# Intermediate Report

## Virtual Assistant For College


Submitted By:

Athira S Pillai(SCT15CS015)
Disha Dinesh Majithia(SCT15CS021)

Guided By:
Smt. Kavitha K V
Assistant Professor
Department Of Computer Science

# CHAPTER 1

# INTRODUCTION

A chatbot is a program that is used to participate in conversations with humans. It uses an appropriate interface for input and output and with the use of AI techniques it can provide realistic answers so the user will think that the communication taking place is with another human. The implementation of such systems varies from using keyword matching, string similarity or complex natural language processing techniques. More sophisticated chatbots could learn from the user input. Nowadays chat bots are used widely in web applications in order to provide help or information when it is asked by the users.

This chapter covers an introduction to the project including the context, a description of aims and objectives, a description of what has been achieved, contributions and the structure of the report.

## 1.1 CONTEXT

Although the admissions process works properly as it is, it is very difficult and time consuming to contact a member of staff of the university. However, the problem would be partially solved if the applicant could talk to a convincing chatbot, able to respond to their concerns with information about admissions. The chatbot should be able to converse with any user in a similar fashion :

Chatbot : Hey, how can I help you?
User: What is the minimum rank I need to have in KEAM to get admissions?
Chatbot : What is the branch of your interest?
User : CS
Chatbot : 7000-8000 to get a merit seat.

## 1.2 AIM

The aim of this project is to contribute to the solution of the problem of direct communication between students and the university.

The main objectives of the project are as follows:

- Database: To develop a database where all the relevant information about questions, answers, keywords, logs and feedback will be stored.

- Algorithm: To develop a keyword matching algorithm and a sentence similarity algorithm and combine them in order to retrieve the best possible answer.
- Interface: To develop a web interface which aims to give the ability to potential students to submit questions to a chatbot and get convincing replies.

**1.3 GOALS**

1. Create an easy to use chatbot with user friendly interface
2. Make the responses of the chatbot as human-like as possible
3. Give appropriate solutions when a response is not known

Work done to achieve goals:

- Various algorithms were referred to, to find the most suitable algorithm
- A combination of two algorithms was what was decided upon as the best algorithm
- As many possible questions were collected from students for chatbot learning
- The basic structure of database was decided on
- Basic learning of python language

# CHAPTER 2

# PROBLEM DEFINITION

In the case of students who are about to join a new course, they are always filled with endless queries and rather than flooding a single person with all those to answer over and over again, it is apt to segregate usual common queries along with their solutions at a single place which is easily accessible by all and as the college has no provision for any such system the need for a virtual assistant is imminent in this scenario.

## 2.1 OVERVIEW

The role of the system is to provide a chatbot that will be able to answer questions related to the admission procedure. It will provide a web interface for the users to interact with the system and an administration interface.

A user is anyone who would like to visit the website and engage in a conversation. As well as talking the user should be able to submit a log of whether he is satisfied with an answer and produce a link. Other than communicating, the user should be able to rank the system. The rank of the system should be a five star ranking system where one is poor and 5 is excellent. A user should also be able to write a review using the feedback form.

The administrator of the system shall be able to log in using a user name and a password. The responsibility of the administrator will be to maintain the system by adding questions and answers to the database and by updating current information sets when necessary. Furthermore he should be able to view the user ranking, feedback messages and logs.
There are several security issues which need to be taken into consideration when designing the system. These include personal and sensitive information. The data should not be accessed without authorisation and authentication. Disclosure or leak of data should be protected against various types of attacks and be encrypted and password protected.

## 2.2 FUNCTIONAL REQUIREMENTS

1. Chatting
    a. The system should allow users to chat.
    b. The system shall inform the user if an answer is not available.

