

## International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 7, Issue 4, April 2018

# A Web Based College Enquiry Chatbot with Results

Sagar Pawar, Omkar Rane, Ojas Wankhade, Pradnya Mehta

Department of Computer Engineering, MMCOE, Pune, India

**ABSTRACT:** A chatbot (also known as a talkbot, chatterbot, Bot, IM bot, interactive agent, or Artificial Conversational Entity) is a computer program which conducts a conversation via auditory or textual methods. The chatbot has information stored in its database to identify the sentences and making a decision itself as response to answer a given question. The college enquiry chat bot will be built using algorithms that analyzes queries and understand user's message. This System will be a web application which provides answer to the query of the student very effectively. Students just have to put their query to the chat-bot which is used for chatting. The system will use bigram and sentence similarity algorithms to give appropriate answers to the user. If the answer is found invalid, then there is a system to declare the answer as invalid. These invalid answers can be deleted or modified by the administrator of the system.

### General Terms

Natural Language Processing, Sentence Similarity.

**KEYWORDS:** Bigram, Information Retrieval and Search.

## I. INTRODUCTION

We sometimes pass our time by chatting with different chatterboxes available on internet, which is often aimed for such purposes or just entertainment [5]. The chatbots have embedded knowledge which helps them to identify the user's query and give a response to it [8]. The college enquiry chat-bot project is built using algorithms that analyses user's queries and understand user's message. This system is a web application which provides answer to the query of the student. Students just have to query through the bot which is used for chatting. The user can query any college related activities through the system. The user doesn't have to personally go to the college for enquiry. The system analyses the question and then answers to the user. The system answers to the query as if it is answered by the person. With the help of algorithms, the system answers the query asked by the students. The system will also have an online notice board on which any Text notices or PDF documents can be displayed. This will help the user to be updated with the important notices. Not much time will be wasted by the user to search for the important notices.

## II. LITERATURE SURVEY

In this paper bigram is used for calculating the sentence similarity. The machine has been embedded with the knowledge to identify the sentence and making a decision itself as response to answer a question [1].

In this study of paper the statistics of pattern matching on the text data and statistics of compressed pattern matching on compressed form of same text data are compared [2].

Here the information repository is in the form of a connected graph where the nodes contain information and links interrelates the information nodes. The design semantics includes AIML (Artificial Intelligence Markup Language) specification language for authoring the information repository such that chat bot design separates the Information repository from natural language interface component [3].

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 7, Issue 4, April 2018

The paper describes a novel method is proposed where bigram is applied to quantify the text and improved information gain algorithm are used to create appropriate feature during text categorization [4].

The paper illustrates the implementation and semantic enhancement of domain oriented question answer system based on pattern matching chatbot technology developed within industrial project FRASI [5].

In this paper a modular knowledge base is equipped with the conventional architecture. It helps in building a specific module that deals with a particular feature of the conversation.

This enhances agent's interaction capabilities [6].

In this paper conversation of agent based on modular knowledge representation is proposed. It has a dynamic and flexible behavior. Because of the modularity of the architecture it allows a concurrent and synergic use of different techniques making it possible to adapt to the specific characteristics of the domain. It has set of modules which is automatically triggered through a component [7].

OntBot uses appropriate mapping techniques to transform ontologies and knowledge into relational database and then use that knowledge to drive its chat. Uses rule matching to match the sentence [8].

Here the growing problem of malicious chatbots are taken into consideration and provide a supporting evidence to distinguish between human and chatbots [9].

Traditional chatbot knowledge base are hard constructed and time consuming. Automatic chatbot knowledge acquisition from online is used here. It uses rough set and ensemble learning for that purpose [10].

### III. PROPOSED SYSTEM

The goal of the system is to help the students to stay updated with their college activities. The main motive of the project is to reduce the work load on the college's office staff and reduce the response time to a user's query. The student when visits the website first registers him/herself and after which he can visit the enquiries section for making queries to the chatbot. The chatbot has information in the form of pattern-template stored in the database [2]. Use of SQL is made for handling the database. The input query undergoes tokenization, bigram and sentence similarity score phases discussed in the following section.

