International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

Database Access for Non-technical Users using ML

Mr. Yogeshwar J. Shirsath[1] and Dr. Rachna Somkunwar[2]

Computer Engineering, Dr. D.Y. Patil Engineering, Pimpri, Pune, India

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

This system has been developed to enable secure access of data to a voice-based user interface by enabling voice-based authentication and integration with an existing Natural Language Processing system. We address the question of how to improve the fetching the results from query results. conventional predefined query forms aren't capable of satisfy diverse advert-hoc queries from users on those databases. Here, we propose a machine learning-based technique to generate the SQL query based on the user's voice, a novel database query form interface, which can dynamically generate query forms.

Keywords: *NLP, Languages and compilers, Optimization, Verification, Voice Recognisation, Machine-independent microcode generation*

Introduction

Natural Language Processing is an area of application and research that explores how computers can be used to understand and manipulate tongue speech or text to do helpful things. The foundation of NLP lies in many disciplines, namely, computer and knowledge sciences, linguistics, mathematics, electrical and electronic engineering, AI robotics, and psychology. NLP researchers aim to gather knowledge on how human beings use and manipulate natural languages to perform desired tasks so that appropriate tools and techniques can be developed. Applications of NLP include many fields of study like multilingual and cross-language information retrieval, machine transaction, language, text processing and summarization, user interfaces, speech recognition, computing, and expert systems

While the natural language may be the easiest system for people to learn and use, it has proved to be the hardest for a computer to understand. The goal of NLP is to enable communication between people and computers without resorting to the memorization of procedures and complex commands.

In other words, NLP may be a technique, which may make the pc understand the languages naturally employed by humans. In this project, we are translating English query into an SQL query using semantic grammar. The system will accept users' query in natural language as an input. The program will check whether the query is valid.

Then we'll generate tokens by performing the division of the question clause. Each token represents a single word in the user's query. The tokens from the query clause are compared with clauses already stored within the dictionary. The dictionary needs to be constantly updated. Then the algorithm scans the tokens and tries to seek out attributes present within the query. Then we discover all the tables within the database which contain the attributes by comparing syntax and semantics. Then we build the ultimate SQL query and execute it on the database and return the result dataset to the user.

II. LITERATURE SURVEY

Natural language processing can be done in two way communication with device one is written communication, as well as verbal communication with device written communication, is much easier than verbal communication. In written communication syntax, semantic, morphological and lexical analysis. Whereas in verbal communication includes all the processes in written as well as additional process include additional knowledge about phonic as well as enough added information to handle the further ambiguities that arise in speech[1].

This paper places interest in some emerging capabilities for incremental speech understanding and processing in virtual human dialogue systems. This work is a part of an in-progress effort that intent to enable practical spoken dialogue with virtual humans in multiparty arbitration scenarios. These scenarios are designed to allow the trainees to practice their

intervention skills by engaging in face-to-face spoken negotiation with one or more virtual humans. A very important think about achieving naturalistic behavior in these arbitration scenarios, which ideally should have the virtual humans representing composite reasoning, fluid turn-taking, and responding to factors like emotions and trust, is for the virtual humans to begin to understand [2].

The current custom in virtual human dialogue systems is to use limited-domain speech synthesis or skilled human recordings. Both approaches lead to a good show but at an elevated cost. To determine the simplest trade-off between performance and price , we evaluate human and synthesize voices concerning naturalness, likability, and conversational aspect. Varying the sort , length, and content of utterances, and take under consideration the age and language of ratters also as their expertise with speech synthesis. The results suggest that a professional human voice can surpass both synthesized voices and an amateur human voice. Also, a high-quality general-purpose voice or an honest limited-domain voice can execute better than parttime human recordings. As expected, in most cases, the high-quality general-purpose voice is rated above the limited-domain voice. There is also a non-statistically significant trend that has been observed for long or negative utterances to receive low ratings [3].

