

A NATURAL LANGUAGE UNDERSTANDING KNOWLEDGE-BASED CHATBOT OVER LINKED WEB DATA

Ravindra Rao¹, Smt. Patil Rekha²

¹MTech Student, Dept of CSE & PDA College of Engineering, Kalaburagi, Karnataka, India

²Assistant Professor, Dept of CSE & PDA College of Engineering, Kalaburagi, Karnataka, India

Abstract --- With the abrupt development of the web today, the information in web pages is structured and tagged to ensure that it can be read directly by computers, which is available publicly in the type of knowledge bases (KBs). Ensuring this information is machine-readable, easily reachable, and valuable for a customer is one of the goals of the chatbot over linked web data in real-time. Developing a knowledge-based chatbot over linked data leads to many challenges, including user question understanding, multilingual recognition, and supporting multiple knowledge bases. To deal with these challenges, we propose a solution; 1st we design and develop interactive user interface system architecture, 2nd, we plan a machine learning(ML) method for natural language understanding(NLU) to be familiar with user queries via text/voice message methods, then take out the necessary keywords. Information for those keywords is obtained by making queries to the linked website's knowledge base to get desired information using the web scraping technique. For the experiment, we use generic medicine data as a domain, and it's added to the frequently asked queries dataset in the system knowledge base, so the existing system knowledge base gets extended automatically by the chatbot intelligent features, such as addressing the user's queries, feedback messages, and the system failed responses. This system can be widespread with the new domains for an extensive range of topics.

Key Words: Chatbot, Knowledge Bases (KBs), Keyword, Linked Data, Machine Learning (ML), Natural Language Understanding (NLU), Web Scrapping.

I. INTRODUCTION

A Chatbot is a software application that instantly understands what user needs are, gives an idea about them instantly to their preferred outcome, with as little work for the user as possible. Like a virtual assistant for customer experience contact points. It allows users to interact with them using natural language via text or text-to-speech methods. Chatbots are very popular since their beginning in 1960. They have boomed during a previous couple of years like Siri by Apple, Google Assistant by Google, and Cortana by Microsoft. Early Chatbots such as Eliza, Alice, were planned and developed based on text messages. But within the past few years, many big multinational companies have invested in R&D to come up with AI Chatbots to know user questions, then analyze and

respond correctly to those questions by using AI and NLP techniques. Chatbots learn context and frequently getting the better answer for questions in the future, to name a few examples are Chatbot apps that built on E-commerce websites, Chatbot apps on social media, Chatbot apps for customer support and further innumerable examples.

In the circumstances of linked web data, the objective of a chatbot is to get back functional and successively knowledge bases (KBs) from one or more websites via natural language understanding (NLU) and web scraping techniques. Therefore the objective is achieved by modifying natural language queries into keywords -then, the information for a keyword is obtained by making an inquiry query to the URL of a linked website to get back all the relevant information the user needs. Many systems have come up with, but they require lots of data to train the system, which is expensive and inaccessible freely to generate. In recent times, with the expansion of the web, much data is freely available for the development of linked data; rising improvement on Chatbots within the industry has been seen in following a line of investigation. On the other hand, they are still facing plenty of issues, like a better understanding of customer queries, multilingual aspects, and multiple knowledge bases. So during this paperwork, we focused to address some of the above issues.

II. REVIEW OF LITERATURES

In the reference [1], the author proposed a program operating within the MAC time-sharing system at MIT which makes firm kinds of natural language conversation between man and computer possible. Input sentences are examined based on decomposition rules which were triggered by keywords appearing in the input text. Responses were generated by reassembly rules associated with elected decomposition rules. It was one of the early Chatbots motivated by the Turing test projected in 1950[3]. In the reference [2], the author proposed a system that uses AIML format, experiments to create different prototypes of ALICE automatically based on a corpus technique, a description of developed software that converts readable text (corpus) into AIML format is presented alongside with describing the different corpora used. These trials discovered that the possibility of generating useful prototypes without the need for sophisticated natural language processing or ALICE Chatbot. In reference [4], the author proposed a system that uses an inclusive NL engine to understand the user query meanings, and it merges chat text with query. About [5], the author proposed a system for students having a medical background for answering their

