**ABSTRACT**

To start a good life healthcare is more important. But it is very difficult to obtain the consultation with the doctor in case of any health issues. The proposed idea is to create a health care Chabot system using Artificial Intelligence that can diagnose the disease and provide basic details about the disease before consulting a doctor. The system provides text (or) voice assistance that means user can use own convenient language. Bot will provide which type of disease you have based on user symptoms and appeared doctor details respective to user disease. The Chabot will clarify the users symptoms with serious of questions and the symptom conformation will be done. The disease will be categorized as minor and major disease. Chatbot will reply whether it is a major or minor disease. If it is a major disease user will be suggested with the doctor details and analagesics for further treatment and also provides food suggestion that means which type of food you have to take. The user can achieve the real benefit of a chatbot only when it can diagnose all kind of disease and provide necessary information. A text-to-text diagnosis Bot engages patients in conversation about their medical issues and provides a personalized diagnosis based on their symptoms. Hence, people will have an idea about their health and have the right protection.

TABLE OF CONTENTS

LIST OF FIGURES

**CHAPTER 1**

**PREAMBLE**

**a. INTRODUCTION**

Now a days, health care is very important in our life. Today‘s people are busy with their works at home, office works and more addicted to Internet. They are not concerned about their health .So they avoid to go in hospitals for small problems. it may become a major problem. So we can provide an idea is to create a health care Chabot system using AI that can diagnosis the disease and provide basic information about the disease before consulting a doctor. Which helps the patients know more about their disease and improves their health User can achieve the all kind of disease information. The system application uses question and answer protocol in the form of chatbot to answer user queries. The response to the question will be replied based on the user query. The significant keywords are fetched from the sentence and answer to those sentences. If match is discovered or significant answer will be given or similar answers will be displayed. Bot will diagnosis which type of disease you have based on user symptoms and also gives doctor details of particular disease.It may reduce their health issues by using this application system. The system is developed to reduce the healthcare cost and time of the users as it is not possible for the users to visit the doctors or experts when immediately needed.

Artificial Intelligence is based on how any device perceives its Environment and takes actions based on the perceived data to achieve the result successfully. It is the study of intelligent agents. The term "artificial intelligence" is applied when a machine mimics "cognitive" functions that humans associate with other human minds, such as "learning" and "problem solving. Artificial Intelligence gives the supreme power to mimic the human way of thinking and behaving to a computer. A chatbot (also known as a talkbot, chatterbot, Bot, IMbot, interactive agent, or Artificial(ConversationalEntity) is a computer program which conducts a conversation via auditory or textual methods. These programs are designed to provide a clone of how a human will chat and thereby it acts as a conversational partner rather than humans. For various practical purposes like customer service or information acquisition, , chatbot is being used in the dialog system. Mostly chatbots uses natural language processing for interpreting the user input and generating the corresponding response but certain simpler systems searches for the keyword within the text and then provides a reply based on the matching keywords or certain pattern. Today, chatbots are part of virtual assistants such as Google Assistant, and are accessed via many organizations' apps, websites, and on instant 2

messaging platforms. Non-assistant applications include chat bots used for entertainment purposes, for research, and social bots which promote a particular product, candidate, or issue.

Chatbot‘s are such kind of computer programs that interact with users using natural languages. For all kind of chat bots the flow is same, though each chatbot is specific in its own area knowledge that is one input from human is matched against the knowledge base of chatbot. Chatbot‘s work basically on Artificial intelligence so using this capability we have decided to add some contribution to the Health Informatics.The high cost of our healthcare system can often be attributed to the lack of patient engagement after they leave clinics or hospitals. Various surveys in this area have proved that that chatbot can provide healthcare in low costs and improved treatment if the doctors and the patient keep in touch after their consultation. To answer the questions of the user chatbot is used. There is very less number of chatbots in medical field.

