Available online at http://www.journalcra.com

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

International Journal of Current Research
Vol. 9, Issue, 09, pp.57575-57577, September, 2017

ISSN: 0975-833X

RESEARCH ARTICLE

QUESTION-ANSWERING SYSTEM USING NATURAL LANGUAGE PROCESSING WITH NLIDB APPROACH

Prof. Pooja Malhotra, *Yash Kapadia, Krishna Saboo and Ankita Sarda

K.J. Somaiya College of Engineering, Vidyavihar-77

ARTICLE INFO

Article History:
Received 14th June, 2017
Received in revised form
25th July, 2017
Accepted 22nd August, 2017
Published online 29th September, 2017

Key words:

Natural language processing (NLP),
Natural language interface for database (NLIDB), SQL,
Question-answering.

ABSTRACT

With increasing information, lot amount of data is stored in relational databases in organized manner. Structured query language (SQL) is primarily used for extracting information from a relational database and allows us to manage and access process on database. But this is insufficient as not all users are accustomed to write SQL queries if the user question is in Natural language (NL). Hence there is a need to develop a system where non-sophisticated users can access data freely without going into technicality by asking question in Natural language. This paper focuses on creating semantic analyzer for automatic Question-answering system for domain specific database. It provides user with the relevant answers to the user questions using Natural Language processing (NLP) and Natural language interface for database (NLIDB). NLIDB are the systems that translate a natural language sentence into a database query. It contains the stepwise description on conversion of question to simple SQL query without using any clauses. It portrays completely automatic, reliable, fast way to query a database. Hence, Natural language processing techniques applied on English text is converted to SQL query using series of steps like lowercase conversion, tokenization, chunking, generation of SQL query and mapping the query to the database.

Citation: Prof. Pooja Malhotra, Yash Kapadia, Krishna Saboo and Ankita Sarda, 2017. "Question-answering system using natural language processing with NLIDB approach", *International Journal of Current Research*, 9, (09), 57575-57577.

INTRODUCTION

Databases are one of the major sources of Information. With the increase in large chunks of data and need for easy and fast retrieval of data, information is stored in databases is relational model and then can be retrieved by querying the database using Structured Query Language (SQL). Hence SQL acts as hotspot for retrieval of information from database. But this is inconvenient for users with no knowledge of generation of these complex SQL queries. Artificial Intelligence (AI) and Linguistics can be combined to develop programs that can help to understand and produce information in a natural language. These information retrieval systems can be used in various educational and informational institutes, in companies to manage and retrieve information. NLIDB approach is used where the information seeker uses natural language for submitting the query and retrieve the results from databases, hence this save the user to master the database language/schemas formulate the queries. NLIDB uses various processes to transfer the raw English text to the database query. This query is then fired on database for efficient information retrieval. The process description used in the paper is as follows:

***Corresponding author:* Yash Kapadia,**
K.J. Somaiya College of Engineering, Vidyavihar-77

Thus NLIDBS are built to optimize the search results and produce information with more accuracy. The present research extends the existing work further by processing more complex queries along with ambiguity removal. Here, we will study how to preprocess the user question and the preprocessed question is passed to the lexical analyzer where the question is converted to the appropriate SQL query using production rules. And finally this generated query is mapped to the knowledge base for information retrieval.

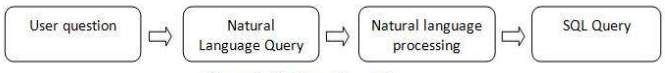


Figure 1: Problem Description

System Implementation

Main components of a question-answering system are: lexical analyzer and NLP to SQL conversion. After the preprocessing of the user question which includes tokenizing, lower case conversion, escape word remover, removing ambiguous attributes, lexical analyzer does the further processing by generating various production rules which converts the user question to the relevant SQL query. Finally, it maps the generated SQL query to the relational Database. (Garima Singh and Arun Solanki, 2016)