query, by using the open-source AIML. The author set out a broadly-offered Unified Medical Language System as a knowledge source domain producing results and converting NL queries into related Structured Query Language queries. These Structured Query Language queries were executed on the knowledge base, and responses were shown to the user in the text messages. In the reference [6], the author proposed a system included with new functions for training English users that can do interaction in English with the English trainee's and produces responses based on the trainee's input, the discussion circumstance, inference, and common sense knowledge. In the reference [7], the author proposed a system, which was constructed using the open-source architecture of AIML to determine, famous person application of Chatbot technology could improve student-content interaction in distance education. Altogether, the findings specify that famous person applications of Chatbot technology may be promising as a teaching and erudition tool in distance and online education. In the reference [9], the author proposed a system related to healthcare that gives responses and instructions related to pregnancy to pregnant patients, mothers, and families with young children, by understanding their requirements/queries. Also, in the reference [10], the author proposed a generic medicine advisor system to prescribe generic medicines for child patients or their parents based on their dialogue with the system, the author used the left-right parsing algorithm to achieve the results. In the reference [11], the author proposed a system for Psychiatric Counseling in Mental Healthcare by using the patient exact and constant emotion identification and collecting clinical psychological responses for giving better satisfactory responses to the patient. Here system was made to order methodologies to understand counseling contents based on high-level natural language understanding (NLU) and emotion recognition based on a multi-modal approach. In the reference [12], the author proposed a system that uses "artificial intelligence" their design is to build a text to text interactive Chatbot that holds patients in conversation about their medical concern and assign essential information and diagnosis based on their symptoms. So that the user will get a plan about their health and have the correct protection. In the reference [13], the author proposed an intelligent and humane Chatbot that provides restaurant allergy accommodation information based on user allergies, to improve the quality of life for young adults with food allergies. It aims to reduce the user's inquiry overload, improve their overall dining-out experiences, and support their social life. In the reference [14], the author proposed a customer service-based Chatbot that supports large-scale and freely obtainable e-commerce data. It takes the benefit of data from product descriptions in addition to user-generated content from e-commerce websites, which is more practical and cost-effective when answering repetitive questions. In the reference [15], the author proposed a system that is based on ontology approach, proposed to replica and function Chatbots. It uses suitable mapping methods to revolutionize ontologies and knowledge into a relational database to drive its chats. In the reference [16], the author proposed a system that aims to simulate friendly conversations using the Egyptian Arabic dialect. Further, reference [8], is an open-domain Chatbot engine designed that integrates the combined results of information retrieval (IR) and Seq2Seq, based generation models.

From the literature survey, the majority of systems are incomplete in terms of design, confidentiality, making use of publically freely available data, and understanding of user

questions naturally. Lots of issues were handled, while few other issues are not handled at all. So this paperwork is to present a chatbot over a linked website knowledge base that takes benefit of extensive, publicly available information.

III. EXISTING SYSTEM

Most of the present systems were implemented based on user conversation that focuses on either a rule or AI- techniques, However, they overlooked to make use of large-scale publically available knowledge bases and the power of linked web data and the semantic web technologies, and other challenges include user queries understanding, multilingual aspect, inadequate in terms of design, confidentiality and natural question understanding.

IV. PROPOSED SYSTEM

The proposed system is built on a modular approach that takes advantage of semantic web, linked data, web scraping, and machine learning techniques. It can perform tasks such as analyzing queries, collecting and responding to user's feedback and analyzing failed responses, comparing with existing FAQs dataset, all get-together information from multiple knowledge bases, showing to the user via text or text-to-speech method. To enhance functionality, users can interact with the system via a text or voice-based message approach.

The proposed system is developed using Python with Django web framework and HTML. Can operate on an independent system to enhance the response time of information reclamation. The FAQ, failed responses and user feedbacks are stored within the database (sqlite3) in an unspecified way for nonstop learning and future enhancement.

OBJECTIVES

- The main objective of this system is to recover helpful and appropriate information from one or many KBs by using NLU and web scraping technologies.
- With the growth of the semantic web, a large scale publically available structured data has become available on the web in the form of knowledge bases (KBs). Making these data available and helpful for users is one of the main objectives of Chatbots over linked web data.
- Design and build a user-friendly chatbot for dialogue management and user interactions;
- Ensuring that the system is scalable and flexible by adding the other knowledge bases, for supporting new languages, and focus on diverse tasks.

