

An Interactive Chatbot Designed for Books

J. Nivetha¹, K. S. Kirubha Hari², B. Keerthi³, Caroline⁴

^{1,2,3}Student, Department of Computer Science and Engineering, SRMIST, Chennai, India

⁴Assistant Professor, Department of Computer Science and Engineering, SRMIST, Chennai, India

Abstract—A chatbot or chatterbot, acts as an conversational interface between an individual and the computer system. They are computer programs capable of carrying out near - natural conversation with people. With recent advances in artificial intelligence and machine learning, chatbots are becoming popular in most domains .In this work, we propose a chatbot as a mobile based application which answers queries from a book or multiple books, fed into the system that may contain few to many pages. A database is used for storing knowledge. Pattern matching is used to find the answers for the input queries. Bigram is used for effective query translation so that the response is identified optimally. This is used for simple statistical analysis of text in our application. The chatbot interface is available to the user as an android application created using Android Studio. The proposed system will prove to be efficient in accessing knowledge without wasting any time.

Index Terms— artificial intelligence, machine learning, database, pattern matching, bigram, Android Studio

I. INTRODUCTION

Artificial Intelligence is the most volatile domain that is attracting a lot of researchers and programmers in recent times. The advancing world, depends on automation of daily routines, which require artificial intelligence for optimal work. Researchers claim that advancements in this domain, could one day conquer the world.

Books have been proved to be the great source of vast knowledge. They contain various information ranging from definition to diagrams. So, when a particular definition or an answer is required, the amount of time spent on searching is wasted. The technologically developed world has made mankind fulfil one's need in much more easier way to get the desired things.

Among various innovations in this domain, chatbots also known as chatterbots, are the most prominent in machine learning. So, the paper proposes a mobile application that powers a chatbot which will respond efficiently to the queries related to a selected book. The chatbot will be very optimal and pave an easy way to access the knowledge of the book in a much lesser time.

The chatbot is designed accept simple queries and it would aim to give the specific answers required pertaining to the given query. For an example, if the user requires an particular definition from the book, a simple message carrying the query will be sent , for which the chatbot will respond with the definition, page number and further explanation related to the

definition . The query will be translated into a simpler form using pattern matching algorithm, efficiently the bigram algorithm and the perfect response will be decided using basic decision tree algorithm.

II. EXISTING SYSTEMS AND RELATED WORKS

Chatbots have been existing since the emergence of artificial intelligence. They have proved to be efficient in various fields such as management, complaints registering and response. Most of the chatbots rely on knowledge that is provided to the bot and do not learn from frequent replies and conversations.

[6] proposes a chatbot which is based on static database. The chatbot is designed to respond to queries in Indonesian language. It is designed effectively to overcome challenges like misspellings, picking up of keywords and result identification. It also implements sentence similarity measurement for effective response.

[5] classifies all the query in three categories for the effective query management. It groups queries under atomic category, i.e the query will have only one response. Other classification is default category where the query will have multiple response and final classification is recursive category where both query and response are dynamic. This effective classification of query makes the working of chatbot simpler and consumes less time.

[3] deals with intelligent and unintelligent chatbots. It aims to develop an intelligent chatbot with features like real time chat system, recommendation of answers,solves queries, cross platform performance and keywords based prediction.

[1] aims on developing a intelligent question answering system, which uses WordNet and English Corpus to identify synonyms of certain words which are not present in directive. It also deals with multi staged queries and aims on proving multi stages responses.

[2] provides challenges in developing a chatbot. Natural language processing and complex queries are prominent challenges faced by chatbot developers. The proposed system uses WordNet or English Corpus to overcome this particular challenge.

[4] provides basic information based on the decision making trees. Decision making trees have proven to be optimal in case of query response. The proposed system implements a classification tree for effective performance and K means clustering tree is used to respond to less accurate query response.

III. METHODOLOGY

Pattern matching algorithm is the base of the chatbot. When a query is accepted, it should be analysed for keywords. The identified keyword must be matched with database for the efficient answer. Identification of keyword is done using Bigrams. The pattern matching is carried out by sentence similarity measurement.

A. Bigrams

The probability of measuring a sentence is given mathematically by the function (1)

$$\rho (|\Psi|) \tag{1}$$

So, if the query is “What is food chain” , $(\rho(\Psi_1)=\text{“what”}) * (\rho(\Psi_2)=\text{“is”}) * (\rho(\Psi_3)=\text{“food”}) * (\rho(\Psi_4)=\text{“chain”})$ will be the elaborated generalization of bigram[6].

Further the extended function generates a set of two lettered bigrams like

{“wh”, “ha”, “at”, “tf”, “fo”, “oo”, “od”, “ch”, “ha”, “ai”, “in”},

which makes the pattern matching more dynamic and proficient. The use of N-grams is also promoted. The disadvantage of using N-grams is that it consumes much time.

