alone - between the program and a real human.[1]

Some notably successful NLP systems developed in the 1960's were SHRDLU, a natural language system working in restricted "blocks worlds" with restricted vocabularies, and ELIZA, a simulation of a Rogerian psychotherapist, written by Joseph Weizenbaum between 1964 to 1966. Using almost no information about human thought or emotion, ELIZA sometimes provided a startlingly human-like interaction.[2]

Up to the 1980's, most NLP systems were based on complex sets of hand-written rules. Starting in the late 1980s, however, there was a revolution in NLP with the introduction of machine learning algorithms for language processing. This was due both to the steady increase in computational power resulting from Moore's Law and the gradual lessening of the dominance of Chomskyan theories of linguistics (e.g. transformational grammar), whose theoretical underpinnings discouraged the sort of corpus linguistics that underlies the machine-learning approach to language processing[3].

In 1990's ,the first evaluation campaign on written texts seems to be a campaign dedicated to message understanding in 1987 (Pallet 1998 Then, the Parseval/GEIG project compared phrase-structure grammars (Black 1991).A series of campaigns within Tipster project were realized on tasks like summarization, translation and searching (Hirschman 1998). In 1994, in Germany, the Morpholympics compared German taggers. Then, the Senseval&Romanseval campaigns were conducted with the objectives of semantic disambiguation. In 1996, the Sparkle campaign compared syntactic parsers in four different languages (English, French, German and Italian). In France, the Grace project compared a set of 21 taggers for French in 1997 (Adda 1999)[4].

C. Problem Definition

Why Natural Language Processing is a Critically Needed Technology?

Time Consumed and wasted while searching process
Anyone who has used a search engine to perform market, consulting, or financial research, can tell you the pain of spending hours looking for the answer to a seemingly simple question. Add up all the questions a researcher must ask and the hours really rack up.
Just how big is the search problem? Furthermore an Accenture study found that 50% of information retrieved in search by middle managers is useless."
Money wasted if can't find appropriate result
According to International Data Group the average Knowledge worker makes $60,000 per year out of which

$14,000 is spent on search. Knowledge workers spend 24% of their time on search. Here is a quote from Network World, "A company that employs 1,000 information workers can expect more than $5 Million in annual salary costs to go down the drain because of the time wasted looking for information and not finding it, IDC research found last year.

Prior knowledge of complex DBMS languages required

When any person requires data from the database, for which he should possess a prior knowledge about complex database languages like SQL, Oracle, etc. In order to acquire that kind of knowledge, he is forced to learn those complex database languages.

When NLP comes into the picture

Now imagine you are on the phone with your firm's senior risk managers (your boss's boss's boss) and you are asked a question that you don't know the answer to? Imagine if you could type a short question into a search box and come up with an answer in time to provide an intelligent and correct response to the question? That is the power of natural language processing, you type in a question in "natural language" and be provided with an instant result containing the answer that saves the day.

III. SCOPE OF THE PROJECT

The scope of the proposed system is as follows:

To work with any RDBMS one should know the syntax of the commands of that particular database software (Microsoft SQL, Oracle, etc.).

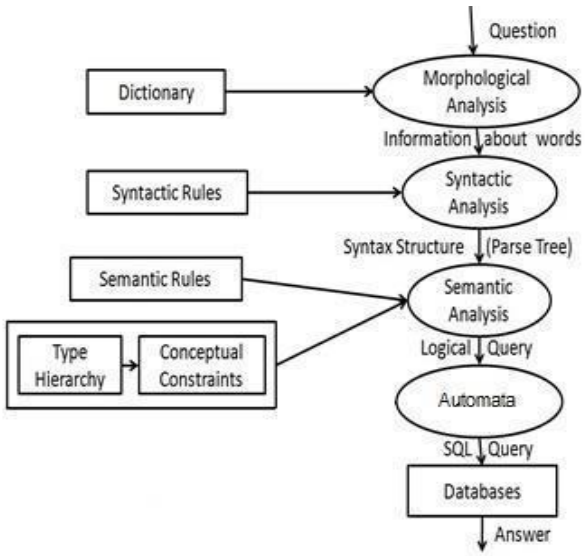
Here the Natural language processing is done in English i.e., the input statements have to be in English language. Input from the user is taken in the form of questions ("WH form" like what, who, where, etc).