V. SYSTEM DESIGN AND METHODOLOGY

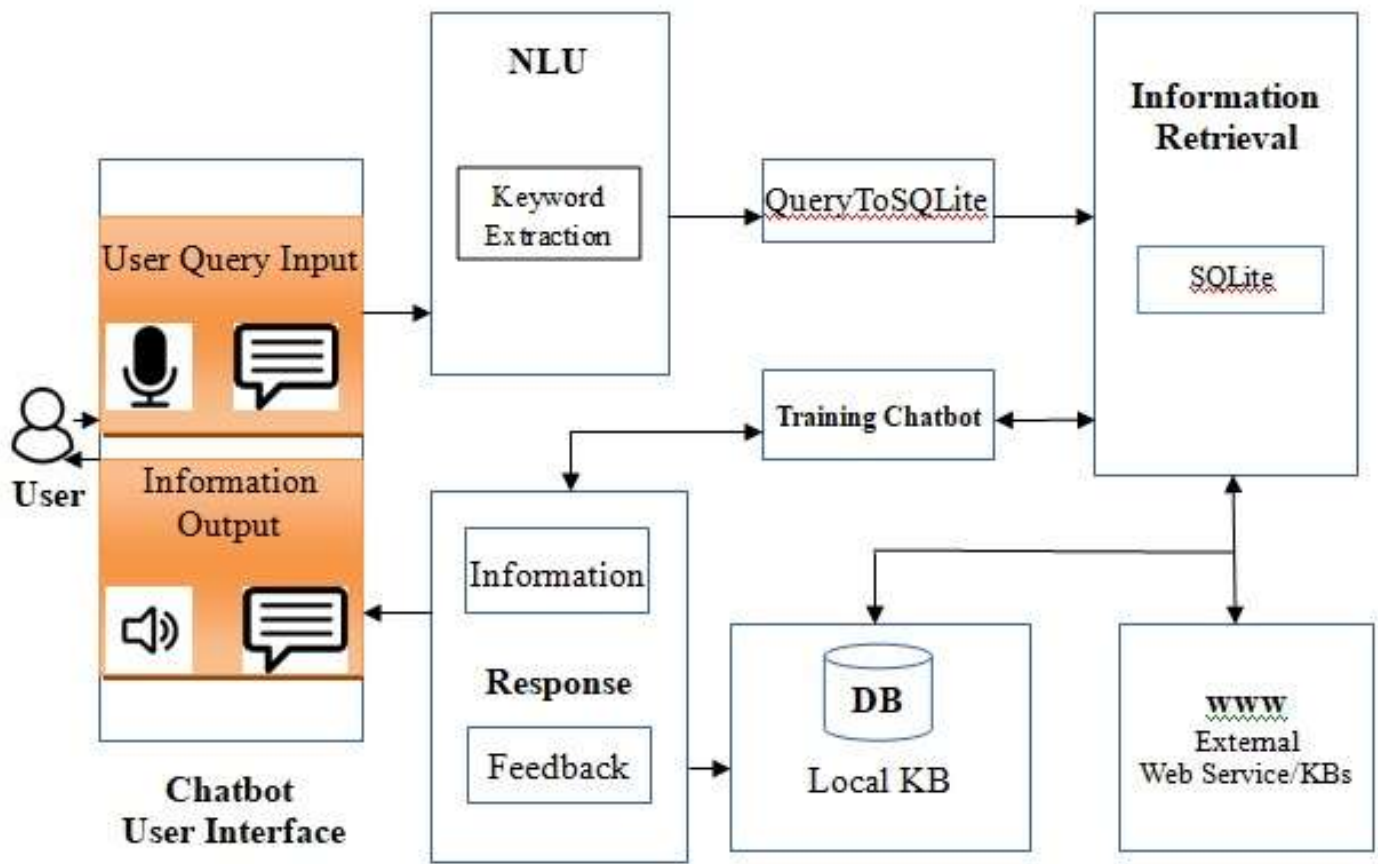


Fig-1: Architecture of Chatbot system.

The Fig-1 shows the architecture diagram of a Chatbot system, which constitutes the following 3 modules explained in detail as follows:

1. Natural Language Understanding
2. Information Retrieval
3. Response

1. Natural Language Understanding (NLU)

NLU is the first module within the system; here the user interaction begins with the system via a text/voice-based message, which can be progressed by this module. Voice is one of the Key techniques of computational linguistics that evolve methodologies, which facilitate the identification of spoken words into text by computers. This method can assist disabled users who couldn't use other devices to communicate with chatbots, and it also improves their typing speed.