This paper aim to explore business applications of chatbots, also on propose several extent metrics to guage practice, usability and overall quality of an embodied conversational agent. Based on these metrics we examine existing Polish-speaking commercial chatbots that, firstly, work in the B2C subdivision. Secondly, reach the widest possible range of users. And lastly, they are most probably the most advanced commercial deployments of their creators. The system analyses various aspects of functioning of each personified conversational agent: optical look, sort of operation on the web site , speech amalgamation unit, inbuilt knowledge domain , presentation of data and supplementary functionalities, conversational abilities and perspective sensitiveness, personalization options, personality traits, emergency responses in unforeseen situations, possibility of rating chatbot and also the web site by the user [4].

In this paper, the author introduces a stochastic model for dialogue systems based on the Markov decision process (MDP).

In this paper, the author proposes a formal quantitative model for man-machine dialogue systems. First, the author introduces a general formalization of such systems in terms of their state space, action set and strategy. Within this Markov decision process (MDP), the framework shows that the matter of dialogue strategy designs are often stated as an optimization problem, and solved by a spread of methods, including the reinforcement learning approach. The author working some preliminary results on learning a dialogue strategy for an Air Travel Information System. [5].

In this paper, the author has claimed that for traditional information retrieval tasks such as document retrieval, speech recognition errors generally are either inconsequential or can be dealt with using simple techniques. The author has softened that somewhat with the acknowledgment of those recognition errors. The author working on speech recognition based on user input. When there are only a few words available, there is no opportunity for repetition and context to compensate for errors [6].

Proposed System

In the first phase, the data is collected in the form of speech and stored as an input for subsequent phase for processing. In the second phase, the input voice is continuously processed and converted to text. In the next phase, the converted text is analyses and processed using Python Script to identify the response to be taken against the command. Finally, once the response is identified, the output is generated based on the database generation.

The architecture depicts the system. The basic flow of the system is in the figure 1. The user gives text or voice input. Voice commands are converted to text through speech API and converted to text. Text input is just stored within the database for further process. It is recognized and matched with the commands available in the database. Once the command is found, its respective task is executed as text through the user interface as output.

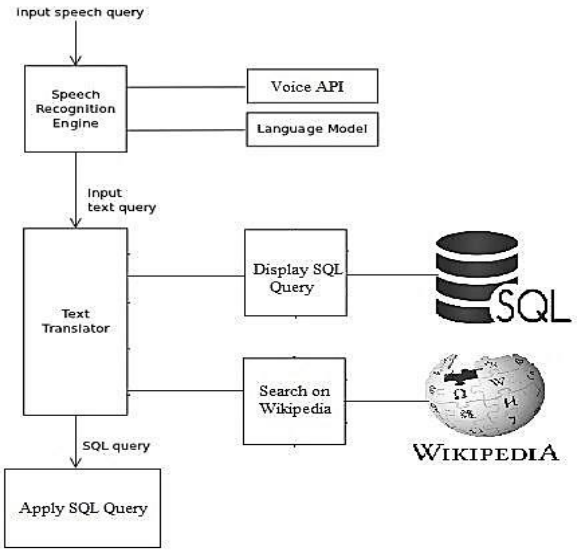

Fig 1. System architecture

In the first phase, the info is collected in the form of speech and stored as an input for subsequent phase for processing. In the second phase, the input voice is continuously processed and converted to text. In the next phase, the converted text is analyses and processed using Python Script to identify the response to be taken against the command. Finally, once the response is identified, the output is generated based on the database generation. The above architecture depicts the system. The basic flow of the system is as above within the figure. The user gives text or voice input. Voice commands are converted to text through speech API and converted to text. Text input is just stored within the database for further process. It is recognized and matched with the commands available in the database. Once the command is found, its respective task is executed as text through the user interface as output.

A. *System feature:*

System Feature 1
The system will help non-technical people in learning SQL queries using voice and text.

System Feature 2
User can fire their queries for the demo database using voice or text.

System Feature 3
User can also search their query in Wikipedia.

B. *System Algorithm*

Following will be our algorithm step

1. Scanning the database:
Here we will go through the database to get the table names, column names, primary and foreign keys.