The main purpose of the scheme is to build the language gap between the user and health providers by giving immediate replies to the Questions asked by the user. Today‘s people are more likely addicted to internet but they are not concern about their personal health. They avoid to go in hospital for small problem which may become a major disease in future. Establishing question answer forums is becoming a simple way to answer those queries rather than browsing through the list of potentially relevant document from the web. Many of the existing systems have some limitation such as There is no instant response given to the patients they have to wait for experts acknowledgement for a long time. Some of the processes may charge amount to perform live chat or telephony communication with doctors online.This system allows computer to communication between human to computer by using natural language processing (NLP). There are three analyses which understand natural language i.e. identification of main linguistic relations is completed to parse subject into object of the sentences. After that Chatbot is an Entity which imitate human discussion in its particular accepted set-up together with a text or vocal language with techniques such as Natural Language Processing (NLP). The aim of this system is to replicate a person‘s discussion. The development of chatbot application canbe done with making a user interface to send input and receive response. It is a system that interact with user by keeping the track of the state of interaction and recollecting the preceding commands to give functionality. The medical chat-bots can be developed by using artificial algorithms that scrutinize user's queries and recognize it and give reply to related query. A big disease can start from small problems such as headache which feels normal but it may beginning of big disease such as brain tumor 3

most of the disease can be identified by common symptoms so the disease can be predicted if the patient body is analyzed periodically[6].The system give response by use of an efficient Graphical User Interface such that if actual person is chatting with the user. chatterbot that can be used in various fields like education, healthcare, and route assistance[8]The central part of the chat-bots includes MySQL. It is an interactive system solve users query regarding medicine. so they can get correct guidance for treatment through android app by using Google API.

Computers give us information; they engage us and help us in a lot of manners. A chatbot is a program intended to counterfeit smart communication on a text or speech. Yet, this paper concentrates only on text.

These systems can learn themselves and restore their knowledge using human assistance or using web resources. This application is incredibly fundamental since knowledge is stored in advance. The system application uses the question and answer protocol in the form of a chatbot to answer user queries. This system is developed to reduce the healthcare cost and time of the users, as it is not possible for the users to visit the doctors or experts when immediately needed.

The response to the question will be replied based on the user query and knowledge base. The significant keywords are fetched from the sentence and answer to those sentences. If the match is discovered or the significant answer will be given or similar answers will be displayed.

The complex questions and answers present in the database are viewed and answered by an expert. Here the users can personally ask any questions regarding healthcare, as not much time will be wasted by the user for consulting a doctor. The input sentence of the chat pattern is stored in an RDBMS. The chatbot would coordinate the input sentence from the user question with the knowledge base. Each query is compared with the knowledge database of the chatbot. The important keywords are extracted from the given input sentence and the sentence similarity is found. The keyword ranking and sentence similarity are found using the N-gram, TF-IDF, and cosine similarity. The interfaces are standalone built using the JAVA programming language.

WHAT IS HEALTHCARE CHATBOT?

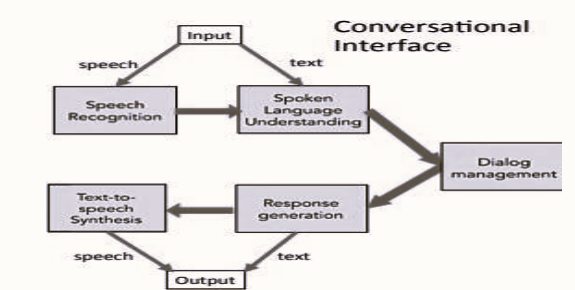
In general terms, a bot is nothing but a software that will perform automatic tasks. In other terms, a bot is a computer program that is designed to communicate with human users through the internet. The most natural definition of a chatbot is – a developed a program that can have a discussion/conversation with a human. Healthcare chatbots are software integrated into either a website or mobile app and possess any combination of the following attributes:

**Natural language processing** :The ability for a chatbot or other form of artificial intelligence to understand aspects of human language including syntax and semantics from users of various backgrounds.

**Contextual AI**: contextual AI takes into consideration the intent of the user as they ask questions and submit responses, thus ensuring the best relevant conversation takes place.

**Chatbot Architecture**

Previously chat bots solely supported a single adjacency pair, also known as a one-shot conversation. However, modern chatbots can sustain multiple adjacency pairs, remembering states and contexts between conversations and have the capability to associate data in different adjacency pairs which is related. This is the bots ability to preserve the conversation. A chatbot consists of four main parts: front- end, knowledge-base, back-end and corpus which is the training data. The front end is accountable for enabling communication between the bot and the user. The NLU utilises Artificial intelligence methods to identify the intent and context of the user input.



An appropriate response is generated from the user’s intent. The knowledge base defines the chatbot knowledge, which is created within the NLU and supported by the back-end, the back-end applies the domains corpus to produce the knowledge base. Input can be supplied to the chatbot in the form of text or speech. The Input is sent to the dialog management system which is the NLU in this case, which determines an appropriate response and amends the chatbots state accordingly to carry out the required action. The chatbot will produce text and speech responses in the form of both text and speech.