To begin during this component, 1st the user interaction language is detected. Then the "Keyword Extraction" subcomponent is employed to require out the important keywords from the queries with the assistance of a selected parser and filters. If the queries belong to local KBs (Database), then the sub-component "QuerytoSQLite" transforms the Query to SQLite query and returns the acceptable response from the FAQ dataset. If the query didn't

belong to the local KBs (Database) then the query is going to be prepared to linked web URL knowledge base to urge the specified information using the web scraping technique, then the "Response" module displays the response to the user via text or text-to-voice message method.

a. Multilingual

Most interactive systems are proposed for English users only because most of the information is obtainable in English on the web. So there is constantly a need to handle multilingual user inputs on interactive systems to serve a broad range of users. We use a language-detect '*googletrans (Google Translation API)*' library to spot the user input language automatically. Later the detected language (Ex: Kannada, Hindi, Telugu, Tamil, Marathi, French, Spanish, German, etc.) are going to be utilized by the "Response" module to display the message to a user in this language; if the specified information isn't found within the detected language then the information to the user is going to be presented in English.

b. Keyword Extraction (KE)

KE is a text examination method with the repeated detection of terms that best-representing user queries. It helps to seek out the essential keywords which user is looking at to get the information. Within the KE process, we use stop words (default for the English language) with the text parser and a

filter technique to filter the foremost frequent words in the user query and then permit the chatbot to select the foremost essential keywords from the user queries.

2. Information Retrieval

IR is 2nd module within the system, here we propose a variety of approaches used to retrieve the response for a given query, SQLite Query generation, and web scraping linked web knowledge bases.

a. SQLite - Query Generation

Here we build a query to SQLite after understanding and processing the query. The Queries to SQLite represent a possibility of user queries within the given FAQ dataset in KBs (Database). The main intention is to create a possible Query that contains information about user input. If the SQLite query didn't match with the FAQ dataset within the system knowledge base then the query is formed to the linked web URL to urge the desired response using the web scraping technique.

b. Web Scrapping

Web scraping is a technique of extracting desired information from a webpage; it parses the webpage and finds the precise location of the tag within the HTML page under which the whole text of the keyword is enclosed.

Below is the **Information Retrieval Algorithm** (based on *Sequential Search and Conditional Logic Algorithm*) used for retrieving the information from multiple knowledge bases in real-time for a given user-entered query.

Algorithm: Information Retrieval Algorithm

Input: *User_Query, Stop_point*

Output: *Information*

Key_words = NULL;

Input = User_Query;

Query = Input_process(Input);

Key_words = Parser(User_Query);

If *Local_KBFAQ contains Key_words* **then**

Information = QuestionToFAQ(Key_words);

Else

Query = QuestionToWebsiteURL(Key_words);

Information = Get_WebSiteResponse(Stop_point; Query);

If *Information* **then**

UpdateLocal_KBFAQ(User_Query, Information);

Else

Information = " Sorry I will get back to this";

UpdateLocal_KBFailedresponse(User_Query);

End

End

Return *Information*

3. RESPONSE

The response is 3rd module within the system; here once the information is retrieved from knowledge bases for the user queries, then it will be shown to the user within the information box as a text/voice message method.

a. Information box

The information box represents an area for displaying the information for keywords acquire by creating queries to numerous knowledge bases in real-time.

b. Feedback

It is a sub-system in response module, where a user can give feedback on the overall performance of the system or anything that must communicate to the supervisor of the system, the feedback submitted by the user are going to be saved to KBs (Database), that may be considered for measuring the quality of chatbot together to improve and progress the performances of the system for future use.

The more detailed working of the system is as shown within the following flowchart in Fig-2 below.