2. Searching
    a. The system should allow users to search for information about admissions.
    b. The system should allow users to search for information about tuition fees.
    c. The system should allow users to search for information about accommodation(for girls).
3. Logs
    a. The system should maintain a log of the current question and answer if the user is not satisfied.
4. Feedback
    a. The user should be able to leave feedback, which is comprised of a text message and a rating
5. Administrative system
    a. Information management: The administrator should be able to add, update and delete questions, answers and keywords.
    b. Log management: The administrator should be able to view and delete logs.
    c. Feedback management: The administrator should be able to view and delete feedbacks.

## 2.3 NON-FUNCTIONAL REQUIREMENTS

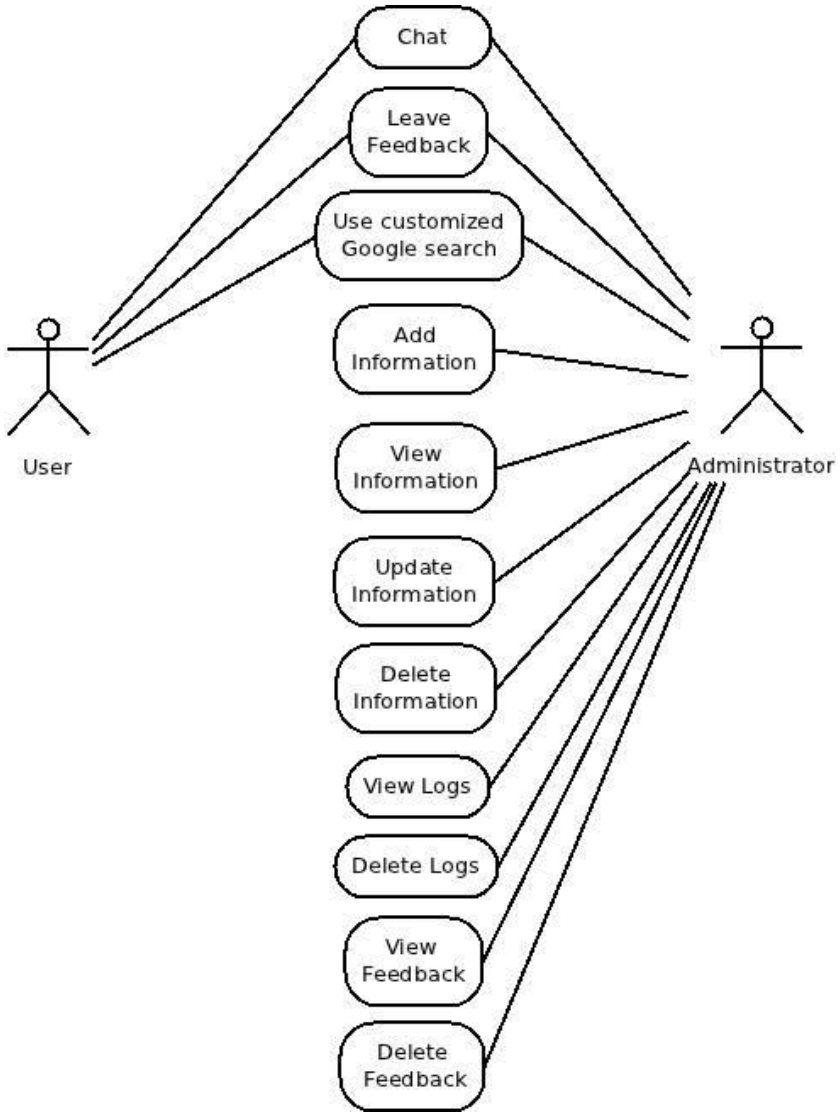
1. User Interface
    a. The system shall maintain an easy to use interface across all functionality and for all users
    b. The client's user interface should be compatible with all commonly used browsers, such as Internet explorer, Firefox, Google chrome and Safari.
2. Security
    a. The administrative system should be protected from unauthorized access.
    b. The database should protected from attacks and unauthorized access.
    c. The interface should be protected from attacks.
    d. All passwords should be stored as a secure hash of the administrator password

3. Maintainability
    a. The system should be easy to maintain.
    b. There should be a clear separation of HTML and interface
4. Exception handling
    a. Exceptions should be reported effectively to the user if they occur.

