AMITY UNIVERSITY

-----------------UTTAR PRADESH----------------

Project Report ON
Wikipedia Article Extraction Using Voice Commands

Submitted to
Amity Institute of Information Technology
In partial fulfilment of the requirements for the award of
The degree.

Bachelor of Computer Applications

Submitted to
Dr. Jitender Tanwar

Submitted by
Aryan Kaushik
Enrollment No. A1004821027

TABLE OF CONTENTS

Declaration

I Aryan Kaushik of BCA hereby declare that the term paper titled " Wikipedia Article Extraction Using Voice Commands" which is submitted by me to Dr. Jitender Tanwar , Amity Institute of Information Technology, Amity University Uttar Pradesh, Noida, in partial fulfilment of the requirement of the degree of Bachelor of Computer Application, has not been previously formed the basics for the award of any degree, diploma or other similar title or recognition the author attests that permission has been obtained for the use of any copyrighted material appearing in the Dissertation/Project report other than brief excerpts requiring only proper acknowledgement in scholarly writing and all such use is acknowledged.

Name of student=Aryan Kaushik
BCA-2A
A1004821027

TO WHOM IT MAY CONCERN

This is to certify that Mr. /Ms. ARYAN KAUSHIK, Enrollment No. A1004821027 of course is doing Term Paper at Amity Institute of Information Technology, Noida, AUUP from 20th June 2022 to 17th July, 2022 under my supervision.

Jitender Tanwal

Faculty Signature
Faculty Name

ACKNOWLEDGEMENT

It is a high privilege for me to express my deep gratitude to those entire faculty members who helped me complete the project, especially my internal guide Dr **Jitender Tanwar** who was always there in my hour of need. My special thanks to all other faculty members, Batch mates & Senior of Amity Institute of Information technology, Amity University, Uttar Pradesh for helping me with the completion of the project work and its report submission.

Name= Aryan Kaushik
Enrolment .no=A1004821027

Wikipedia Article Extraction Using Voice Commands

Abstract :-

This project uses python as primary language. This project will aim to create an application which will enable person using it to extract information from Wikipedia, and the application will read important points from it to the user and the whole article will be displayed. Moreover this project can be expanded into many sectors and it can be more than just a Wikipedia article extractor.

Introduction:-

This project is made using the python language which is best for web scraping projects and to reduce the complexity of the code. This project will make the process of searching an article from Wikipedia easier by giving you the option to extract the article from Wikipedia by just a voice command.

Different Creators have made different modules to help the programmer to create projects with ease. These modules are used in this project to make it with ease, these modules were downloaded externally using the PIP command provided by python developers.

Different functions are created in this project and each function performs different operations, these operations are created for the smooth functioning of the project.
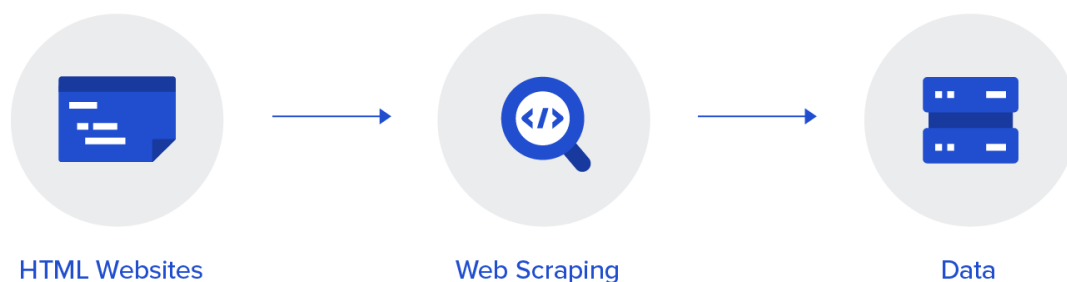
How Wikipedia works and What is web scraping:-

1. **Wikipedia Working :-** Wikipedia works as an encyclopaedia which is widely accessible to people who have access to internet. Wikipedia is an online platform which provides different articles on various topics. These articles are written and edited collaboratively. Anonymous volunteers from all around the world help write and edit the Wikipedia articles. There are some sections of Wikipedia which are restricted to edit to prevent disruption or vandalism. Anyone is allowed to add or edit words, references, images, and other media here. The content must not be copyright restrictions and contentious material about living people. It must conform with Wikipedia's policies, including being verifiable against a published reliable source. If someone damages the contents of the Wikipedia, then their software is built to change it easily. The changing of the article is just one EDIT button away.