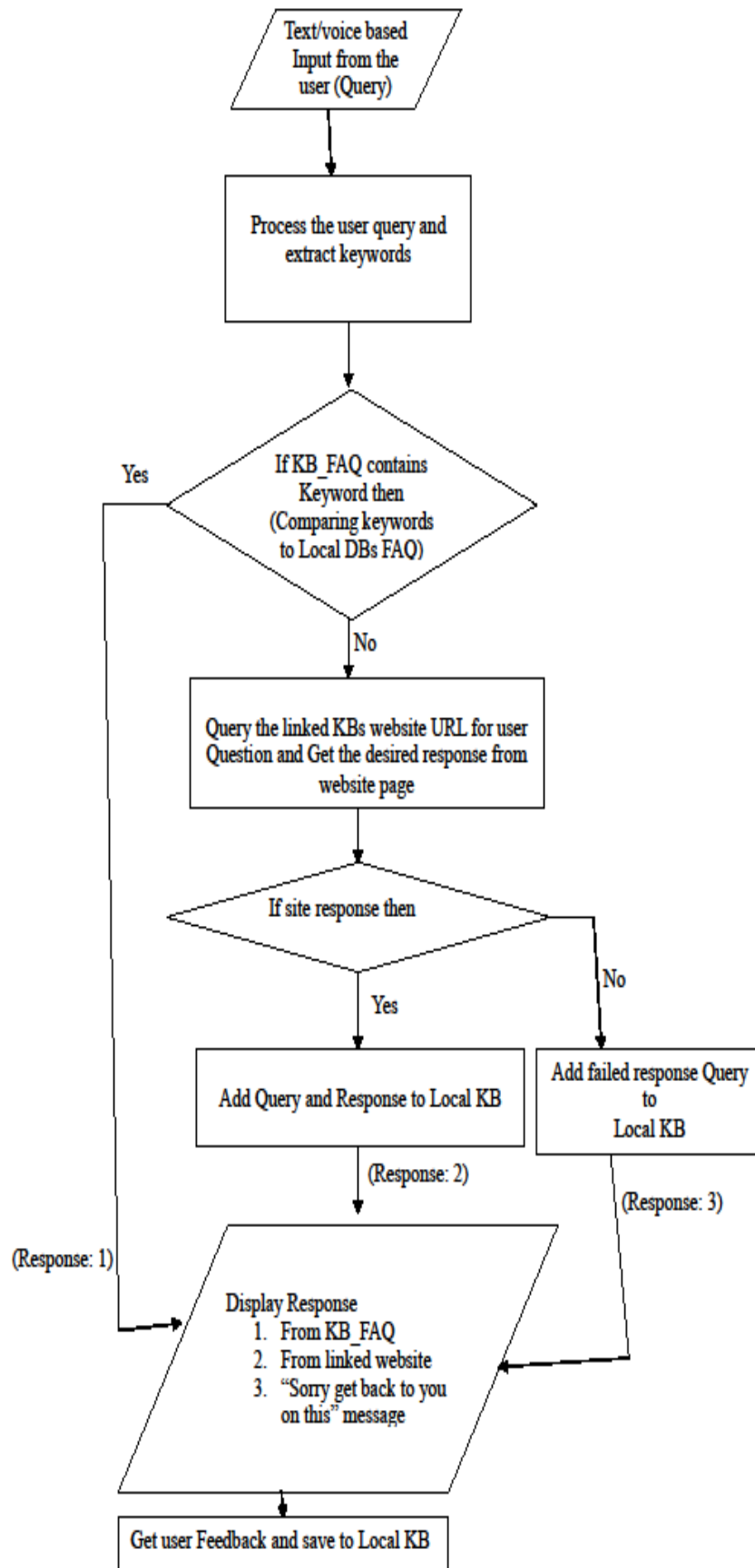


Fig-2: Working mechanism of the Chatbot System

VI. EXPERIMENTAL RESULTS

Here the user interaction starts with the system asking questions via text/voice message method and obtain the response from local knowledge bases and display either via text or text-to-voice message. If there is no response available in the local knowledge base, then the system will look into linked web knowledge bases and get the desired response using the web scraping technique and display the information to the user either via text or text-to-voice message. The same is updated to the local knowledge base by adding questions asked by the user and its response. So that when next time the same question is asked / repeated by any other user, response to the query is returned from the local knowledge base rather than searching in the linked web knowledge base. Hence improving the information retrieval time and increasing the efficiency. If there is no response obtained in the previous two cases, then it will display the message "Sorry I will get back to this" and Log the failed response for user query into the database. When the system administrator login to the system it will be displayed as a list of failed responses, then the admin addresses these failed responses offline. Admin also addresses the feedback messages submitted by registered users, so that the system performance and overall chat quality are improved for future uses. User also gets to visualize their earlier chat history with the system within the information box that is stored in the system database.

This chatbot is tested for the following languages English, Kannada, Hindi, Telugu, Tamil, Marathi, Spanish, French, German, etc. the results show that the chatbot correctly identifies the user natural language, process it, and returns the results in the user language, this chatbot can be extended and tested further to support other new languages.

In summary, this chatbot contributes to a better understanding of user queries in the context of linked data by answering different user queries. This has many following strengths, such as, It provides a secured user-friendly an interactive interface for a wide range of user conversation, it instantly reply to user repeated questions via text/voice message, it supports multilingual, it's easy to extend and flexible to add new knowledge bases and lastly very flexible/customized to support new languages.