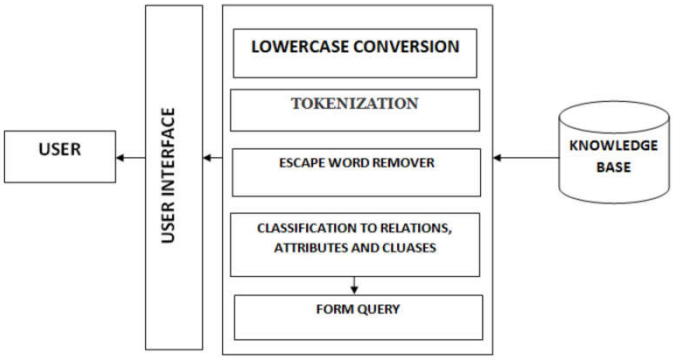


Figure 2: System Implementation

Table 1 Escape words list.

A	An	The	Select
find	which	whose	Is
Of	A	With	To
for	Are	And	What

Example: List of all the students from IT department

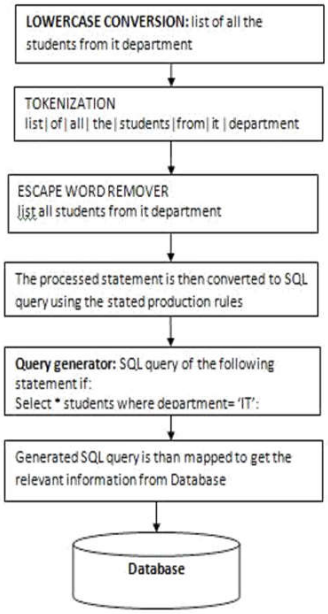


Figure 3: Processing of User query

Table 2: Rules of attribute in relation Student

Rules	Rule Description
Name	Attribute for relation 'student'
Email id	Attribute for relation 'student'
Date of birth	Attribute for relation 'student'
Address	Attribute for relation 'student'
Department	Attribute for relation 'student'
Mobile number	Attribute for relation 'student'

Table 3: Rules of attribute in relation Faculty

Rules	Rule Description
Name	Attribute for relation 'student'
Email id	Attribute for relation 'student'
Branch	Attribute for relation 'student'
Position	Attribute for relation 'student'

Example

Input: List of all the students

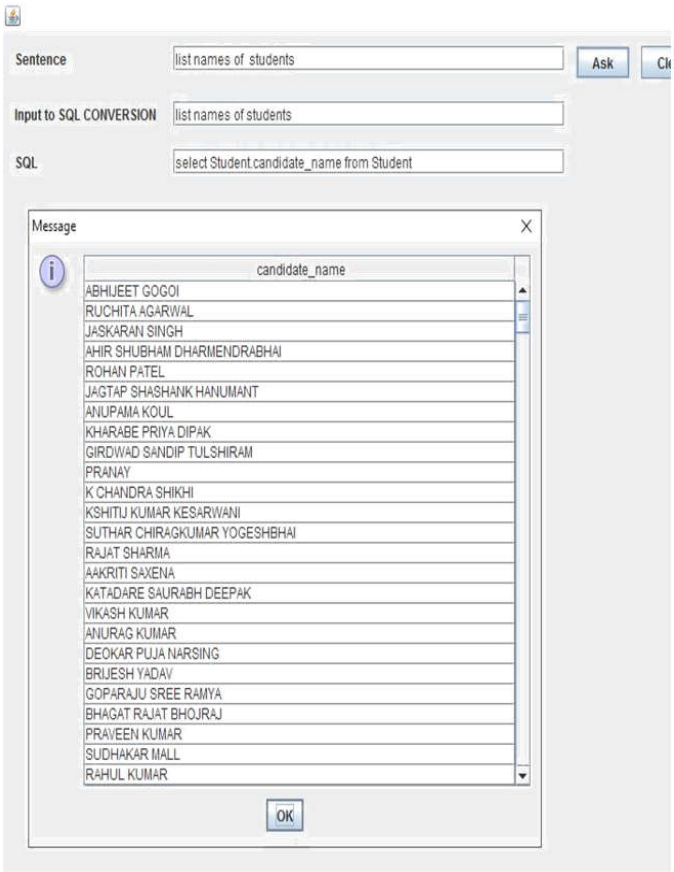
Production Rule:

```
list {operation="select";}
students | student {table_name="Student";}
all {attribute_name="*";}
names {attribute_name="Student.candidate_name";}
\n        {}
.         {}
```