B. Sentence Similarity Measurement

Sentence similarity is giving score for semantic relation between two sentences or strings. So, if there are two sentences or strings, sentence similarity can be determined by the similarity of two sentences or strings. The higher score of the sentence semantic similarity, the more similar the meaning of two sentences. The score of the sentence semantic similarity is from 0 until 1. The equation of sentence similarity represented by eq. (2).

$$\zeta(s_1 \in s_2) \cup \zeta(s_2 \in s_1) / \zeta(s_1) \cup \zeta(s_2) \tag{2}$$

where s_1 = sentence or keyword identified in the query, s_2 = piece of sentence identified in the database. The ζ stands for the cardinality of the bigram [6] set produced by bigram algorithm after identification of keyword.

So for an example let us consider two words “deflection” and “defection” and perform the given algorithms.

So the inputs are made into sets, name $S_1=\{d,e,f,e,c,t,i,o,n\}$ and $S_2=\{d,e,f,l,e,c,t,i,o,n\}$. By applying bigrams

$S_1=\{\text{“de”}, \text{“ef”}, \text{“fe”}, \text{“ec”}, \text{“ct”}, \text{“ti”}, \text{“io”}, \text{“on”}\}$, where the cardinality of the set $\zeta(S_1)=8$.

Similarly, $S_2=\{\text{“de”}, \text{“ef”}, \text{“fl”}, \text{“le”}, \text{“ec”}, \text{“ct”}, \text{“ti”}, \text{“io”}, \text{“on”}\}$, and the cardinality of the set $\zeta(S_2)=9$.

Applying the formula
 $S_1 \in S_2 = 6$ and $S_2 \in S_1 = 6$

Applying (8), we get,
 $6 \cup 6 / 8 \cup 9 = 12 / 17 = 0.70$.

So the sentence similarity score for the words “deflection” and “defection” is 0.7 which can be accepted by the proposed algorithm.

IV. SYSTEM ARCHITECTURE

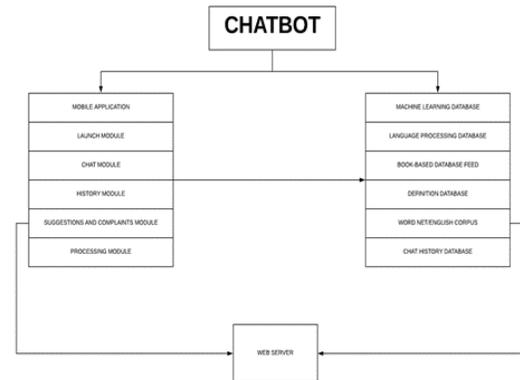


Fig. 1. System architecture

The proposed system consists of two important fragments, which are the mobile application and the database. They are embedded with each other by the use of SQLite Database features offered by application development platforms.

V. MODULE DESCRIPTION

A. The Chatbot Module

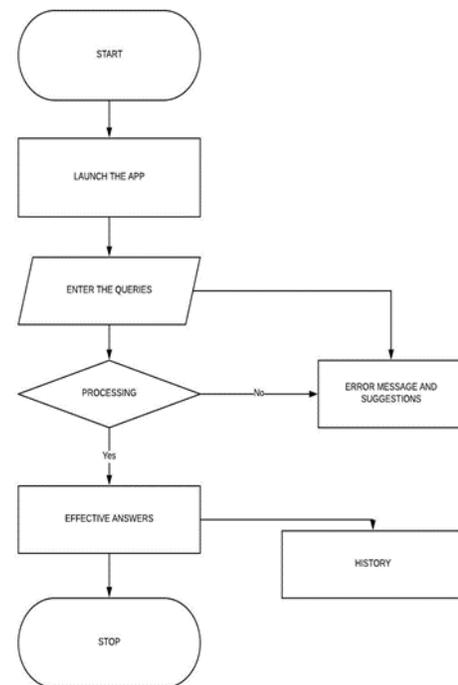


Fig. 2. Chatbot module

The chatbot module is the first prominent module of the proposed system. The module is powered by a mobile application which is designed by Android Studio (any latest version). This module consists of the other five main modules namely the launch module, the processing module, the chat

module, the history module and the complaints and suggestions module.

The launch module is supported by MainActivity.java that is developed in the application. Launching of chatbot is associated with it. The processing module is responsible for all functions of the chatbot. Query processing using bigrams algorithm, pattern matching, and query responding are the most significant steps carried out by the processing module.

The Activity.xml file powers the chat module. This particular module controls the presentation of chatbot on screen. It has another counterpart-the history module. This modules saves previous responses and sends it back to the database for effective query response. This module is also connected to web, so that all responses are inter connected which again improves the quality of query response.

The final module is the complaints and suggestion module, which is directly connected to the web. The web has a separate database to store history of every response and suggestions, which will be updated to the application in stipulated intervals. The complaints will be rectified at the server end periodically which improves the efficiency at a greater rate.