5. Ethics
    a. The system shall not store or process any information about its users.

## 2.4 USE CASE MODEL

The use case diagram describes the functionality of the system as designed from the requirements and can be found below.




Actor Documentation

User: Someone who uses the system to engage in a conversation

Administrator: The administrator of the system. Allowed to carry out administrative tasks.

# CHAPTER 3

# LITERATURE REVIEW

Generally speaking a bot is any software that performs an automated task, however we are interested in the class of bots that live in chat platforms called chatbots. A chatbot is an instant messaging account that is able to provide services using instant messaging frameworks with the aim of providing conversational services to users in an efficient manner. The past research on chatbots (also known as conversational agents) can be examined using quantitative bibliometric analysis. Bibliometric analysis is a fundamental and powerful method to explore the patterns and future trends of a research topic. There are various platforms to create a chatbot and we select one based on the functionality of our chatbot. One such platform to create chatbots is IBM Watson. It has three main ingredients Intent, Entities and Dialog. NLP is used to have computer-human interaction using natural human languages. Machine learning (ML) helps the chatbot to learn from the data and experience. Databases are used to store the huge amount of information used by the chatbot. It helps in learning as well as forming responses. NLP and ML also happen to be the biggest challenges during chatbot creation. Chatbot with complex queries handling need high attention in using singular and plural forms, need to take care of synonyms, hyponyms, and finally, the sentimental analysis should be done carefully. In short, chatbot is ecosystem and moving quite fast and with the passage of time new features are added in the existing platform. Recent advancements in the machine learning techniques may able to handle complex conversation issue such as payments correctly.

# CHAPTER 4

# DESIGN

## 4.1 DATABASE DESIGN

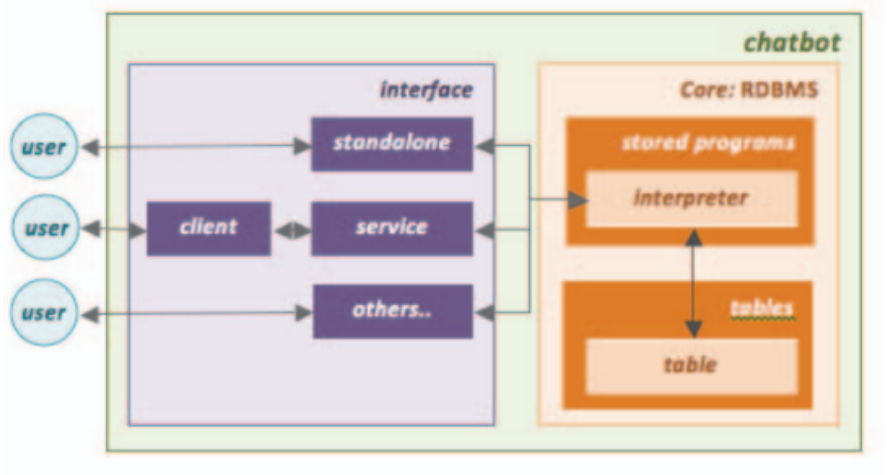
A relational database management system (RDBMS) is a program set that is employed to define, manage, and process a database. The database is a structure that is built to be functioned as data storage. MySQL is a server database or RDBMS software that can manage the database and can store data in many numbers. It can be accessed by multi-user and can do multi-threaded. There are some methods applied to the pattern similarity process. A sentence-similarity measurement scores which employed is to obtain similarity level between both input and pattern. This process is done in the RDBMS. Before entering design process, it needs to know global architecture of the chatbot.

