



Voice Based SQL Query Generation with NLP Process Implementation Using Machine Learning

Savita Kandharkar and Rushali Deshmukh

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

May 5, 2022

Voice based SQL Query Generation with NLP Process Implementation using Machine Learning

Savita V Kandharkar¹

¹Pune University, Computer Engineering,
JSPM's, Rajarshi Shahu College of Engineering kandharkarsavita@gmail.com

Rushali Deshmukh²

²Pune University, Computer Engineering,
JSPM's, Rajarshi Shahu College of Engineering radeshmukh_comp@jspmrscoe.edu.in

Abstract: This system has been developed to change secure access of information to a voice-based programme (UI) by facultative voice-based authentication associate degreed integration with an existing tongue process (NLP) system. we tend to address the question of a way to improve the winning the results from question results. ancient predefined question forms don't seem to be ready to satisfy varied ad-hoc queries from users on those databases. Here, we tend to propose Machine learning based mostly technique to get the SQL question supported user voice, a unique information question type interface, that is in a position to dynamically generate question forms.

Keywords: natural language processing, Languages and compilers, improvement, Verification, Voice Recognition, Machine-independent code generation

I. INTRODUCTION

While linguistic communication could also be the best system for individuals to find out and use, it's proven to be the toughest for a laptop to know. The goal of human language technology is to alter communication between individuals and computers while not resorting to acquisition of complicated commands and procedures. In alternative words, human language technology may be a technique, which may create the pc perceive the languages naturally utilized by humans. during this project, we tend to ar translating English question into a SQL question mistreatment linguistics descriptive linguistics. The system can settle for users question in linguistic communication as Associate in Nursing input. The program can check whether or not the question is valid or not. Then we are going to generate tokens by playacting the division of the question clause. every token represents one word within the users question. The tokens from the question clause ar compared with clauses already hold on within the lexicon. The lexicon has to be perpetually updated. Then the rule scans the tokens and tries to search out attributes gift within the question. Then we discover all the tables within the information that

contain the attributes by scrutiny syntax and linguistics. Then we tend to build the ultimate SQL question and execute it on the information and come back the result dataset to the user.

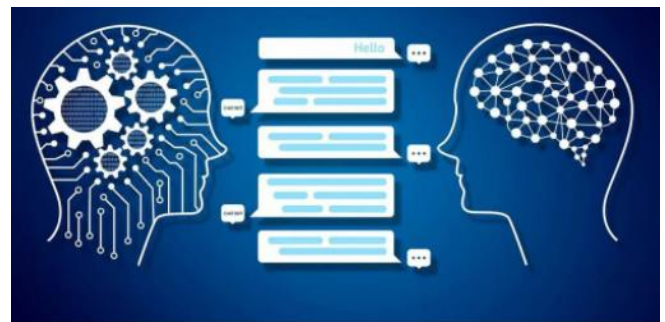


Fig1. Natural Language Processing

II. LITERATURE SURVEY

Natural language methodology area unit exhausted a mix of suggests that communication with device one is communication moreover as verbal communication with device communication is far more easier than the verbal communication. In communication syntax, semantic, lexical and morphological analysis is finished. Whereas in verbal communication embraces all methodologyology| the method} in written further} as a lot of method embody a lot of data relating to linguistics additional} as enough additional information to handle the extra ambiguities that arise in speech[1].

This paper places associate interest in some rising capabilities for progressive speech understanding and methodology in virtual human dialogue systems. This work may be a element of associate current effort that aims to vary wise spoken dialogue with virtual humans in multiparty arbitration eventualities .These eventualities

unit designed to allow trainees to follow their intervention skills by collaborating in face-to-face spoken negotiation with one or more virtual humans. a serious trust achieving realistic behaviour in these arbitration eventualities, that ideally have to be compelled to have the virtual humans representing fluid turn-taking, composite reasoning, and responding to factors like trust and emotions, is for the virtual humans to start out to know [2].