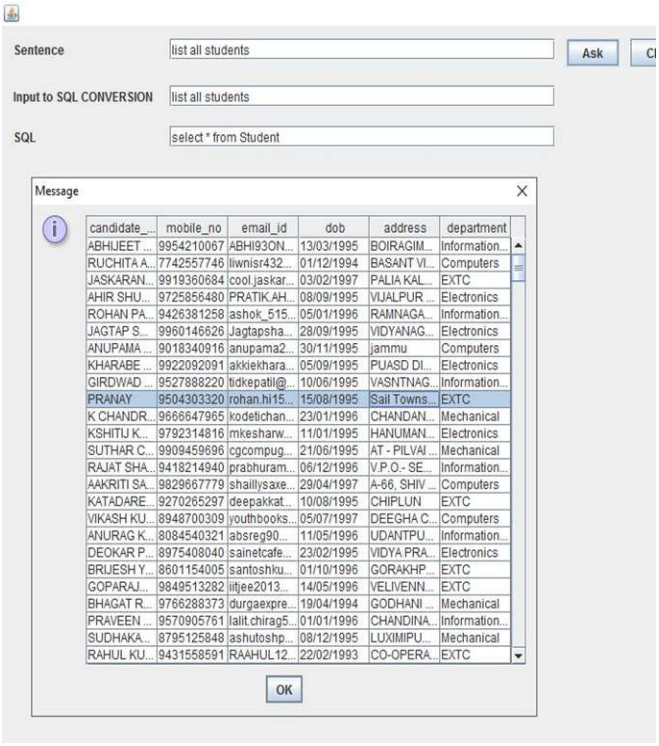
Output: Select * from students

Results and final output

Output 1:



Output 2



candidate_...	mobile_no	email_id	dob	address	department
ABHIJEET ...	9954210067	ABHI93ON...	13/03/1995	BOIRAGIM...	Information...
RUCHITA A...	7742557746	liwnisr432...	01/12/1994	BASANT VI...	Computers
JASKARAN...	9919360684	cool.jaskar...	03/02/1997	PALIA KAL...	EXTC
AHIR SHU...	9725856480	PRATIK.AH...	08/09/1995	VIJALPUR ...	Electronics
ROHAN PA...	9426381258	ashok_515...	05/01/1996	RAMNAGA...	Information...
JAGTAP S...	9960146626	Jagtapsha...	28/09/1995	VIDYANAG...	Electronics
ANUPAMA ...	9018340916	anupama2...	30/11/1995	jammu	Computers
KHARABE ...	9922092091	akkiekhara...	05/09/1995	PUASD DI...	Electronics
GIRDWAD ...	9527888220	tidkepatil@...	10/06/1995	VASNTNAG...	Information...
PRANAY	9504303320	rohan.hi15...	15/08/1995	Sail Towns...	EXTC
K CHANDR...	9666647965	kodetichan...	23/01/1996	CHANDAN...	Mechanical
KSHITIJ K...	9792314816	mkesharw...	11/01/1995	HANUMAN...	Electronics
SUTHAR C...	9909459696	cgcompug...	21/06/1995	AT - PILVAI ...	Mechanical
RAJAT SHA...	9418214940	prabhuram...	06/12/1996	V.P.O.- SE...	Information...
AAKRITI SA...	9829667779	shaillysaxe...	29/04/1997	A-66, SHIV ...	Computers
KATADARE...	9270265297	deepakkat...	10/08/1995	CHIPLUN	EXTC
VIKASH KU...	8948700309	youthbooks...	05/07/1997	DEEGHA C...	Computers
ANURAG K...	8084540321	absreg90...	11/05/1996	UDANTPU...	Information...
DEOKAR P...	8975408040	sainetcafe...	23/02/1995	VIDYA PRA...	Electronics
BRIJESH Y...	8601154005	santoshku...	01/10/1996	GORAKHP...	EXTC
GOPARAJ...	9849513282	iitjee2013...	14/05/1996	VELIVENN...	EXTC
BHAGAT R...	9766288373	durgaexpre...	19/04/1994	GODHANI ...	Mechanical
PRAVEEN ...	9570905761	lalit.chirag5...	01/01/1996	CHANDINA...	Information...
SUDHAKA...	8795125848	ashutoshp...	08/12/1995	LUXIMIPU...	Mechanical
RAHUL KU...	9431558591	RAAHUL12...	22/02/1993	CO-OPERA...	EXTC