A limited Data Dictionary is used where all possible words related to a particular system will be 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 system must be stable and can be operated by people with average knowledge.

IV. PROPOSED SYSTEM



A. Algorithm Of NLP

The working of algorithm is explained with below example.

Text file:

Employee related sentences specifying id name designation etc.

Step1

Input from the User to Natural Language Query Processor

"Wh" question related to employee databaseor questions with the keywords like list,give,display

Step2: Once given input question, the program actually checks for a valid question.

it checks whether we have inputted a proper question.

Step 3:

If there are no utterances in sentence then it will display "No answer found"

Step4:

Syntax graphs are matched

This is syntax graph for the question "What is the maximum salary of employee"

Step5:

If syntax nodes match, then meanings of words associated with syntax nodes are compared

Here meanings of the words mean "Lexemes" and lexemes mean, if employee is noun in questions, and if it is noun in input text then river lexeme is matched.

Step6:

If both syntax and meanings are equal, and if the utterance are considered to be equal, then matching score is incremented.

The more the matches of lexemes, the more the score and the more score gets the output answer. Once it has found matches we will have the output.

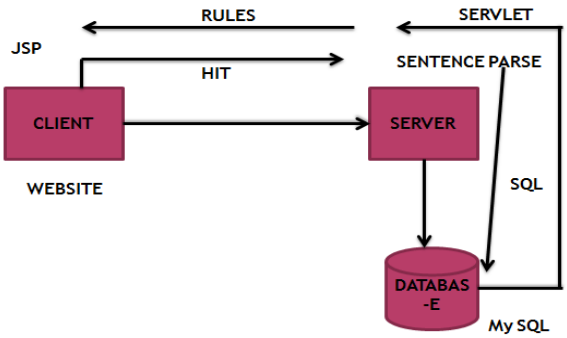


Fig 1 : Working of Natural Language Processing

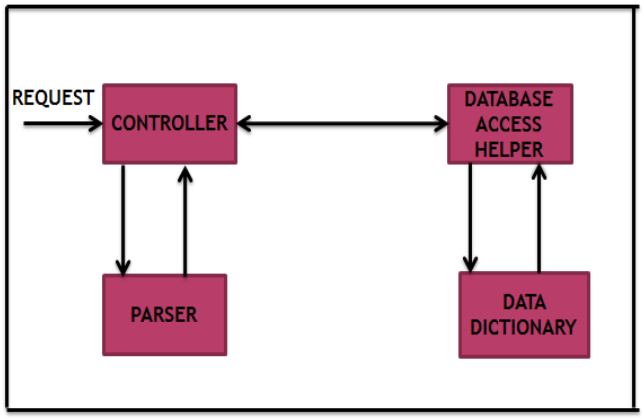


Fig 2 : Architecture Of NLP

V. EXPECTED RESULT

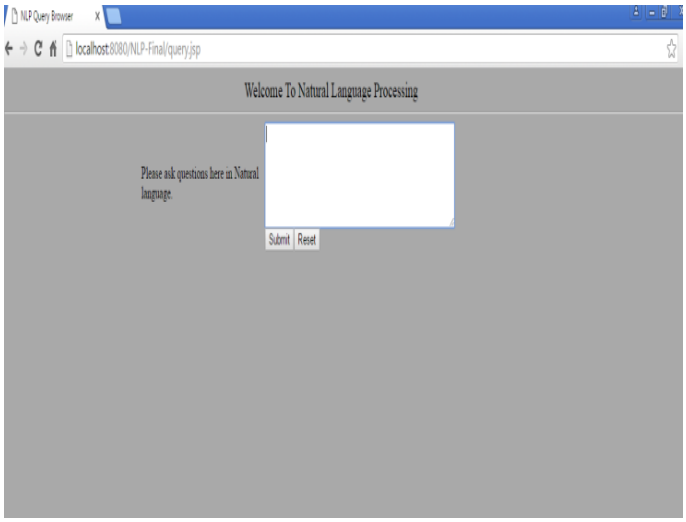


Fig 3 : WebPage for NLP

```
NLPController.java  Analytic.java  DBManager.java  query.jsp
110
111             ////////////////////////////start select clause //////////////////
112             if(!nounList.isEmpty() && !fromList.isEmpty()){
113                 for (String from : fromList) {
114                     List<String> columnList = DBManager.dictionary.get(from);
115                     for (String noun : nounList) {
116                         if(columnList.contains(noun)){
117                             selectList.add(noun);
118                             removeList.add(noun);
119                         }
120                         for (String column : columnList) {