The system architecture is as shown below:

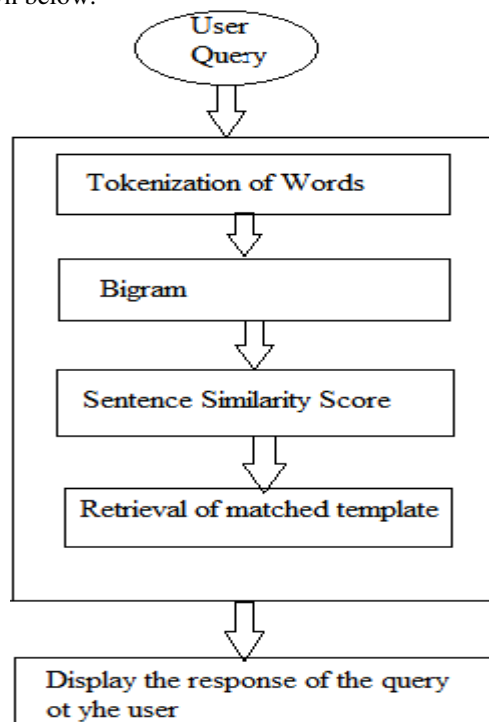


Fig. 3.1 System Architecture of Chabot.

## International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 7, Issue 4, April 2018

The system has 3 modules. One in which authorized user inputs his query and on that query tokenization, bigram, sentence similarity scores are applied and retrieving or updating of the appropriate template from database is done in the second module. The third module displays the response to the user.

### a. Tokenization

In this step, a complete sentence is divided into words, that is, the sentence is broken down into separate words.

Example: If user inputs a query say, "I read a book."

The output of tokenization step will be: { I, read, a, book. }

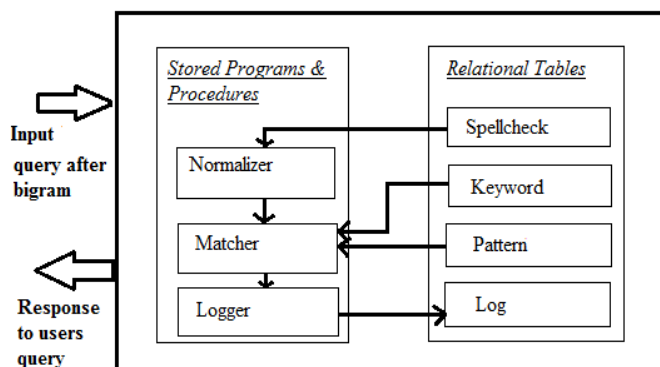


Fig. 3.2 Core of Chabot.

The output of bigram algorithm goes through normalization. Also spelling checking is done along with normalization. After normalization, the keywords from the output step of bigram are matched with the patterns stored in the database. Appropriate template is returned after matching and log of it is stored in the database.

Sr.No	Users Input Query after Bigram	Output of Normalizer with spellcheck
1	I read	I read
2	read a	read a
3	a book.	a book

So now for the pattern in the database, the matcher matches the keywords of that pattern along with that of the output of the normalizer, and using sentence similarity as described later, retrieves the template having the highest sentence-similarity score, as shown below :

Sr. No .	Input Sentence	Pattern in the database	Sentence Similarity Score	Output Of Matcher
1	I read a book	I read a good story book	0.5	Appropriate template of "I read a good story book".
		I read horror books.	0.33	

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 7, Issue 4, April 2018

## b. Bigram

Bigram means sequence of two adjacent words. The probability of calculating sentence could be represented mathematically as [11]:

$$P(W) = \prod_{i=1}^{W_i+1} P(w_i | w_0 \dots w_{i-1}) \quad \dots(1) \quad \text{where } W = \text{sentence given,}$$

$P(W)$  = Probability of sentence.,  $w_i$  = a word in the sentence at position  $i$ .

Bigram model adds one word of context

$$P(w_i | w_0 \dots w_{i-1}) \approx P(w_i | w_{i-1}) \quad \dots(2)$$

So, the (2) equation can be enacted a linear interpolation using Witten-Bell smoothing algorithm. The Witten-Bell smoothing algorithm is applied to predict the probability of bigram model with zero count or  $P(w_{i-1}|w_i) = 0$  [12]. From Good-Turing estimation, the total mass of counts with a zero count in distribution is the number of things with one count.

