جامعة دمياط
كلية الحاسبات والذكاء الاصطناعى

Chatbot voice input

Our project is a chatbot web site which is responsive and can add to your knowledge in different fields like AI , Neural Networks Introduction , etc, using flask for backend and HTML , CSS , JavaScript for front

Team Members

Ahmed Abdelfattah Romia	CS	800167475
Abdelsalam abu elwafa	CS	800167536
Mohamed Mohamed elesawy	CS	800167581

1. Introduction

the research area of developing a chatbot with voice feature using NLP and NN holds great significance for both industry practice and knowledge advancement. It opens up new avenues for creating more interactive and effective conversational agents, benefiting businesses and users alike.

2. Problem Statement

this research project aims to address the specific problem of developing a chatbot with voice feature using NLP and NN, and it focuses on investigating constraint management in construction projects and integrating that knowledge into the chatbot's functionality

3. Dataset

Our datasets are provided from the python library NLTK and they are as following:

```
nltk.download('treebank')
nltk.download('brown')
nltk.download('conll2000')
nltk.download('universal_tagset')
```

4. Methodology

For NLP:

We Used TFIDF Technique for slicing out BOW (Bag of Words) in a text file and finding the most alike answer and wanted by the end-user by the help of some libraries provided by NLTK like:

-TfidfVectorizer

-CosineSimilarity

For NN:

We Used Bidirectional LSTM technique with number of epochs reaches 6 epochs and batch size 256.

Training ratio:0.75 , testing ratio:0.1 , validation ratio:0.15.

This technique is considered as RNN technique (Recurrent Neural Network)

Our Neural Network technique finds the tags of the words in order to determine if the sentence entered by the end-user is going to be saved in the dataset or not for next processes.

For example:

If it was an information like:

“Elephants have four legs” this sentence is going to be saved in my dataset.

Else if it was a sentence which contains a question or an invalid information like:

“What do you know about football ?“ this is not going to be saved in the dataset

5. Results

The accuracy of the model reaches 90% after only 6 epochs and loss 30%

Loss can decrease after more epochs.

After 20 epochs loss percent reaches 10%.

6. Reference List

We used datasets provided by NLTK (Universal Tags dataset)

We read the documentation of keras library as a reference in order to use it’s built-in functions

Our Code Link on GitHub : https://github.com/ahmedromia-oss/ChatBot