In this custom in virtual human dialogue systems is to use arch human recordings or limited-domain speech synthesis. every approaches cause wise show but at associate elevated value. to figure out the foremost effective trade-off between performance and worth, we've got got Associate in Nursing inclination to perform associate analysis of an individual's and synthesize voices with relation to naturalness, informal side, and likability. varied the kind, length, and content of utterances, and take into thought the age and language of ratters moreover as their expertise with speech synthesis. The results counsel that knowledgeable human voice can surpass every associate amateur human voice and synthesized voices. Also, a high-quality general voice or associate honest limited-domain voice can execute on the so much facet part-time human recordings. for sure, in most cases, the high-quality general voice is rated quite the limited-domain voice. there is likewise a non-statistically necessary trend that has been discovered for long or negative utterances to receive lower ratings [3]. The aim of this paper is to explore business applications of chat bots, moreover on propose several extent metrics to guage follow, usability ANd overall quality of associate embodied informal agent. On the thought of these metrics we've got got Associate in Nursing inclination to seem at existing Polish-speaking business chat bots that, firstly, add the B2C subdivision. Secondly, reach the widest potential vary of users. And lastly, unit possibly the foremost advanced business deployments of their creators. The system analyses varied aspects of functioning of each personified informal agent: optical look, variety of operation on Infobahn electronic computer, speech uniting unit, intrinsic cognition, presentation of knowledge and supplementary functionalities, informal skills and perspective sensitiveness, temperament traits, personalization choices, emergency responses in unforeseen things, risk of rating chatbot and along Infobahn electronic computer by the user [4].

In this paper the author introduce a random model for dialogue systems supported man of science decision technique (MDP). In this paper author propose associate

degree correct quantitative model for man-machine dialogue systems. First, author introduce a general rationalisation of such systems in terms of their state house, action set and strategy. among this man of science decision technique framework show that downside{the matter} of dialogue strategy designs area unit expressed as associate improvement disadvantage, and resolved by a variety of the approach, what is more as a results of the reinforcement learning approach. The author operative some preliminary results on learning a dialogue strategy for associate travel system. [5].

In this paper author have claimed that for ancient information retrieval tasks like document retrieval, speech recognition errors generally unit either inconsequential or area unit addressed victimization simple techniques. Author has softened that somewhat with the acknowledgment those recognition errors. The author functioning on the speech recognition supported user input. once there unit entirely variety of of words getable, there is not any likelihood for repetition and context to make amends for errors [6].

The implemented system is basically a Virtual Assistant that is strictly faculty adjusted. The implemented system entertains the queries of a student regarding faculty connected issues. Authentication mechanism is utilized by the implemented system for student identification. The authentication mechanism includes countersign protection. SQLite is utilized for countersign security. The implemented system was created in golem Studio Platform [7].

The implemented system is basically a Virtual Assistant that is strictly faculty adjusted. The implemented system entertains the queries of a student regarding faculty connected issues. Authentication mechanism is utilized by the implemented system for student identification. The authentication mechanism includes countersign protection. SQLite is utilized for countersign security. The implemented system was created in golem Studio Platform[8].

The Hardware device captures the audio request through electrical device and processes the request therefore the device can reply to the individual victimization in-built speaker module. as associate example, if you raise the device 'what's the weather?' or 'how's traffic?' victimization its intrinsic skills, it's up the weather and traffic standing severally and then returns the response to the patron through connected speaker[9].

This paper presents a replacement approach for sensible search. Overall in world there are many of us United Nations agency use assistant. The paper presents applications of virtual assistant that helps in providing probability for humanity in varied domains. This paper additionally describes provocation of applying virtual Assistant technology [11].

In this paper author propose a way to come up with a whitelist mechanically exploitation queries issued

throughout internet application tests. Author projected methodology uses the queries generated throughout application tests. it's freelance of specific applications, that yields improved timeliness against attacks and flexibility for multiple applications [12].

III. PROPOSED SYSTEM

In 1st section, the information is collected within the type of speech associated keep as an input for consecutive section for process. In second section, the input voice is endlessly processed and reborn to text. In next section the reborn text is analyses and processed mistreatment Python Script to spot the response to be taken against the command. Finally once the response is known, output is generated supported the information generation. The below design depicts the system. the fundamental flow of the system is as on top of within the figure. User provides text or voice input. Voice commands square measure reborn to text through speech API and reborn to text. Text input is just keep within the information for more method. it's recognized and matched with the commands obtainable within the information. Once the command is found, its individual task is dead as text through program as output.

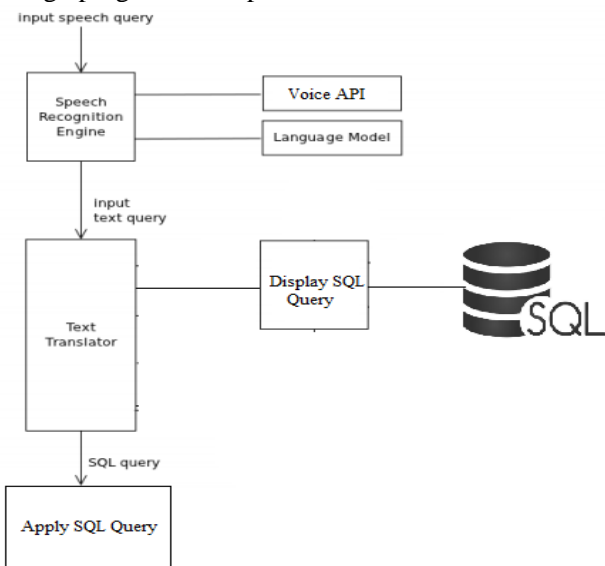


Fig 1. System architecture

A. System feature:

System Feature 1

The system can facilitate non-technical folks in learning sql queries mistreatment voice and text.

