Natural Language (English) To MongoDB Interface

Mahesh D. Gadekar, Bhagwan M. Jadhav, Ansar S. Shaikh, Raju B. Kokare

***Abstract*— As today's world is moving from offline to online. All the process now are been done with the help of computer's, any data we require is present on the internet so almost all IT applications are storing into and retrieving information from Databases. Retrieving information from the database requires knowledge of technical languages such as NON Structured Query Language (NOSQL). The main purpose of Natural Language Query Processing is for an English sentence to be interpreted by the computer and appropriate action taken. Asking question to databases in natural language is a very convenient and easy method of data handle, especially for casual users who do not understand complicated database query languages such as NOSQL. Various analyses shows user is not restricted to formulate any kind of query so this system provides result to users any type of query he fires to the system accurately and efficiently, even if any user make spelling mistake the system will autocorrect the spelling and experimental result show spell checking technique is efficient than Microsoft spell checker.**

Index Terms— **Query Mapping, Spelling Error Correction**

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. The applications that will be possible when NLP capabilities are fully realized are impressive computers would be able to process natural language, translating language accurately and in real time information from a variety of data sources, depending on the users requests. The main goal of this system is to provide communication between human and computer without recalling any sort of database DDL or DML query syntax. The more general goal of such search interface or NLP system is to make the computer able to understand the natural language so that user can address the system as if they are addressing some other person and get the expected result they want.

This paper proposes a system which gives interface between user and computer, in the form of database query language. Also the spelling corrections of misspelled words in query are getting correct. First we summarize some classic NLDBI system. Consequently we discuss the overall system architecture of the natural language database interfaces, some implementation details and experimental results.

Manuscript received March, 2015.

Mahesh D. Gadekar, *Computer Department, PGMCOE Wagholi, Pune City Pune, Country India, Mobile No +919766896070*

Bhagwan M. Jadhav, *Computer Department, PGMCOE Wagholi, Pune. City Pune, Country India, Mobile No +918796580655*

Ansar S. Shaikh, *Computer Department PGMCOE Wagholi, Pune City Pune, Country India, Mobile No +919545031659*

Raju B. Kokare, *Computer Department, PGMCOE Wagholi, Pune City Pune, Country India, Mobile No +919764383487*

A.Word Pair Mining Technique

Search session consist of set of user queries fired by a single user within a short time period. Many of the search session consisted of misspelled word and there corrected spelling. We segment the query stream from user into sessions. If the time between the two queries exceeds a certain period of time then we put session boundary between the one or more query. The time period between the two queries is kept short, the reason behind this that it is observed that user firing query is corrects the misspelled word immediately after recognizing the mistake. The following table shows some example of correct and misspelled word.

Misspelled Word	**Correct Word**
Newpepa	Newspaper
Boke	Book
Collage	College
Colmn	Column

II. RELATED WORK

The very first NLP system for database system is as old as any of the NLP research. Some of the related systems till date are given below that provided database interface for users:

A.Lunar

Woods built one of the first natural language question answering systems (LUNAR) to answer questions about the Apollo 11 moon rocks for the NASA Manned Spacecraft Center while he was at Bolt Beranek and Newman (BBN) in Cambridge, Massachusetts. The program used an Augmented Transition Network (ATN) parser and Woods' Procedural Semantics.

B.Lifer/Ladder

It was one of the first good database NLP systems. It was designed as a natural language interfaces to a database of information about US Navy ships. This natural language interface system, as described in a paper by Hendrix (1978), used a semantic grammar to parse questions and query a distributed database. The LIFER/LADDER system can only support simple one-table queries or more than one table queries with easy join conditions.

C.Chat-80

The system CHAT-80 is one of the most referenced NLP systems in the eighties. The system was implemented in Prolog language. The CHAT-80 was an quite impressive, efficient and sophisticated system.The database of CHAT-80 consist of facts about 150 of the countries world and a small set of English language vocabulary that are enough for querying the database.

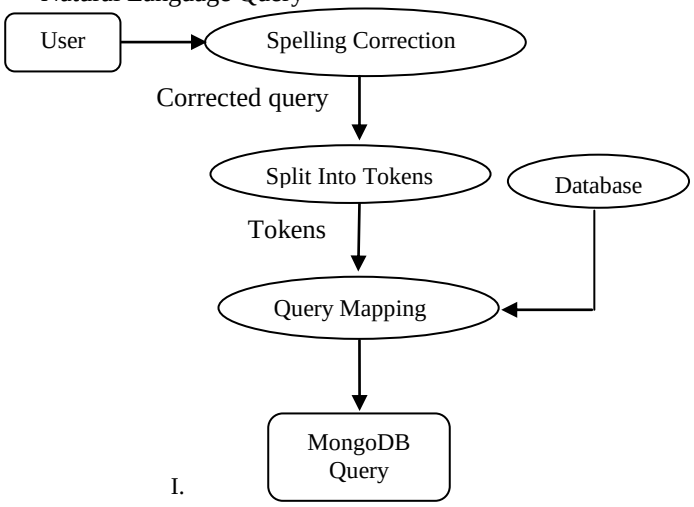
III. PROPOSED WORK

A.System Description

A brief description of the system is given. Let us consider a database say MongoDB. Within this MongoDB database we have stored some collections which are properly normalized. Now if the user wish to access the data from the collection, user has to be technically sound in the NOSQL language to make a query for the MongoDB database. We will perform the following steps for the transformation of query from natural language to database query (NOSQL) sequentially as listed in the following points:

