

A Project Report
On
"Career path chatbot"
A System for counselling

BY:

Fatmah Maid Alharthi	43708440
Hanan Atya Alzhrani	43706483
Bedoor Awad Alzaidi	43708107
Eman Saleh Alrubeie	43706129

3rd Bsc CS

2019-2020

Table of Contents

1. Abstract	3
2. Introduction	4
1.1. What is a chatbot?.....	4
1.2. Common uses of chatbot.....	4
1.3. Natural Language Processing	5
1.4. Related Works.....	5
2. Problem Statement.....	6
3. Methodology.....	6
3.1. NLTK Library in python	6
3.2. System Implementation	7
3.3. Testing & Result.....	10
4. Conclusion and future works	11
5. Bibliography.....	12

Table of Figures

Figure 1: The Framework of chatbot.....	7
Figure 2: 1st result of testing chatbot.....	9
Figure 3: 2nd result of testing chatbot.....	9
Figure 4: 3rd result of testing chatbot.....	9

Abstract

ChatBot is an automated answering system that simulate human conversation. We aim to develop a chatbot that provides a career counseling by analyzing users' questions and answering these questions as if it were answered by an expert or a consultant. This Chatbot helps users to know enough information about any career path including skills needed, work environment, salary, etc, thus, the user decision about their desired career path can be supported using the proposed chatbot.

1. Introduction

1.1. What is a chatbot?

Chatbot is an automated conversational agent that interacts with users in a natural language based on artificial intelligence software. When a user inputs any content, the chatbot processes it then generate an appropriate response. [1]

There are two types of chatbot which are Rules-based chatbot, machine learning based chatbot. The rule-based chatbot is bounded in such way that it can only respond to specific commands. If it received unexpected input, it will not response to it.

However, the machine learning-based chatbot has the ability to understands language, not only the commands. This type gets smarter as it learns from the conversations it has with the users. [2]

The real power of chatbots presented by its ability to provide an immediate answer to the questions, even faster than human, by being available 24/7 and by its ability to efficiently handle multiple chats at the same time without any delay [3].

1.2. common uses of chatbot

There are multiple uses of chatbot including :

Customer Support : Different companies are using chatbots to provide 24/7 customer service. The goal of customer support is to quickly provide answers and handle customer complaints or simply track the order.

Weather : Chatbot can used to ask about the actual conditions.

Personal Assistance: Chatbots helps users to perform everyday tasks e.g. sending email , reminder appointment ,etc.

These are just examples of common uses of chatbots, There are limitless possibilities for what chatbots can be used for. [4]

1.3. Natural Language Processing

Natural language processing (NLP) is the study of the ability to understand human language and process it in the same way that a human does.

NLP enables computer to meaningfully interpret the inputs given by the users. For example, in the case of chatbots it understands the intent of the input from the users and then creates responses based on an analysis similar to the human being.

Recent NLP engines uses machine learning to parse user input in order to take out the important objects and understand user intent. Without NLP the AI that analyze the input language is limited.[5]

NLP is applicable in several way ranging from speech recognition, language translation, classifying documents to extract information. [6]

1.4. Related Works

The authors of this work [8] propoed the design of an expert system for educational guidance for students available after SSC & HSC, which helps the students to have the basic idea of the possible career opportunities. The system will help the users who have passed SSC and HSC to select their field of interest or a field that would be best for them in order to build up their future. It also will establish an automated process similar to a one-to-one meeting with a career counselor and will aid to plan a career that takes into account the student's grade, IQ, hobbies, interests, and other predominant specifications entered by the user at the time of registration .

In this paper [9], they proposed the design of a chatbot which is going to help new students to solve all the problems they face and the questions which arises in their mind during and after the admission. In particular, the paper investigates the implementation of ALICE chatbot system as an application named as college enquiry chat bot .

2. Problem Statement

Many people need to consult someone about their career. The number of experts who can help is insufficient for the consulting requests, furthermore they cannot be available 24 /7.

The need of a career counselor to be informatic and 24/7 available is high, thus we propose the developing of a career path counseling chatbot .

Our chatbot will solve this problem by provide a great career consultation with an immediate response, with privacy to the user and with 24/7 availability.

This chatbot will be built using Computer science algorithms such as Artificial Intelligence.

3. methodology:

3.1. NLTK Library in python:

The Natural Language Toolkit (NLTK) is an open-source library in Python programming language. NLTK is one of the most powerful NLP libraries. It contains packages to make the machine understand human language and response to it. It contains Text Analysis processes including tokenization, parsing, classification, stemming, tagging and more.