2. **Web Scraping –** Assume you read some information or article on internet, and you want to use the article or the information provided on the website in your project. That is when web scraping comes into play, it is a technique to extract the data from a particular website. This data can be in any form text, tabular format, numerical data, etc. This can

also extract any particular form of data from a website. This technique is basically used to extract large amounts of data from a website. It has many uses for example:- Train a Machine Learning Algorithm, Stock Market Analysis, etc. Web scraping is an automatic system to gain large quantities of data from websites. utmost of this data is unshaped data in an HTML format which is also converted into structured data in a spreadsheet or a database so that it can be used in colourful operations. There are numerous different ways to perform web scraping to gain data from websites. These include using online services, particular API's or indeed creating your law for web scraping from scrape. numerous large websites, like Google, Twitter, etc.

HTML Websites → Web Scraping → Data

The Modules and how they work:-

Before getting to know all the modules we must know which Language will I be using. I chose Python as a programming language as web scraping is easy and it provides various modules to make my code easy.
So the list of all the modules required and used in the project are:-

1) **Datetime Module:-** In Python, date and time are not a data type of their own, but a module named datetime can be imported to work with the date as well as time. Python Datetime module comes built into Python, so there is no need to install it externally. It contains 6 main categories :-

   1. Date – This displays current Date, it follows Gregorian calendar

   2. Time – This displays current Time according to the 24hr clock format.

   3. Datetime – This displays the current date and time.

   4. Timedelta - A duration expressing the difference between two date, time, or datetime instances to microsecond resolution.

2) **Speech Recognition:-** Speech Recognition is an important feature in several applications used such as home automation, artificial intelligence, etc. This article aims to provide an introduction on how to make use of the Speech Recognition library of Python. This is useful as it can be used on microcontrollers such as Raspberry Pi's with the help of an external microphone.

3) **Pyttsx3:-** It is a module which is used to convert the speech to text. This library can also work offline while other libraries require the user to be online. For every operating system there

are different voices. Like Windows- "SAPI5", Mac- "nsss", Linux- "espeak".

4) **Wikipedia Module:-** In this article, we will see how to use Python's Wikipedia module to fetch a variety of information from the Wikipedia website. This basically uses web scraping to extract the article from the website of Wikipedia. This returns the result in the format of a string. Ex:-

> **Syntax :** *wikipedia.summary(title, sentences)*
>
> **Argument :**
>
> *Title of the topic*
>
> *Optional argument: setting number of lines in result.*
>
> **Return :** *Returns the summary in string format.*

Python code of the project:-

wikipedia_voice.py ✕

wikipedia_voice.py > ...

```python
from datetime import datetime
import speech_recognition as sr
import pyttsx3
import wikipedia
engine = pyttsx3.init('nsss')
voices = engine.getProperty('voices')
engine.setProperty('voices', voices[1].id)



def speak(audio):
    """
    This Funcation speaks every thing that is given as an argument. You can call this funtion manually
    or use a loop as per the situation
    """

    engine.say(audio)
    engine.runAndWait()


def wishme():
    """
    This function wishes you as per the time of the day, For ex:- If it is Afternoon which is between
    12pm to 5pm then this program will wish you 'Good Afternoon'. This function uses the module name
    Datetime to extract the time.
    """


    hour = int(datetime.now().hour)
```