B. The Database Module

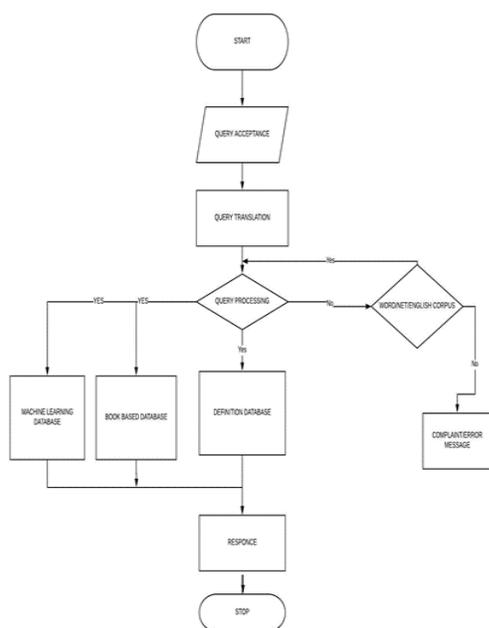


Fig. 3. Database module

The database module is the heart of the chatbot. It contains various other modules which are the core responsibility of working of the chatbot. This module consists of the book-based database feed, the definition database, and the language processing database. The English Corpus/Wordnet database and chat history database are directly connected to the web.

The book based database is fed using MySQL, in such a way the accessing of the database for query responding is made available. It is accessed by the chatbot, after performing pattern

matching using bigrams algorithm. The language processing database is a naïve requirement to deal with natural language processing. It categorizes queries as atomic, default and recursive and reponds from the database. The reponse is improvise using this database. The definition database consists of basic definitions for queries with “what” keyword. The definition database also contains similar words and possible queries.

The English Corpus and WordNet are used on wider basis for accessing synonyms of certain words to match with existing database. In care cases, decision tree algorithms are used for effective keyword identification. The chat history is saved in the database for future response and is periodically updated to the web.

VI. CONCLUSION AND FUTURE SCOPE

Hence, need for information from huge books is made easier using this proposed system. Incorporating the proposed methodology, we can make an effective chabot.

With the tremendous growth in the following domain and the fast paced life, the proposed system will prove to be proficient and less time consuming. The chatbot serves the specific purpose by being specially designed for books. In the future they can be designed for specific and important documents.

The chatbots can be trained to understand big data and retrieve information. They can also be used as assists in the domain of big data analysis. They can also be trained to study reports and derive conclusions. This aspect can be applied to various fields on a higher scope with far sighted advantages.

The chatbot can also be developed with voice over controls. The query can be retrieved form a voice note and specific answers can also be read out by the system. This can help visually challenged groups needing various information from huge books.

As the proposed system supports the language of English only, in the future, it can be designed for various languages across the world. Assisting them with voice over controls can prove to be more helpful and advantageous. Similarly, various algorithms can be developed for natural language processing and effective pattern matching.

Overall the proposed system can be utilized and will pave proficient way of learning. It will in growth of knowledge, without spending tremendous time on information collection.

REFERENCES

- [1] D. Aimless, S. Umatani, "A tutorial for adding knowledge to your robot", Retrieved from, 2004.
- [2] Aimi and Isa", Advances in Computing Communications and Informatics (ICACCI) 2016 International Conference on. IEEE, pp. 2740-2742, 2016.
- [3] H. Al-Zubaide, A. A. Issa, "OntBot: Ontology Based Chatbot", Proc. IEEE of 2011 Fourth International Symposium on Innovation in Information & Communication Technology (ISIICT), pp. 7-12, 2011.
- [4] C. Erdogan, H. Nusret Bulus, B. Diri, "Analyzing The Performance Differences Between Pattern Matching and Compressed Pattern Matching on Texts", Proc. IEEE of 2013 International Conference on Electronics Computer and Computation (ICECCO), pp. 135-138, 2013.

- [5] S. Ghose, J. J. Barua, "Toward The Implementation of A Topic Specific Dialogue Based Natural Language Chatbot As An Undergraduate Advisor", Proc. IEEE of 2013 International Conference on Informatics Electronics & Vision (ICIEV), pp. 1-5, 2013.
- [6] G. Neubig, NLP Programming Tutorial 2 - Bigram Language Models, Nara Institute of Science and Technology (NAIST).
- [7] Y. Bin, P. Cunlin, L. Dan, "Chinese Text Feature Extraction Method Based on Bigram", Proc. IEEE of 2013 International Communications Circuits and Systems (ICCCAS), pp. 342-346, 2013.
- [8] Bayu Setiaji, Ferry Wahyu Wibowo, "Chatbot Using A Knowledge in Database", 7th International Conference on Intelligent Systems Modelling and Simulation, 2016.
- [9] Bonnie Chantarotwong, "The Learning Chatbot", Fall, 2006.