The chatbot consists of core and interface accessing that core. The core is in RDBMS being database. The database consists of tables to store knowledge, while the interpreter is a stored program of function and procedure sets for requiring of pattern matching. The interface could be a standalone application that can be employed by user for chatting or conversation. It can also be employed by service that needs additional client application to converse with the user. This application in the interface side can be expanded more over as user needs it and can also be written using other programming languages. The development of chatbot core that is in the RDBMS covers knowledge storage and pattern matching process. The boundaries of chatbot in this method have some requirements:

1. Chatbot should be able to differ each conversation session that is running, so it has to store data due to the conversation session such as session of identity (sessionid), user name on the session. The sessionid must always be sent together with user input by application in the interface side along conversation process
2. Chatbot must store knowledge in the pattern-template form
3. All user inputs must be free of misspellings, punctuation, and must be in lower case, so to anticipate these cases the chatbot should be able to do normalization of the input that doesn't fit.
4. For the purposes of misspellings correction, the chatbot should have a list of misspelled words and the correction stored in the tables of database
5. The chatbot should be able to pick up the keywords from the user input, so the chatbot should have a list of keywords which is stored in the tables of database.

6. The chatbot should be able to do a search template using a sentence-similarity measurement scores between both pattern and input. The searching of pattern is narrowed based on the result identification as described at point 4
7. The chatbot should keep a conversation log containing sessionid, time, input, and response.

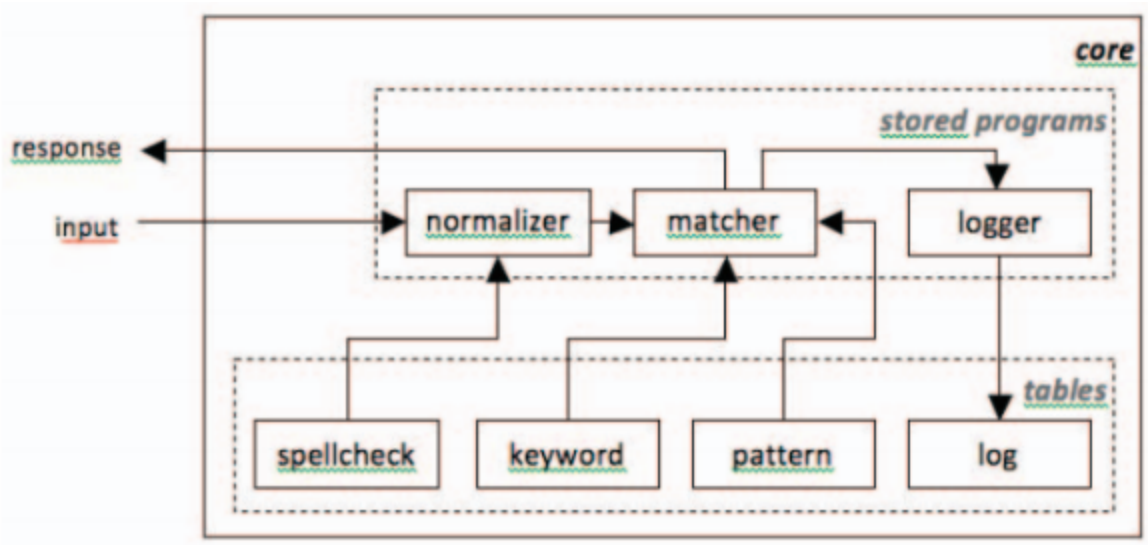
Some points those have been identified above provide a global description scheme of the chatbot core related with conversation processing




.