VII. CONCLUSIONS

The usage of chatbot is extremely customer friendly and this will be utilized by anyone who knows the way to type within their language in the application. Our proposed chat system responds to supported user queries taking advantage of global, openly freely accessible linked knowledge bases data. Multilingual, voice to text, moreover it influences the technologies of ML and NLU, and repeated questions, with dialogue management. From usability investigation, it views that the proposed chatbot is enhanced throughout the user experience in terms of question answering and performance. It is more suitable for information recovery, Information acquisition, question understanding, and continual knowledge. The future work will be extended to further knowledge bases (extending to new domains, for a wide-ranging of topics) and new languages; the same is incorporated with 3rd party services.

REFERENCES

- [1] J. Weizenbaum, "ELIZA_A Computer Program For The Study Of Natural language Communication Between Man And Machine", Commun. ACM, vol. 9, no. 1, pp. 36_45, Jan. 1966, DOI: 10.1145/365153.365168.
- [2] E. Atwell and B. AbuShawar "ALICE Chatbot: Trials And Outputs," Com-putación y Sistemas, vol. 19, no. 4, Dec. 2015.
- [3] A. M. Turing, "I_Computing Machinery And Intelligence," Mind, vol. LIX, no. 236, pp. 433_460, Oct. 1950, DOI: 10.1093/mind/LIX.236.433.
- [4] R. Chatbot. (2008). [Online]. Available: <http://brilligunderstanding.com/rosedemo.html>
- [5] H. Kazi, and Z. Memon and B. S. Chowdhry "Medchatbot: An UMLS based chatbot for medical students," Int. J. Comput. Appl., vol. 55, no. 17, pp. 1_5, 2012.
- [6] J. Jia, "CSIEC: A computer-assisted English learning chatbot based on textual knowledge and reasoning," Knowl.-Based Syst., vol. 22, no. 4, pp. 249_255, May 2009.
- [7] D. Y.-O.-N.Marchand, B. Heller, M. D. Proctor, B. Cheung, "Freudbot: An investigation of chatbot technology in distance education," in Proc. World Conf. Multimedia, Hypermedia Telecommun., 2005.
- [8] H. Chen, J. Huang, and W. Chu, M. Qiu, F.-L. Li, S. Wang, X. Gao, Y. Chen, W. Zhao, "AliMe chat: A sequence to sequence and rerank based chatbot engine," in Proc. 55th Annu. Meeting Assoc. Comput. Lin-guistics (Short Papers), vol. 2, 2017, pp. 498_503. [Online]. Available: <https://www.aclweb.org/anthology/P17-2079>
- [9] A. Melpignano, L. Vaira, M. A. Bochicchio, and M. Conte, F. M. Casaluci, "MamaBot: A system based on ML and NLP for supporting women and families during pregnancy," in Proc. 22nd Int. Database Eng. Appl. Symp.IDEAS, 2018, pp. 273_277.
- [10] B. M. B. Francisco B. E. V. Comendador, , J. S. Medenilla, ,and T. B. E. Serac, "Pharmabot: A pediatric generic medicine consultantchatbot," J. Autom. Control Eng., vol. 3, no. 2, pp. 137_140, 2015.
- [11] K. Oh, H.-J. Choi, and D. Lee, B.Ko "A chatbot for psychiatric counseling mental healthcare service based on emotional dialogue analysis and sentence generation," in Proc. 18th IEEE Int. Conf. Mobile Data Manage. (MDM), May 2017, pp. 371_375.
- [12] I. Divya, K. Devi, and P. Ishwarya "A self-diagnosis medical chatbot using arti_cial intelligence," J.WebDevelop.Web Designing, vol. 3, no. 1, pp. 1_7, 2018.
- [13] T. Hsu, T. Liu, C.Wang, J. Zhao and K. Liao "AllergyBot: A chatbot technology intervention for young adults with food allergies dining out," in Proc. CHI Conf. Extended Abstr. Hum. Factors Comput. Syst. CHI EA, 2017, pp. 74_79, DOI
- [14] L. Cui, C. Tan, C. Duan, S. Huang, F. Wei, and M. Zhou, "SuperAgent: A customer service chatbot for E-commerce websites," in Proc. ACL, Syst.Demonstrations, 2017, pp. 97_102.
- [15] A. A. Issa and H. Al-Zubaida, "OntBot: Ontology-based chatbot," in Proc.Int. Symp. Innov. Inf. Commun. Technol., Nov. 2011, pp. 7_12.
- [16] N. Habash and D. A. Ali, "Botta: An Arabic dialect chatbot," in Proc.COLING, 2016, pp. 208_212.