2. Input:
We will take a sentence as an input from the user (using voice) then convert it into text.

3. Tokenize and Tag:
We will tokenize the sentence and using POS tagging to tag
the words

4. Syntactic Parsing:
Here we will attempt to map the table name and column name with the given natural query. Also, we'll attempt to identify different attributes of the query.

5. Filtering Redundancy:
Here we will try to eliminate redundancy like if while mapping we have created a joining requirement and if they are not necessary then we remove the extra table.

6. Query Formation:
Here we will form an entire SQL query based on MySQL syntax.

7. Query Execution:
Here we will execute the query on the database to get results.

Mathematical Model

Input-Output:
U= {I, O, f, S, F}
Where,
I= {I1, I2}
I1= {l1,l2 ..... ln} where n sql query
I2= i.e. sql query voice based

O= {O1,O2,O3, O4, O5}
O1= Voice process
O2= Sql Query Generate
O3 = Apply SQL query
O4 = SQL query detection
O5 =Voice Generation

f= {f1,f2} f1= preprocess
(voice, sql query) f2 = analysis
(sql query)

S: Success:
SQL query successfully apply

F: Failure:
Algorithm not working properly
Voice command failure

NDD: For a particular input the computer will give different output on different execution**,** so our system is non deterministic

Discussion

In this project, the system translating English query into an SQL query using semantic grammar. The system will accept the user's query in natural language as an input. The program will check whether the query is valid. Then we'll generate tokens by performing the division of the question clause. Each token represents a single word in the user's query. The tokens from the query clause are compared with clauses already stored within the dictionary. The dictionary needs to be constantly updated. Then the algorithm scans the tokens and tries to seek out attributes present within the query. Then we discover all the tables within the database which contain the attributes by comparing syntax and semantics. Then we build the ultimate SQL query and execute it on the database and return the result dataset to the user.

Conclusion

The use of Natural Language brings ease for any human being. This system helps the user to easily retrieve data

from the database using simple English language. The user needn't learn complex command language like SQL. We can add more synonyms for column names and table names so that the system can handle more queries. The system also stores the successfully executed queries based on voice generation. This system provides some recommendations so that it is helpful for the user. In the future, we can add some strong recommendation framework in this system so that the user will have to take fewer efforts.

Acknowledgment

I wish to precise my profound because of all who helped us directly or indirectly in making this paper. Finally, I wish to thank all our friends and well-wishers who supported us in completing this paper successfully I am especially grateful to our guide Dr.Rachna Somkunwar for him time to time, very much needed, valuable guidance. Without the full support and cheerful encouragement of my guide, the paper would not have been completed on time.

References

[1] F.Siasar djahantighi, M.Norouzifard, S.H.Davarpanah, M.H.Shenassa, "Using Natu- Ral Language Processing In Order To Create Sql Queries", in Proceedings of the International Conference on Computer and Communication Engineering, Kuala Lumpur, May 2008, pp. 600-604.

[2] Vismaya V, Darvin Reynald J, "Natural Language Processing Using Python ", International Journal of Scientific Engineering Research Volume 8, Issue 5, May-2017 19 ISSN 2229-5518[IEEE]

[3] TulikaNarang, "Natural Language Processing Techniques Applied in Information Retrieval-Analysis and Implementation in Python", International Journal of Innovations Advance- ment in Computer Science IJIACS ISSN 2347 – 8616 Volume 5, Issue 4 April 2016.

[4] Anuradha Mohite, Varunakshi Bhojane, "Natural Language Interface to Database Using Modified Cooccurrence Matrix Technique 2015 International Conference on Pervasive Computing (ICPC)

[5] Levin E., Pieraccini R., Eckert W. "Learning dialogue strategies within the Markov decision process framework", Automatic Speech Recognition and Understanding IEEE Proceedings, pp. 72-79, 1997.[IEEE]

[6] Liddy, E. D. In Encyclopedia of Library and Information Science, 2nd Ed. Marcel Decker, Inc.

[7] J. Allan. Perspectives on information retrieval and speech. In Information Retrieval Techniques for Speech Applications: LNCS 2273, pages 1–10, 2002.