wikipedia_voice.py ✕

wikipedia_voice.py > ...

```python
        if hour>=0 and hour<12:
            speak("Good Morning")
        elif hour>=12 and hour<17:
            speak("Good Afternoon")
        else:
            speak("Good Evening")


        speak("I'm wikipedia article extractor, tell me which article to fetch")



def takecommand():
    """
    This function takes voice input and converts it into text, Then that text is used to extract the
    article from Wikipedia. This uses google speech to text program.
    """

    r = sr.Recognizer()

    with sr.Microphone() as source:
        print("Listening....")
        r.pause_threshold = 1
        audio = r.listen(source)

    try:
        print("Recognizing....")
        query = r.recognize_google(audio, language='en-in')
        print(f"User said: {query}\n")
```
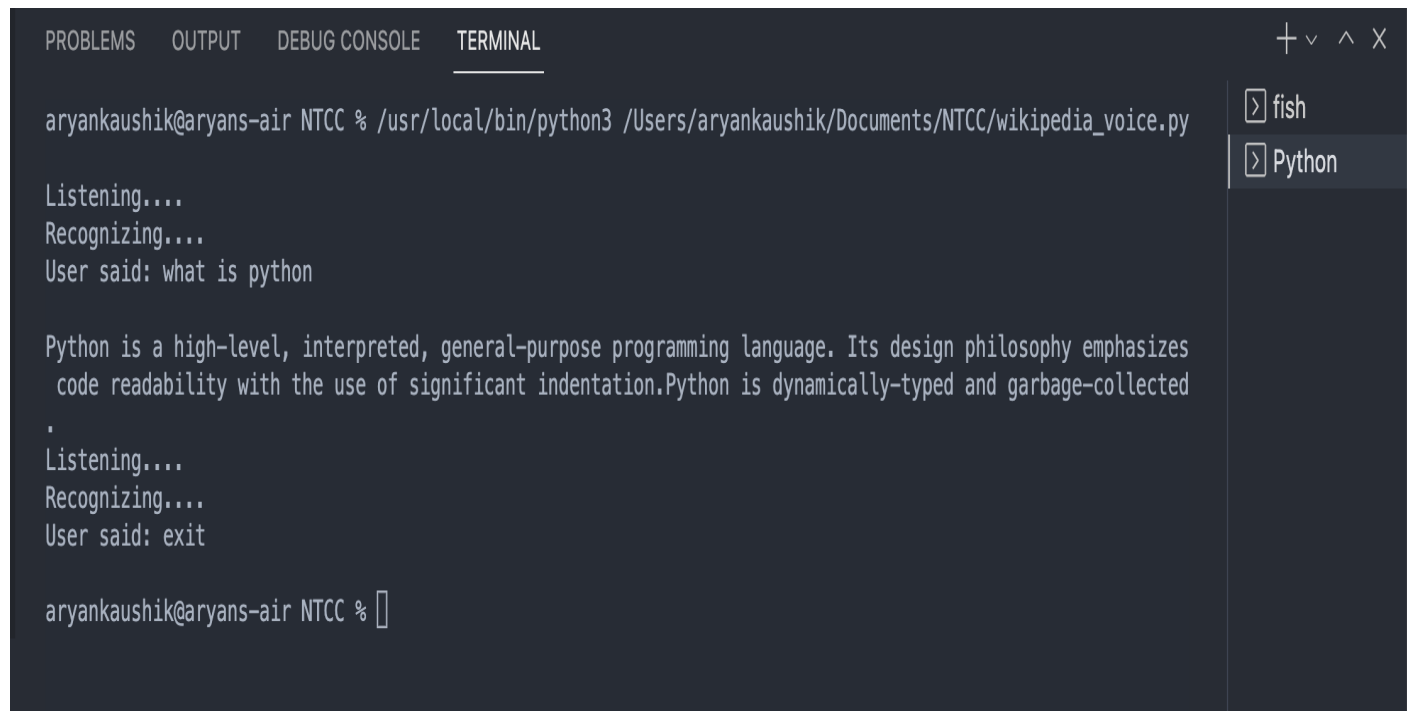
```python
        except Exception:
            print("Can you say that again please")
            return "None"


    return query

if __name__ == '__main__':
    wishme()

    """
    This is the main funtion which controls which function to call. In this an infinite loop is
    executed which continues to extract the articles as per your convinience. You can exit this loop
    by saying Exit, and it will exit with a quote of "Have a good day".
    """
    while True:

        query = takecommand().lower()

        if query == 'exit':
            speak("Have a good day!")
            break

        speak("Searching Wikipedia...")
        results = wikipedia.summary(query, sentences = 2)
        speak("According to wikipedia")
        print(results)
        speak(results)
```