NLTK also includes more than 50 corpora and lexical sources such as Open Multilingual Wordnet and Problem Report Corpus.[7]

3.2. System implementation

The chatbot works in 3 steps :

First Reading input and then preprocessing the text finally Generating Response. Before these steps, the chatbot need to read the source content that is a knowledge base of the chatbot. The figure below, shows the design of the chatbot

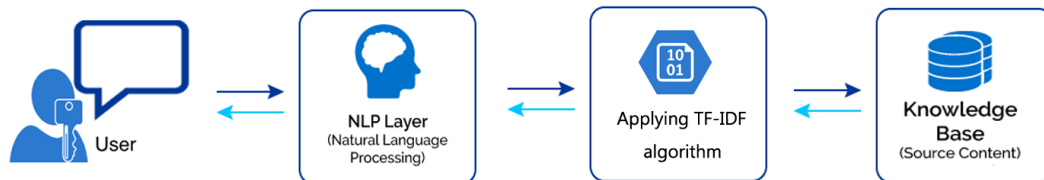


Figure1: The design of chatbot

The explanations of these steps is given below in more detail along with their implementation code.[10]

Reading in the source content

First, we will feed the content that we want our chatbot to say by write the content to the

chatbot.text file that works as a knowledge base of the chatbot and the chatbot will use the file during the processes.

```
with open('chatbot.txt', 'r', encoding='utf8',  
errors='ignore') as fin:  
    raw = fin.read().lower()
```

We choosed four fields of computer science ,which are software engineering, cybersecurity, artificial intelligence and web development, to be the career paths that the chatbot will consult about it . Information about these fields has been included in the source file. At the end, we will test the chatbot by asking it to give counseling about one of these fields.

Reading data and Pre-processing the raw text

Chatbot reads the user's input then Pre-processing the text (user input) using NLTK library to tokenization the text and return tokens.

Tokenization is a way to split text into tokens. These tokens can be paragraphs, sentences, or words.

```
lemmer = WordNetLemmatizer()
def LemTokens(tokens):
    return [lemmer.lemmatize(token) for token in tokens]

remove_punct_dict = dict((ord(punct), None) for punct in string.punctuation)
def LemNormalize(text):
    return
LemTokens(nltk.word_tokenize(text.lower().translate(remove_punct_dict)))\
```

Generating Response

The chatbot creates a response for input question using the Tf-Idf approach.

Term Frequency Inverse Document Frequency (TF-IDF): is approach to rescale the frequency of words by how often they appear in all documents where:

Term Frequency (TF) : is a scoring of the frequency of the word in the document.

$TF = (\text{Number of times the term } t \text{ appears in a document}) / (\text{Number of terms in the document})$

Inverse Document Frequency (I D F): is a scoring of how rare the word is across documents.

$IDF = 1 + \log(N/n)$, where N is the number of documents and n is the number of documents a

term t has appeared in.

Tf-IDF weight: is a statistical measure used to estimate how important a word is to a document

in a collection, often used in information retrieval and text mining.

The Tf-IDF weight is the product of TF & IDF.

TF-IDF is a conversion applied to texts to get two vectors in vector space. We can then obtain the Cosine similarity of any pair of vectors.

Cosine similarity is a measure of similarity between two non-zero vectors.

Using it we can find out the similarity between any two documents . We use it to obtain the Cosine similarity of input & source .

Tf-IDF can be implemented in scikit learn as:

```
from sklearn.feature_extraction.text import TfidfVectorizer
```

From scikit learn library we used the TfidfVectorizer module to convert a collection of raw documents to a matrix of TF-IDF features. Also we used cosine similarity module from scikit learn library ,it will be used to find the similarity between words entered by the user and the words in the corpus.

We define a response function that will search the user's statement for familiar words and returns one of the multi possible responses. If there is no input matching any of the keywords, it will return "sorry! I can't understand you".

```
def response(user_response):
    robo_response = ''
    sent_tokens.append(user_response)
    TfidfVec = TfidfVectorizer(tokenizer=LemNormalize, stop_words='english')
    tfidf = TfidfVec.fit_transform(sent_tokens)
    vals = cosine_similarity(tfidf[-1], tfidf)
    idx = vals.argsort()[0][-2]
    flat = vals.flatten()
    flat.sort()
    req_tfidf = flat[-2]
    if (req_tfidf == 0):
        robo_response = robo_response + "I am sorry! I don't understand you"
        return robo_response
    else:
        robo_response = robo_response + sent_tokens[idx]
        return robo_response
```

3.3. Testing and Result

```
chatbot2
C:\Users\Stars\PycharmProjects\CareerPathChatbot\venv\Scripts\python.exe C:/Users/Stars/PycharmProjects/CareerPathChatbot/chatbot2
COBO: My name is Cobo. I will answer your queries about CS Career path. If you want to exit, type Bye!
hello
COBO: hi
How much the salary of AI engineer?
COBO: the average salary for an ai jobs are around $100,000 to $150,000.
ok what about the salary of software a software engineer?
COBO: -facebook: average salary for software engineers $ 224,003
-dropbox: the average salary for software engineers is $ 250,068
-snapchat: average salary for software engineers $ 243,915.
great can you tell me about the work environment for the software engineering?
COBO: -amazon: average salary for software engineers $ 134,608

the work environment for the software engineering : computer software engineers typically work in well-lit offices in comfortable surroundings or in computer laboratories.
oh thank you
COBO: for example, a customer may want a form to capture an end user's e-mail to request additional information, provide a newsletter, or thank a customer for their business.
thanks
COBO: You are welcome..

Process finished with exit code 0
```