121                             if(noun.contains(column) && !selectList.contains(column)){
122                                 selectList.add(column);
123                                 removeList.add(noun);
124                             }
125                             if(column.contains(noun) && !selectList.contains(column)){
126                                 selectList.add(column);
127                                 removeList.add(noun);
128                             }
129                         }
130                     }
131                 }
132             }
133             nounList.removeAll(removeList);
134             removeList.clear();
135             ////////////////////////////end select clause //////////////////
136         }
137
138         ////////////////////////start select function //////////////////
139         if(!adjList.isEmpty()){
140             for (Map.Entry<String, String> entry: agregateMap.entrySet()) {
141                 for (String adj : adjList) {
142                     if(entry.getKey().contains(adj)){
```

Fig 4 : Code for Select Query

```
NLPController.java  Analytic.java  DBManager.java  query.jsp
136         }
137
138         ////////////////////////start select function //////////////////
139         if(!adjList.isEmpty()){
140             for (Map.Entry<String, String> entry: agregateMap.entrySet()) {
141                 for (String adj : adjList) {
142                     if(entry.getKey().contains(adj)){
143                         function=entry.getValue();
144                     }
145                 }
146             }
147         }
148
149         ////////////////////////end select function //////////////////
150
151         ////////////////////////select * if select list empty //////////
152         if(selectList.isEmpty()){
153             for (String noun : nounList) {
154                 for (String string : starList) {
155                     if(noun.contains(string) && !selectList.contains("*")){
156                         selectList.add("*");
157                         break;
158                     }
159                     if(string.contains(noun) && !selectList.contains("*")){
160                         selectList.add("*");
161                         break;
162                     }
163                 }
164             }
165         }
166
167         ////////////////////////select * if select list empty //////////
168
```