**b. Problem Statement**

**CHAPTER 2**

**SYSTEM DESIGN**

INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

* What data should be given as input?
* How the data should be arranged or coded?
* The dialog to guide the operating personnel in providing input.
* Methods for preparing input validations and steps to follow when error occur.

OBJECTIVES

1.Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.

2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.

3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus, the objective of input design is to create an input layout that is easy to follow.

OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system‘s relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.

2. Select methods for presenting information. 3.Create document, report, or other formats that contain information produced by the system. The output form of an information system should accomplish one or more of the following objectives.

* Convey information about past activities, current status or projections of the Future.
* Signal important events, opportunities, problems, or warnings.
* Trigger an action.
* Confirm an action.

**CHAPTER 3**

**SOFTWARE REQUIREMENT SPECIFICATIONS**

**3.1 Functional Requirements**

Hardware Requirement

Processor : INTEL CORE i7 2.60 GHz

RAM : 16GB or Higher

Hard Disk : 1TB

Software Requirements

Software Description

**3.2 Non-Functional Requirements**

**CHAPTER 4**

**IMPLEMENTATION/METHODOLOGY**

Natural language processing: Natural language processing involves the reading and understanding of spoken or written language through the medium of a computer. This includes, for example, the automatic translation of one language into another, but also spoken word recognition, or the automatic answering of questions. Computers often have trouble understanding such tasks, because they usually try to understand the meaning of each individual word, rather than the sentence or phrase as a whole. So for a translation program, it can be difficult to understand the linguistic nuance in the word Greek when it comes to the examples My wife is Greek and It’s all Greek to me, for example Through natural language processing, computers learn to accurately manage and apply overall linguistic meaning to text excerpts like phrases or sentences. But this isn’t just useful for translation or customer service chat bots: computers can also use it to process spoken commands or even generate audible responses that can be used in communication with the blind, for example. Summarizing long texts or targeting and extracting specific keywords and information within a large body of text also requires a deeper understanding of linguistic syntax than computers had previously been able to achieve.

How does natural language processing work?

It does not matter whether its processing an automatic translation or a conversation with a chat bot all natural language processing methods are the same in that they all involve understanding the hierarchies that dictate interplay between individual words. But this isn‘t easy – many words have double meanings. Pass for example can mean a physical handover of something, a decision not to partake in something, and a measure of success in an exam or another test format. It also operates in the same conjugation as both a verb and a noun. The difference in meaning comes from the words that surround pass within the sentence or phrase (I passed the butter/on the opportunity/the exam).

These difficulties are the main reason that natural language processing is seen as one of the most complicated topics in computer science. Language is often littered with double meanings, so understanding the differences requires an extensive knowledge of the content in which the different meanings are used. Many users have first-hand experience of failed communication with chat bots due to their continued use as replacements for live chat support in customer service. But despite these difficulties, computers are improving their 63 understanding of human language and its intricacies. To help speed this process up, computer linguists rely on the knowledge of various traditional linguistic fields: The term morphology is concerned with the interplay between words and their relationship with other words.

Syntax defines how words and sentences are put together Semantics is the study of the meaning of words and groups of words Pragmatics is used to explain the content of spoken expressions And lastly, phonology covers the acoustic structure of spoken language and is essential for language recognition.

We are using three algorithms to implement making health care chatbot using NLP technique

1.n-gram algorithm.

2.TF-IDF (term frequency-inverse data frequency).

3.Cosine similarity algorithm.

**N-gram algorithm**

In the fields of computational linguistics and probability, an n-gram is a contiguous sequence of n items from a given sample of text or speech. The item scan be phonemes, syllables, letters, words or base pairs according to the application. The n-grams typically are collected from a text or speech corpus. When the items are words, n-grams may also be called shingles Using Latin numerical prefixes, an n-gram of size 1 is referred to as a "unigram"; size 2 is a "bigram" (or, less commonly, a "diagram"); size 3 is a "trigram". English cardinal numbers are sometimes used, e.g., "four-gram", "five-gram", and so on. In computational biology, a polymer or oligomer of a known size is called a k-mer instead of an n-gram, with specific names using Greek numerical prefixes such as "monomer", "dimer", "trimer", "tetramer", "pentamer", etc., or English cardinal numbers, "one-mer", "two-mer", "three-mer" etc.