Lexical Analyzer

Lower case conversion: User input is converted to lower case irrespective of the input statement. Hence it marks all the text to lower case and helps in minimizing the number of production rule.

Tokenization: Process of breaking a stream of text into words, phrases, and other meaningful elements called as tokens. The generated statement with token is used as input to the further processing in production rules.

Escape word remover: The extra/stop words are removed which are not needed in the analysis of query. The following are the words selected from sentence and removed in escape word remover

Production Rules:

Conversion of user query to SQL

Production rules are primarily pre-described set of rules and behavior. The production is said to be triggered. Hence the executed production rule converts the user statement into SQL query and then is fired on database. (Garima Singh and Arun Solanki, 2016) Production rule consist of two parts i.e. a precondition (IF condition) and action (Then condition). Hence the rules specify the condition-action behavior of user query. (Sangeeth and Rejimoan, 2015)

Conclusion

Natural Language Processing can change the complete working of the computer programming interface. The proposed system is designed to handle the challenges in Natural language processing and make it more robust and flexible for all kinds of queries respective to its domain. It uses semantic matching technique to translate the NL question to the relative SQL query. Various steps such as lower case conversion, tokenization and ambiguity remover are used to convert to SQL query which is mapped to the database to get the required information.

Goals and future work

Goal is to increase its accuracy and make it robust, hence allowing it to handle more difficult and challenging questions and increasing the scope of database. Further complex queries that can be considered are queries that make use of GROUP BY and HAVING clauses. The queries containing GROUP BY clause gather all the rows with some common specification and the HAVING clause allows to specify conditions that filter which group of results appear in the Final results. Following system can be added to our college website for easy searching and information retrieval and hence making it more handy.

REFERENCES

An algorithm to transform natural language into SQL queries for relational databases by Garima Singh, Arun Solanki Department of Computer Science and Engineering, Gautam Buddha University, Greater Noida, India, 2016.

An Intelligent system for information extraction from relational database using HMM by Sangeeth N. and Rejimoan R. 2015.

Androutsopoulos, G.D. Ritchie, and P. Thanisch, "Natural Language Interfaces to Databases - An Introduction, Journal of Natural

Conversion of natural language query to SQL by Prabhdeep Kaur1 and Shruthi J2 1Student, 2Assistant Professor, Department of Computer Science & Engineering, BMS Institute of Technology, Bangalore, India

Hendrix, G., Sacrdoti, E., Sagalowicz, D. and Slocum, J. 1978. "Developing a natural language interface to complexdata." ACM Transactions on Database Systems, Volume 3, No. 2, USA, Pages 105 – 147.

http://www.cs.utexas.edu/users/m llnldata/geoquery.html.

Language Engineering 1 Part 1" volume-I, pp 29-81, 1995.

Parse Tree Database for Information Extraction by Luis Tari, Phan Huy Tu, J¨org Hakenberg, Yi Chen, Tran Cao Son, Graciela Gonzalez and ChittaBaral

Rohit Agrawal, Amogh Chakkarwar, Prateek Choudhary, Usha A. Jogalekar, and Deepa H. Kulkarni, "DBIQS - An Intelligent System for Querying and Mining Databases using NLP", International Conference on Information Systems and Computer Networks, pp 39-43, 2014.

Semantic grammar. International Journal on Computer Science and Engineering, 2(2): 219-223 Rukshan A, Rukshan P, Mahesan S. 2013. Natural Language Web Interface for Database (NLWIDB).

Sreenivasulu M. 2014. Information retrieval using natural language interfaces. International Journal of Computer Applications, 92(12):43-37

****