- First we will accept the string in natural language.
- After accepting the string we will check the query for misspelled words (if any) using word pair mining.
- After that we will split query into tokens.
- After getting tokens we perform the NOSQL mapping for transformation

The above process will be as shown in following figure:



I.

Fig. 1 System Architecture

B.Spelling Correction

In the above process the spelling correction of misspelled word made by the user while entering the queries is done with the help of word pair mining technique that is performed as follows:

1. The queries entered by the user will have same number of word
2. The difference will be of just a single word in the above two queries.
3. For the word pair the word in the first query will be misspelled and the second will be considered as corrected give frequency to each word pair.
4. After this we will be discarding the low frequency word pairs.

The word pair in the above implementing technique will mine from search session at Bing.

C.Query Mapping

After getting the correct query we will tokenize the query, after getting the tokens we will extract the meaningful tokens from the query that will be required for the mapping or transforming the query in database query. The meaningful tokens resembles to the token that match with any of the columns, values, any aggregate function etc. that is entered by user. After extracting the meaningful tokens we than map the NLQ to database query according to our algorithm the following example will make much more clarification about the meaningful tokens.

E.g.:
User Query- name of user having status A
Meaningful Tokens-name (column), status (column), A (field)
After getting the meaningful tokens the query will be transformed in MongoDB query by putting all these tokens in the
MongoDB Syntax. It is as shown in the following example:
After getting the meaningful tokens
MongoDB Query- db.user.find(
{ status: "A" }
{ user_id:1, status:1, _id=0 })

Even if the user makes some grammatical errors in query then the system will provide the grammatically correct suggestions.

IV.FUTURE SCOPE

In the above section we discussed about the proposed system for Natural Language Processing. In this section we will discuss the future scope of the system. As discussed proposed system converts the Natural Language query in NOSQL query, also it corrects the spelling automatically. In future we will be focusing towards reformulation of query and transforming the Natural Language Query in the SPARQL database language required for the semantic search to provide the more accurate result as the user require instead of checking the multiple link as we do at present. Today's generation also use lots of abbreviation so keeping it in mind we will be reformulating the query (for e.g. LA by Los Angeles and TOI by Times Of INDIA) so in the future we will try to implement the above discussed goals.

V.CONCLUSION

In this paper we proposed the search interface that will be applicable for the online applications, provides easy to the user by reducing their part of recalling complex database language syntax. Natural Language Processing can bring powerful enhancements to virtually any computer program interface, because language is so natural and ease to use for humans. The Intelligent Tutoring System, NLDBI is no exception. Alternatives for integrating a database NLP component into the NLDBI were considered and assessed.

REFERENCES

[1] Woods, W.A. et al, "*The Lunar Sciences Natural Language Information System*", BB&N Report 2378, June 1972.
[2]Hendrix, G.G., Sacerdoti, E.D., Sagalowicz, D., Slocum, J. "*Developing a natural language interface to complex data*", in ACM Transactions on database systems, 3(2), pp.105-147, 1978.
[3] Linguistic Technology. English Wizard – Dictionary Administrator's Guide. Linguistic Technology Corp.Littleton, MA, USA, 1997.
[4] Hendrix, G. (1977). "*The LIFER manual a guide to building practical natural language interfaces*". SRI Artificial Intelligence Center, Menlo Park, Calif. Tech. Note 138.
[5] Warren, D., Pereira, F. (1982). "*An efficient and easily adaptable system for interpreting natural language queries in Computational Linguistics*".
[6]Huangi, Guiang Z., Phillip C-Y S. "*A Natural Language database Interface based on probabilistic context free grammar*", IEEE International workshop on Semantic Computing and Systems 2008.
[7] Mrs. Gauri R."*Natural Language Query Processing Using Semantic Grammar*". / (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 02, 2010, 219-223
[8] ELF Software CO. "*Natural-Language Database Interfaces from ELF Software Co*", cited November 1999, available from Internet:
[9] Ana-Maria Popescu, Alex Armanasu, Oren Etzioni, David Ko, and Alexander Yates, "*Modern Natural Language Interfaces to Databases:Composing Statistical Parsing with Semantic Tractability, COLING (2004*)".
[10] "*Natural language Interface for Database: A Brief review*", Mrs. Neelu Nihalani, Dr. Sanjay Silakari, Dr.Mahesh Motwani. IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 2, March 2011 ISSN (Online): 1694-0814