We can use linear interpolation

$$P(w_i | w_{i-1}) = \lambda_2 P_{ML}(w_i | w_{i-1}) + (1 - \lambda_2) P(w_i) \quad \dots(3)$$

Where  $P_{ML}$  is the probability of Maximum Likelihood.

Make the interpolation depend on the context:

$$P(w_i | w_{i-1}) = \lambda_{w_{i-1}} P_{ML}(w_i | w_{i-1}) + (1 - \lambda_{w_{i-1}}) P(w_i) \quad \dots(4)$$

Where,  $\lambda_{w_{i-1}}$  can be calculated as:

$$\lambda_{w_{i-1}} = 1 - \frac{u(w_{i-1})}{u(w_{i-1}) + c(w_{i-1})}$$

$u(w_{i-1})$  = number of unique words after  $w_{i-1}$  .....(5)

## c. Sentence Similarity Score

Sentence Similarity Score is calculated by first finding the intersection between two sentences one which is input query and other the pattern stored in database. The intersection is denoted by:

$$S_1 \cap S_2 \ \& \ S_2 \cap S_1, \text{ where } S_1 \text{ and } S_2 \text{ are sentences.}$$

The formula for calculating the sentence similarity score is:

$$\frac{\text{Count}(S_1 \cap S_2) \cup \text{Count}(S_1 \cap S_2)}{\text{Count}(S_1) \cup \text{Count}(S_2)}, \text{ where } S_1 \ \& \ S_2 \text{ are sentences.}$$

$$\text{Count}(S_1) \cup \text{Count}(S_2)$$

Consider the example of the user input statement:

$$S_1 = \text{"I read a book."} \quad \text{and the pattern in the database as:} \quad S_2 = \text{"I read a good story book"}$$

Using the bigram algorithm we get:

$$S_1 = \{I \text{ read, read a, a book}\} \quad \text{Count}(S_1) = 3$$

$$S_2 = \{I \text{ read, read a, a good, good story, story book}\} \quad \text{Count}(S_2) = 5$$

Now,  $S_1 \cap S_2 = 2$  and  $S_2 \cap S_1 = 2$

Sentence Similarity Score is calculated as:

$$\frac{\text{Count}(S_1 \cap S_2) \cup \text{Count}(S_1 \cap S_2)}{\text{Count}(S_1) \cup \text{Count}(S_2)} = \frac{2 \cup 2}{3 \cup 5} = \frac{4}{8} = 0.5$$

Similarly we can calculate the sentence similarity score for each pattern stored in the database.

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 7, Issue 4, April 2018

## IV. APPLICATIONS

- The chat-bot can be even embedded on the website of tuition or private coaching classes to answer the queries of students.
- Even this chat-bot can be embedded on the website of a hospital for its users to make inquiries about the patients and doctors available in the hospital. Also which specialties are provided can be asked.

## V. RESULTS

The output of the implemented system is shown below.



Fig. 5.1 User Inputs Question

As shown in the above Fig. 5.1, the user has to select the category of the question to which the question is related. Selection of the category is compulsory. After entering the question the user clicks on the Ask button.

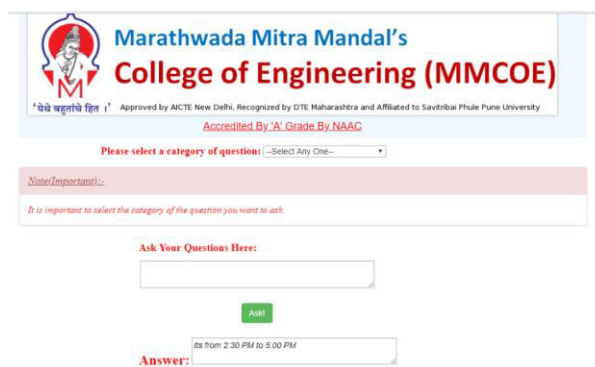


Fig. 5.2 Output to the Users Query.

As shown in the above Fig. 5.2, after clicking on the Ask button the reply to the query is given in the Answer box.