## I. TABLES

The table blocks consist of main tables as scheme and supporting tables i.e. spellcheck, keyword, pattern, and log. The spellcheck stores list of misspelling words from user input and correction. The keyword stores list of keywords that is probably found in the user input. The keywords are used to narrow the range of pattern searching. The pattern stores pattern-template pairing and roles as main knowledge. Each pattern could be paired with one or more templates and each template could be paired with one or more patterns, so it would be divided into three tables i.e. pattern, template, and pattern-template. The pattern is functioned to store pattern containing patternid and pattern, template is functioned to store template containing templateid and template, and the pattern-template roles as table connecting pattern and template tables. The convlog stores conversation history containing sessionid, time, user input, and response given by user. The session stores attributes of such conversation session including sessionid, user name, and other attributes. The array is a temporary table outside the main tables. This table is only supporting for internal operation requirements of array operation function. And tid is used as supporting table for id generator process that is applied in other requiring tables. These tables contain id as id name, counter as id counter, and rtable as table name reference.




## II. STORED PROGRAMS

The stored program is containing stored procedures and functions for patternmatching requirements. The normalizer is a function to norm user input that corrects spellings, eliminates punctuations, and changes into lower case. The matcher is a main function in pattern matching to find appropriate template based-on the sentence-similarity measurement scores between both input and pattern. Before doing pattern-matching process, it needs to take keywords on the input, so it is just pattern with same keywords that will be matched. The logger is a procedure functioned as conversation history storing for user input and response of question. The array is a set of functions and procedures for array tables used for the internal operations of 3 major processes that have been described previously. The array of functions and procedures includes array intersect() which is a procedure for operating the intersection of two arrays, array push which is a procedure for operating the push element to the array, array pop() which is a function for pop operation element of the array, array count() which is a function to count the number of elements in the array, array clear() which is a procedure to clear the contents of the array, and bigram which is a procedure for preparing bigram of a string.

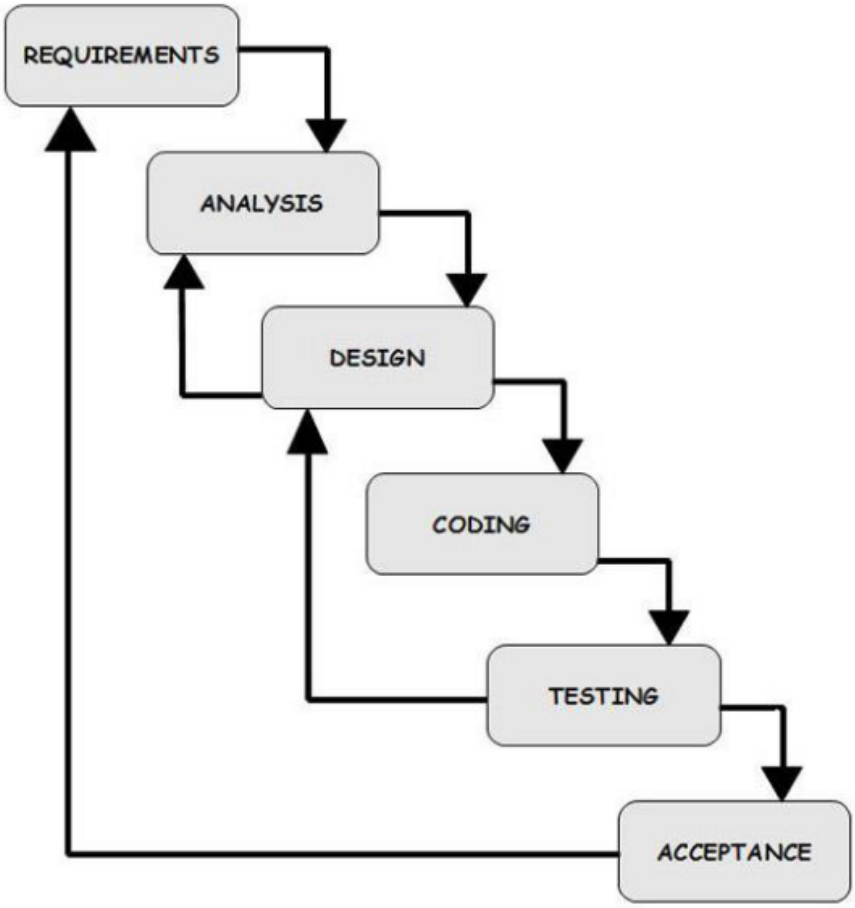
## III. GOOGLE SEARCH

In the case that a user is not satisfied by a system reply then he will be provided with a link from the admissions website. The link will be retrieved using customized Google search, according to the input submitted.