Output of the Code:-

1.


```
PROBLEMS    OUTPUT    DEBUG CONSOLE    TERMINAL                                            + ∨ ∧ X

aryankaushik@aryans-air NTCC % /usr/local/bin/python3 /Users/aryankaushik/Documents/NTCC/wikipedia_voice.py    ⟩ fish
                                                                                                               ⟩ Python
Listening....
Recognizing....
User said: what is python

Python is a high-level, interpreted, general-purpose programming language. Its design philosophy emphasizes
 code readability with the use of significant indentation.Python is dynamically-typed and garbage-collected
.
Listening....
Recognizing....
User said: exit

aryankaushik@aryans-air NTCC % ▯
```
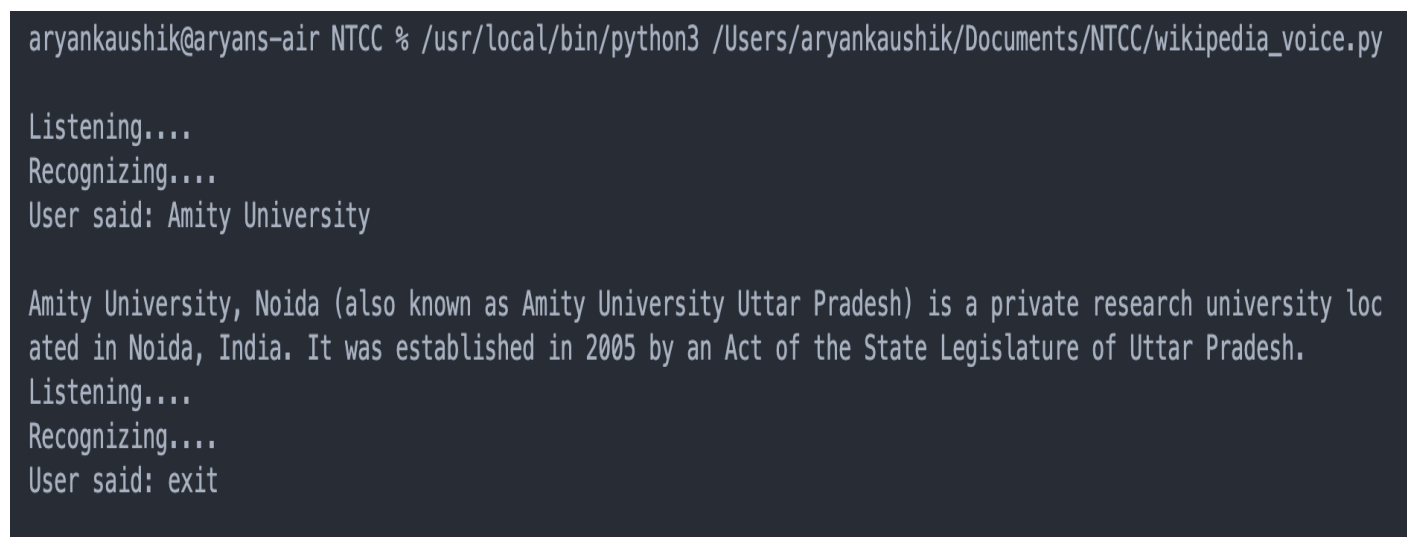
2.


```
aryankaushik@aryans-air NTCC % /usr/local/bin/python3 /Users/aryankaushik/Documents/NTCC/wikipedia_voice.py

Listening....
Recognizing....
User said: Amity University

Amity University, Noida (also known as Amity University Uttar Pradesh) is a private research university loc
ated in Noida, India. It was established in 2005 by an Act of the State Legislature of Uttar Pradesh.
Listening....
Recognizing....
User said: exit
```

3.

```
aryankaushik@aryans-air NTCC % /usr/local/bin/python3 /Users/aryankaushik/Documents/NTCC/wikipedia_voice.py

Listening....
Recognizing....
User said: what is metaverse

In futurism and science fiction, the metaverse is a hypothetical iteration of the Internet as a single, uni
versal and immersive virtual world that is facilitated by the use of virtual reality (VR) and augmented rea
lity (AR) headsets. In colloquial use, a metaverse is a network of 3D virtual worlds focused on social conn
ection.The term "metaverse" originated in the 1992 science fiction novel Snow Crash as a portmanteau of "me
ta" and "universe." Metaverse development is often linked to advancing virtual reality technology due to in
creasing demands for immersion.
Listening....
```

Conclusion:-

Conclusion of this project is that I learned about how advance Python language can be. A project to extract Wikipedia by voice commands was created, and it contains very minimum errors. I used different examples and phrases to check the program if it is working properly. At the end this was a successful project.