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 7, Issue 4, April 2018



The image shows two screenshots of a web interface for Marathwada Mitra Mandal's College of Engineering (MMCOE). The top screenshot shows a question: "Does the college have hostel facility?" with a green "Ask" button and an empty "Answer:" field. The bottom screenshot shows the same interface with the question field empty and the answer field containing the text: "Yes, there is a separate hostel for Boys and Girls".

## VI. CONCLUSION

The proposed system will be used to identify answers related to user submitted questions. The need is to develop a database where all the related data will be stored and to develop a web interface. A database will be developed, which will store information about questions, answers, keywords, logs and feedback messages. A usable system will be designed, developed and deployed.

## VII. FUTURE SCOPE

In the future scope of this project, we can include voice based queries and answers. The users will have to give voice input and the system will give the text output and while giving it, it will give a voice output as well. By just adding text-to-speech and speech-to-text we can add the above functionality to our project.

# International Journal of Innovative Research in Science, Engineering and Technology

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

Visit: [www.ijirset.com](http://www.ijirset.com)

**Vol. 7, Issue 4, April 2018**

## REFERENCES

- [1] Bayu Setiaji, Ferry Wahyu Wibowo, "Chatbot Using A Knowledge in Database", 2016 7th International Conference on Intelligent Systems, Modelling and Simulation.
- [2] C. Erdogan, H. Nusret Bulus, and 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), 2013, pp. 135-138, doi:10.1109/ICECCO.2013.6718247.
- [3] S. Ghose and 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), 2013, pp. 1-5, doi:10.1109/ICIEV.2013.6572650.
- [4] Y. Bin, P. Cunlin, and L. Dan, "Chinese Text Feature Extraction Method Based on Bigram," Proc. IEEE of 2013 International Communications, Circuits and Systems (ICCCAS), 2013, pp. 342- 346, doi: 10.1109/ICCCAS.2013.6765352.
- [5] A. Augello, G. Pilato, A. Machi, and S. Gaglio, "An Approach to Enhance Chatbot Semantic Power and Maintainability: Experiences Within The FRASI Project," Proc. of 2012 IEEE Sixth International Conference on Semantic Computing, 2012, pp. 186-193, doi:10.1109/ICSC.2012.26.
- [6] A. Augello, M. Scriminaci, S. Gaglio, and G. Pilato, "A Modular Framework for Versatile Conversational Agent Building," Proc. IEEE of 2011 International Conference on Complex, Intelligent and Software Intensive Systems (CISIS), 2011, pp. 577-582, doi:10.1109/CISIS.2011.95.
- [7] G. Pilato, A. Augello, and S. Gaglio, "A Modular Architecture for Adaptive Chatbots," Proc. IEEE of 2011 Fifth IEEE International Conference on Semantic Computing (ICSC), 2011, pp. 177-180, doi:10.1109/ICSC.2011.68.
- [8] H. Al-Zubaide and A. A. Issa, "OntBot: Ontology Based Chatbot," Proc. IEEE of 2011 Fourth International Symposium on Innovation in Information & Communication Technology (ISIICT), 2011, pp. 7-12, doi:10.1109/ISIICT.2011.6149594.
- [9] J. P. McIntire, L. K. McIntire, and P. R. Havig, "Methods for Chatbot Detection in Distributed Text-Based Communications," Proc. IEEE of 2010 International Symposium on Collaborative Technologies and Systems (CTS), 2010, pp. 463-472, doi:10.1109/CTS.2010.5478478.
- [10] Y. Wu, G. Wang, W. Li, and Z. Li, "Automatic Chatbot Knowledge Acquisition from Online Forum via Rough Set and Ensemble Learning," Proc. IEEE of 2008 IFIP International Conference on Network and Parallel Computing, 2008, pp. 242-246, doi:10.1109/NPC.2008.24.
- [11] G. Neubig, "NLP Programming Tutorial 2 – Bigram Language Models," Presentation Module of Nara Institute of Science and Technology (NAIST).
- [12] M. Dickinson, "Smoothing," Presentation Module of Dept. of Linguistics, Indiana University, Fall 2009.