**TF-IDF(term frequency-inverse data frequency):**

**Term frequency( tf ):**

Suppose we have a set of English text documents and wish to rank which document is most relevant to the query, "the brown cow". A simple way to start out is by eliminating documents that do not contain all three words "the", "brown", and "cow", but this still leaves many documents. To further distinguish them, we might count the number of times each term occurs in each document; the number of times a term occurs in a document is called its term frequency. However, in the case where the length of documents varies greatly, adjustments are often made (see definition below). The weight of a term that occurs in a document is simply proportional to the term frequency.

Often times, when building a model with the goal of understanding text, you‘ll see all of stop words being removed. Another strategy is to score the relative importance of words using TF-IDF.

The number of times a word appears in a document divided by the total number of words in the document. Every document has its own term frequency.



 Exmaple1: I am suffering from fever.

 Example2:I am suffering from headache.

 Tf=number of occurances/number of words in document

 Example1: Example2:

 Tf(I)= 1/5=0.2 Tf(I)= 1/5=0.2

 Tf(am)=1/5=0.2 Tf(am)=1/5=0.2

 Tf(suffering)=1/5=0.2 Tf(suffering)=1/5=0.2

 Tf(from)=1/5=0.2 Tf(from)=1/5=0.2

 Tf(fever)=1/5=0.2 Tf(headache)=1/5=0.2

**Inverse Data Frequency (IDF):**

Because the term "the" is so common, term frequency will tend to incorrectly emphasize documents which happen to use the word "the" more frequently, without giving enough weight to the more meaningful terms "brown" and "cow". The term "the" is not a good keyword to distinguish relevant and non-relevant documents and terms, unlike the less-common words "brown" and "cow". Hence an *inverse document frequency* factor is incorporated which diminishes the weight of terms that occur very frequently in the document set and increases the weight of terms that occur rarely.

The log of the number of documents divided by the number of documents that contain the word ***w***. Inverse data frequency determines the weight of rare words across all documents in the corpus.



Exmaple1: I am suffering from fever.

Example2: I am suffering from headache.

Idf=log(number of documents/number of documents that containing a word)

Idf(I) = log(2/2) = 0

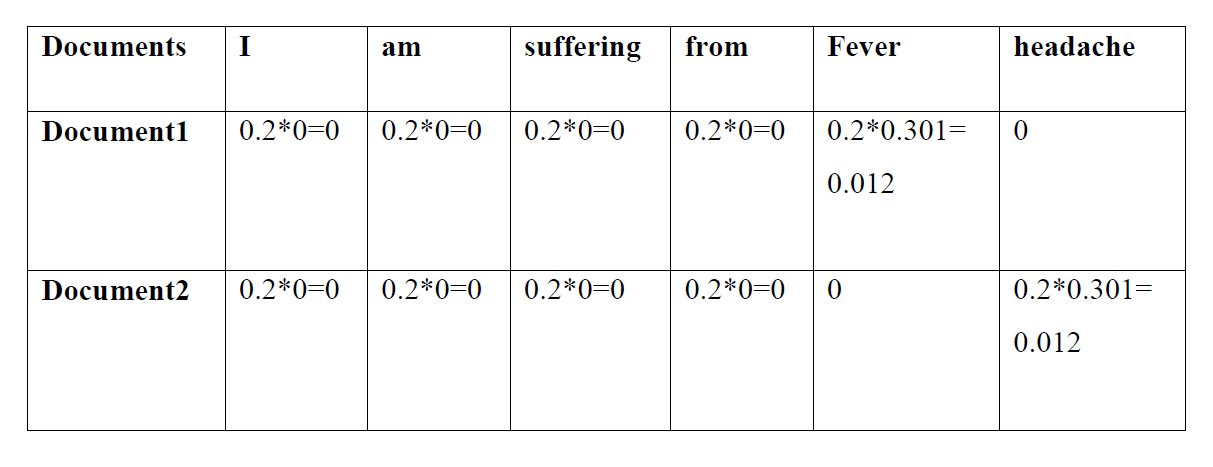
Idf(am) = log(2/2) = 0

Idf(suffering) = log(2/2) = 0

Idf(from) = log(2/2) = 0

Idf(fever) = log(2/1) = 0.301

Idf(headache)= log(2/1) = 0.301

**Tf -idf = Tf\*idf**

**CODE**