Fig 5 : Code for Select Function

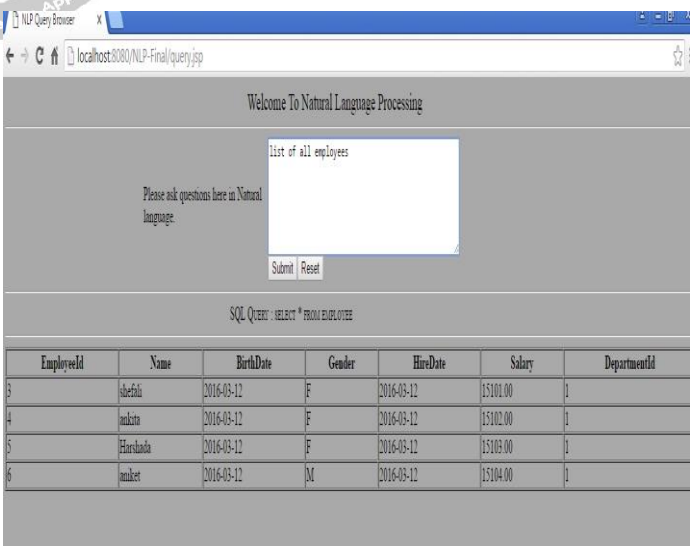


Fig 6 : Query for Employee Details

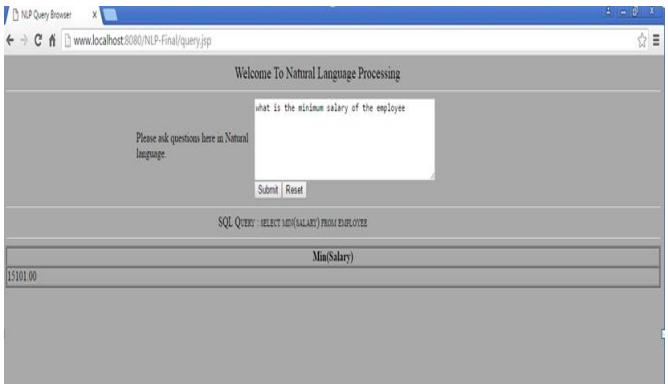


Fig 7: Query for Minimum salary of Employee

VI. CONCLUSION

The main objective of this paper was to throw some light on basic architecture Natural Language Processing, its use in SQL query generation. We planned our project and developed the feasibility report. The project is an economically and technically feasible project. The system developed will be beneficial to the many users which are not familiar with SQL language but want to access the data from the database. By providing an expert system, we are encoding hidden mystery of natural language; The fact that common words tend to have multiple meanings can lead to ambiguity, the expert system can maintains database that represents the state of the world by looking at the context surrounding the sentences and receives the best recognized from the text. We collect the required knowledge for this system from an individual who is experienced in natural language analysis, and embed this knowledge into an expert system as a knowledge base. It finds the most similar entity name to the terms of input sentence based on searching this knowledge base. This paper is presenting the result of using an expert system beside common existing solutions for transforming natural language expressions to SQL query language. Result shows this process can be completely automated.

REFERENCES

[1] J. U. Duncombe, "Infrared navigation—Part I: An assessment of feasibility (Periodical style)," *IJREAM Trans. Electron Devices*, vol. ED-11, pp. 34–39, Jan. 1959.

[2] S. Chen, B. Mulgrew, and P. M. Grant, "A clustering technique for digital communications channel equalization using radial basis function networks," *IJREAM Trans. Neural Networks*, vol. 4, pp. 570–578, Jul. 1993.

[3] R. W. Lucky, "Automatic equalization for digital communication," *Bell Syst. Tech. J.*, vol. 44, no. 4, pp. 547–588, Apr. 1965.

[4] S. P. Bingulac, "On the compatibility of adaptive controllers (Published Conference Proceedings style)," in *Proc. 4th Annu. Allerton Conf. Circuits and Systems Theory*, New York, 1994, pp. 8–16.

[5] G. R. Faulhaber, "Design of service systems with priority reservation," in *Conf. Rec. 1995 IJREAM Int. Conf. Communications*, pp. 3–8.

[6] W. D. Doyle, "Magnetization reversal in films with biaxial anisotropy," in *1987 Proc. INTERMAG Conf.*, pp. 2.2-1–2.2-6.

[7] G. W. Juette and L. E. Zeffanella, "Radio noise currents n short sections on bundle conductors (Presented Conference Paper style)," presented at the IJREAM Summer power Meeting, Dallas, TX, Jun. 22–27, 1990, Paper 90 SM 690-0 PWRS.

[8] J. G. Kreifeldt, "An analysis of surface-detected EMG as an amplitude-modulated noise," presented at the 1989 Int. Conf. Medicine and Biological Engineering, Chicago, IL.

[9] J. Williams, "Narrow-band analyzer (Thesis or Dissertation style)," Ph.D. dissertation, Dept. Elect. Eng., Harvard Univ., Cambridge, MA, 1993.