System Feature 2

User will fireplace their own queries for demo information mistreatment voice or text.

System Feature 3

User may also search their question in Wikipedia.

IV. MATHEMATICAL MODEL

Input-Output:

$U = \{I, O, f, S, F\}$

Where,

$I = \{I1, I2\}$

$I1 = \{I1, I2, \dots, In\}$ 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

V. RESULT

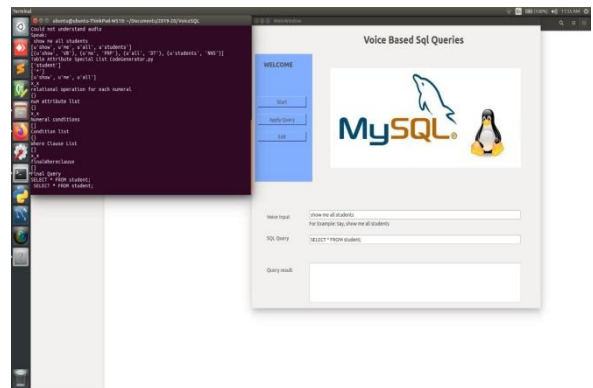


Fig 2. Voice to Query Generate

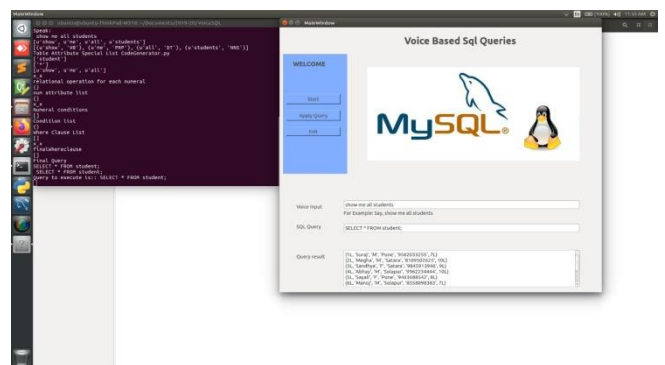


Fig 3. Final Result

VI. CONCLUSION

Use of linguistic communication brings ease for any person. this method helps user to simply retrieve information from info victimization easy English. The user needn't learn complicated search language like SQL. we are able to add a

lot of synonyms for column names and table names in order that system is in a position to handle a lot of queries. The system additionally stores the with success dead queries supported voice generation. this method provides some recommendations in order that it's useful for user. In future we {will|we are able to} add some robust recommendation framework during this system in order that user will have to be compelled to take fewer efforts.

REFERENCES

[1] F.Siasar djahantighi, M.Norouzifard, S.H.Davarpanah, M.H.Shenassa, "USING NATURAL 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] NATURAL LANGUAGE PROCESSING USING PYTHON International Journal of Scientific Engineering Research Volume 8, Issue 5, May-2017 19 ISSN 2229-5518[IEEE]

[3] Natural Language Processing Techniques Applied in Information Retrieval-Analysis and Implementation in Python, TulikaNarang, 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 Co-occurrence 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.

[8] Vishmita Yashwant Shetty, Nikhil Uday Polekar, Sandipan Utpal Das, Prof. Suvarna Pansambal, "ARTIFICIALLY INTELLIGENT COLLEGE ORIENTED VIRTUAL ASSISTANT Department of Computer Engineering, Atharva College of Engineering, Malad (W) Mumbai, India.

[9] Abhay Dekate, Chaitanya Kulkarni, Rohan Killedar, "Study of Voice Controlled Personal Assistant Device" International Journal of Computer Trends and Technology (IJCTT) – Volume 42 Number 1 – December 2016 ISSN: 2231-2803.

[10] Rasika Anerao, Utkarsh Mehta, Akash Suryawanshi, "Personal Assistant for User Task Automation" SSRG International Journal of Computer Science and Engineering (SSRG-IJCSE).

[11] Tushar Gharge, Chintan Chitroda, Nishit Bhagat, Kathapriya Giri, "AI-Smart Assistant" International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 06 Issue: 01.

[12] Komei Nomura, Kenji Rikitake, Ryosuke Matsumoto "Automatic Whitelist Generation for SQL Queries Using Web Application Tests", 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)