Figure 1: 1st result of testing chatbot

```
chatbot2 x
C:\Users\Stars\PycharmProjects\CareerPathChatbot\venv\Scripts\python.exe C:/Users/Stars/PycharmProjects/CareerPathChatbot/chatbot2
COBO: My name is Cobo. I will answer your queries about CS Career bath. If you want to exit, type Bye!
hello
COBO: I am glad! You are talking to me
tell me what do we mean by the science of artificial intelligence?
COBO: artificial intelligence is a branch of computer science where the science of computing in turn leads to the emergence of artificial intelli
ok what are the best universities to study artificial intelligence in ?
COBO: it is unfortunate that this specialization does not exist in most universities, but there are some of the most important universities that
carnegie mellon university, stanford university, massachusetts university of technology, california university, harford university, yale universi
and the university of texas.
if I graduated from Artificial Intelligence, where I can work?
COBO: if you graduated from artificial intelligence you can work in ministries, schools, universities, libraries, criminal centers, companies, in
you can also work in the field of teaching and training others in the use of software and artificial intelligence.
what kind of jobs should I start looking for?
COBO: what kind of jobs should you start looking for?let's take a closer look to ai jobs: machine learning engineer ,data scientist,,business int
and many other jobs.
```

Figure 2: 2nd result of testing chatbot

```
COBO: My name is Cobo. I will answer your queries about CS Career path. If you want to exit, type Bye!
hello
COBO: hi there
do you know how can I become a web developer?
COBO: the skills of the web developer
1-first skill programming (-html, the basic language in building websites
-programming languages used in the web domain (# java - php - python - asp.net and c)
-language of databases and database systems( sql)
-the web developer should can do some simple tasks like creating registration pages, logging in and out, commenting pages, adding articles))
2-continuous learning
3- experimentation and testing
4-knowledge of design basics
5-understand common penetration methods
6- handle photos
7-he web developer should learn to organize the time and good planning of the project
8-patience is another feature that a web developer should enjoy.
How long a web developer work per week?
COBO: most work at least 40 hours a week, but due to the project-oriented nature of the work, they may also have to work evenings and weekends to meet deadlines or solve unexpected technical probl

what is the Functional tasks of a web developer?COBO: I am sorry! I don't understand you
I said
what is the Functional tasks of a web developer?COBO: the functional tasks of a web developer 1-prepare programming plans and maps
2-type the code required to run the site as required
3-test the work of symbols at each stage of development
4-detect and find solutions to code errors
5-following up on modern technologies and trying to apply them in work projects
6-follow up the performance of the site after its launch and make adjustments
```

Figure 3: 3rd result of testing chatbot

4. Conclusion and future works

The developed chatbot showed a good result as it responds to the user input correctly. Even though the chatbot didn't give a satisfied answer for some questions, it answer other questions in a good answer. We aim to improve it by add more paths and make it more intelligent by using other intelligent algorithms.

5. Bibliography

- [1]. wikipedia, "Chatbot " Accessed 8 November 2019,
<https://en.wikipedia.org/wiki/Chatbot>
- [2]. Sharpen. August 23, 2016. The Anatomy of a Chatbot .online(E-book).
- [3]. Sienkiewicz Agnieszka. Oct 29, 2019 "What Is a Chatbot? 11 Benefits Of Chatbots That Will Help Your Business Grow" Accessed 8 November 2019.
<https://www.tidio.com/blog/what-is-a-chatbot/>
- [4]. Kim Larry. Jul 3, 2017 "11 Best Uses of Chatbots Right Now" Accessed 8 November 2019.
<https://medium.com/the-mission/11-best-uses-of-chatbots-right-now-1c27764b7e62>
- [5]. Phillips Casey . May 25, 2018 "Natural Language Processing (NLP) & Why Chatbots Need it" Accessed 8 November 2019.
<https://chatbotsmagazine.com/natural-language-processing-nlp-why-chatbots-need-it-a9d98f30ab13>
- [6]. Navlani Avinash . September 4, 2018 "Text Analytics for Beginners using NLTK"
Accessed 8 November 2019.
<https://www.datacamp.com/community/tutorials/text-analytics-beginners-nltk>
- [7]. Techopedia . "Natural Language Toolkit (NLTK)" Accessed 8 November 2019.
<https://www.techopedia.com/definition/30343/natural-language-toolkit-nltk>
- [8]. F. Ansari, Sh. Saad and Sh. Shareem, "Career Counselling Chatbot" International Research Journal of Engineering and Technology, vol. 6, 2019.
- [9].. B. S.Bani, A. P.Singh, " College Enquiry Chatbot Using A.L.I.C.E," International Journal of New Technology and Research, vol. 3, 2017.
- [10]. Pandey Parul . "Building a Simple Chatbot from Scratch in Python (using NLTK) " Accessed 14 September 2019.
<https://medium.com/analytics-vidhya/building-a-simple-chatbot-in-python-using-nltk-7c8c8215ac6e>