# CHAPTER 5

# METHODOLOGIES

The design and development of the system followed the waterfall model as described below.




The waterfall model follows a series of processes, which are used during development. Usually the stages will require the gathering of requirements and their analysis. The design of the system is the next stage, followed by coding the actual system. Then evaluation, testing and debugging, if necessary, is the next step. Finally the system will either be accepted and therefore maintained or rejected. It is vital to move to the next process of the waterfall model if the previous step has been completed.

## 5.1 ALGORITHMS

### 5.1.1 KEYWORD MATCHING

The keyword matching algorithm will attempt to identify keywords in a sentence. In the case that one or more keywords are found in the user's input text then an answer will be retrieved.

### 5.1.2 PATTERN MATCHING

The pattern-template storage as knowledge in the relational database management system (RDBMS) or called as database allows the use of structured query language (SQL) to handle the process of pattern-matching. The RDBMS already available in many builtin functions or procedures and can be made user-defined stored program which can be called using SQL. This allows many programming languages can be implemented easily as query of the database to send and receive an input response

### 5.1.3 SENTENCE SIMILARITY MEASUREMENT

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

## 5.2 COMBINATION OF ALGORITHMS

In order to achieve the best possible outcome we decided to combine the two algorithms. For example the keyword matching algorithm will search the keywords and the string similarity algorithm will search all pre-stored questions with or without keywords. In order to improve this similar questions have been added that have the same answer. Basically if the keyword matching fails then we try to think the way a simple user would think, asking a question.

# CHAPTER 6

# DATASET

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | Timestamp | What would you like to know about the colle | Please frame a proper response for the question you submitted above. | | | |
| 2 | 11/27/2018 20:21:55 | Where is the college located? | The college is at Pappanamcode, near to Central Works KSRTC. | | | |
| 3 | 11/27/2018 20:22:29 | What is the full name of the college? | Sree Chitra THIRUNNAL College of Engineering | | | |
| 4 | 11/27/2018 20:24:25 | How many departments does the college ha | The college has 6 departments - Computer Science, Electronics and communication, Mechanical En | | | |
| 5 | 11/27/2018 20:30:37 | How far is the college from the nearest railw | It is 4 kms from the nearest railway station which is Thiruvananthapuram Central. | | | |
| 6 | 11/27/2018 20:37:26 | Does the college provide Civil Engineering? | No, the college does not have Civil Engineering department currently. | | | |
| 7 | 11/27/2018 20:38:46 | Who is the principal of the college? | Prof.Dr.K.Prabhakaran Nair is the current Principal. | | | |
| 8 | 11/27/2018 20:47:36 | Who is the Academic Dean of the college? | Dr. Jayasudha J. S is the Academic Dean. | | | |
| 9 | 11/27/2018 20:48:18 | Does the college have lab facilities? | Yes, the college does provide the required lab facilities. | | | |
| 10 | 11/27/2018 20:49:25 | Does the college have any clubs? | Yes, the college has a Literature and Debate Club, Film Club, Cycling Club, Photography Club. | | | |
| 11 | 11/27/2018 20:52:09 | Does the college provide transport services? | Yes, the college has 2 bus which picks up students from different parts of Trivandrum. | | | |
| 12 | 11/27/2018 20:52:43 | How many students are there in each class? | There are 50-60 students in each class. | | | |
| 13 | 11/27/2018 20:53:28 | Are the students allowed to participate in ext | Yes, students are allowed to participate in various Cultural and Sports activities. | | | |
| 14 | 11/27/2018 21:02:19 | Does the college have playground? | No the college does not have a huge playground | | | |
| 15 | 11/27/2018 21:12:16 | Does the college provide hygienic facilities? | Yes, there are several maintenance workers who work hard to maintain the college hygiene. | | | |
| 16 | 11/27/2018 21:14:22 | Are students allowed to work on projects of t | Yes, students are allowed to pursue and also connected with professionals for their help. | | | |
| 17 | 11/27/2018 21:41:41 | Does the college have a canteen service? | Yes, the college has a canteen service from where students can have their breakfast and lunch. | | | |
| 18 | 11/27/2018 21:42:22 | What is a mini canteen? | A service provided for students and faculty members to have tea and snacks during lunch break or e | | | |
| 19 | 11/27/2018 21:43:39 | Is the college working on Saturday? | To cover up portions, classes are scheduled for certain Saturdays otherwise only the faculty has to at | | | |
| 20 | 11/27/2018 21:44:15 | Does the college provide the course Aerospa | No, currently the college does not provide the course. | | | |

Collected using Google Forms from students of SCT College.

# CHAPTER 7

# SCHEDULE

| TASKS | START DATE | END DATE | STATUS |
|---|---|---|---|
| Problem Statement Identification and Formation | | 19-7-2018 | Complete |
| Requirement Analysis and study on present scenario | 3-08-2018 | 14-09-2018 | Complete |
| Study of the desired algorithm | 20-09-2018 | 28-09-2018 | Complete |
| Learn Python programming and use of libraries | 1-10-2018 | | Ongoing |
| Collection of data | 1-10-2018 | 30-11-2018 | Ongoing |
| Build database and the User Interface | 30-11-2018 | 30-1-2019 | Incomplet |
| Build network modules | 1-2-2019 | 20-2-2019 | Incomplet |
| Unit test the bot | 25-2-2019 | 1-3-2019 | Incomplet |
| Integration testing | 2-3-2019 | 10-3-2019 | Incomplet |
| Deploy over network | | 25-3-2019 | Incomplet |

# CHAPTER 8

# CONCLUSION

The main objectives of the project are to develop an algorithm that will be used to identify answers related to user submitted questions. To develop a database where all the related data will be stored and to develop a web interface. The web interface developed will have two parts, one for simple users and one for the administrator. A background research including an overview of the conversation procedure and any relevant chatbots available will be done. A database shall be developed, which stores information about questions, answers, keywords, logs and feedback messages. A usable system shall be designed, developed and deployed to the web server on two occasions. An evaluation of data collected by potential students of the college. Also after received feedback from the first deployment, extra requirements shall be introduced and implemented.

# CHAPTER 9

# REFERENCES


1. B A Shawar and E. Atwell, “Chatbots. Are they really useful?” LDV Forum, vol,. 22,pp 29-49 2007
2. Chatbot for University Related FAQs by Bhavika R Ranoliya, Nidhi Raghuwanshi and Sanjay Singh from Manipal University of Technology
3. Chatbots and Conversational Agents: A Bibliometric Analysis by H.N.Io, C.B.Lee
4. Long Short-Term Memory Networks for Automatic Generation of Conversations by Tomohiro Fujita Wenjun Bai Changqin Quanof Department of Computational Science Graduate School of System Informatics, Kobe University Kobe, Japan (2016)
5. A Generation Method of Back-channel Response to Let a Chatting bot Be a Member of Discussions by Yoko Nishihara, Masaki Ikuta, Ryosuke Yamanishi, and Junichi Fukumoto at 2017 6th IIAI International Congress on Advanced Applied Informatics
6. Learning to predict the adequacy of answers in chat-oriented humanagent dialogs by Luis Fernando D’Haro and Rafael E. Banchs at 2017 IEEE Region 10 Conference (TENCON), Malaysia
7. Development and Implementation of a Chat Bot in a Social Network at Ninth International Conference on Information Technology- New Generations