INTERNSHIP PROJECT REPORT

NEWS READER - SELENIUM

Name of Project: News reader – Selenium project

Name of Intern: Meesaragandla Srikar

Intern's Ph. no.: 9948697191

Intern's Add.: 9-2/2 Near Parasarabharthi school, Kanigiri - 523230, Prakasam (dt) Andhra Pradesh

Name of College: SAVEETHA SCHOOL OF ENGINEERING

College's Add.: Saveetha Nagar, Thandalam, Chennai - 602105

Branch: Computer Science And Engineering

Project Submission Date: 29/04/2021

NEWS READER – SELENIUM PROJECT

The News Reader project main aim is to provide the simple way of listening news instead of reading it in your busy day schedule. The research activities concluded within this project is strongly rely on different types events like some which focuses on News detection, processing , storage and reading the different format news. The overall approach for processing the data follows a sequence of steps as shown below in the flowchart figure(Fig2). News Reader uses an open source and updated libraries, packages to detect the news from the different type of webpages and convert the text to speech format to and able to listen through any music player in your phone or mobile and also through various platforms. For this project we are mainly using following libraries and tools.

Selenium - Selenium is the most popularly used freeware and open source automation tool. The benefits of Selenium for Test Automation are immense. Importantly, it enables record and playback for testing web applications and can run multiple scripts across various browsers.

gTTS - gTTS (Google Text-to-Speech), a Python library and CLI tool to interface with Google Translates text-to-speech API. Writes spoken mp3 data to a file, a file-like object (byte string) for further audio manipulation, or stdout. It features flexible pre-processing and tokenizing.
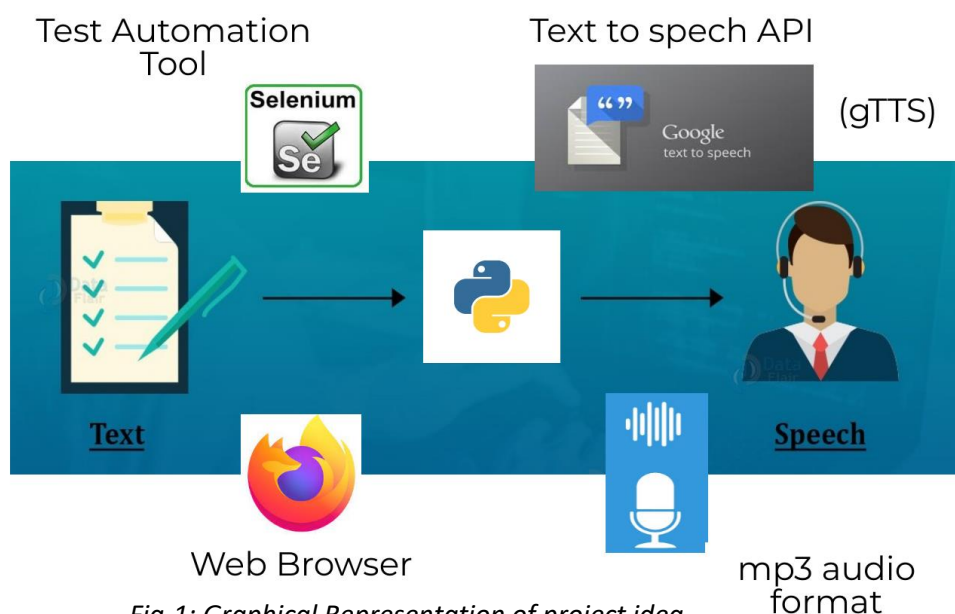

Fig-1: Graphical Representation of project idea

List of Software:

- Selenium : Selenium Libraries

- Google Text To Speech (gTTS) – Converting Text file to Speech (mp3 audio)

- OS module – To play the audio file

- Notepad – To save the news in a text file

- Juypter Notebook – To write the code for the project

- Firefox – web browser used

List of Hardware:

- Laptop

- Keyboard and mouse

START

Installing & importing required packages (Selenium, gTTS)

using webdriver Function

get() function to pass the website link

Using find_elements() method to find the certain class in the page

Open file in appending mode

copying the text using write() mode

save as text file

Run text to speech

Save as audio file(mp3)
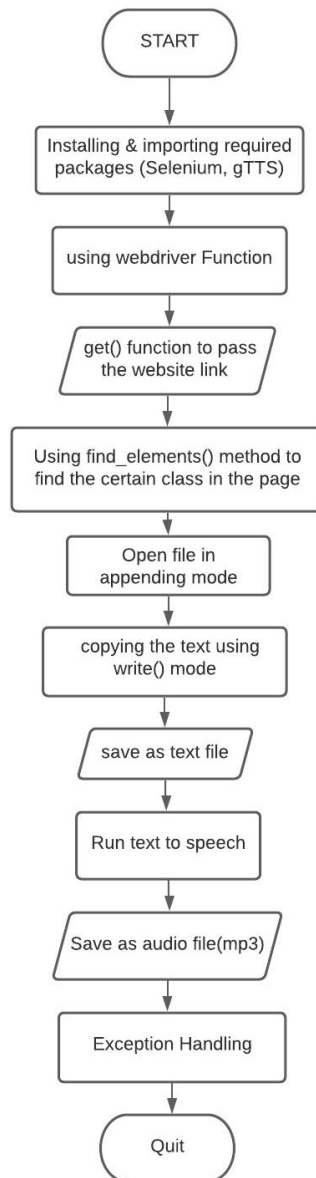
Exception Handling

Quit

Fig-2: Flowchart

- This project is useful to listen news instead of reading, by converting text to audio format.
- This program needs internet connectivity to get news from the desired website.
- User can access various websites and listen the news or stories, audio books etc.

Purpose:

The main goal of proposed system is to view the News from various News websites, and convert the text to audio file, One can view News feed of any other website using this News Reader – selenium project.

Installing required libraries and packages

- !pip install selenium - Run in a juypter notebook

```
In [1]: !pip install selenium    #command to install selenium libr

        Collecting selenium
          Downloading selenium-3.141.0-py2.py3-none-any.whl (904 kB)
        Requirement already satisfied: urllib3 in c:\users\srika\anaconda3\lib\site-packages (from se
        lenium) (1.25.11)
        Installing collected packages: selenium
        Successfully installed selenium-3.141.0
```

- !pip install gTTS

```
In [9]: !pip install gTTS      #Installing another required package google Text To speech

        Collecting gTTS
          Downloading gTTS-2.2.2-py3-none-any.whl (25 kB)
        Requirement already satisfied: requests in c:\users\srika\anaconda3\lib\site-packages (from g
        TTS) (2.24.0)
        Requirement already satisfied: click in c:\users\srika\anaconda3\lib\site-packages (from gTT
        S) (7.1.2)
        Requirement already satisfied: six in c:\users\srika\anaconda3\lib\site-packages (from gTTS)
        (1.15.0)
        Requirement already satisfied: idna<3,>=2.5 in c:\users\srika\anaconda3\lib\site-packages (fr
        om requests->gTTS) (2.10)
        Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in c:\users\srika\anac
        onda3\lib\site-packages (from requests->gTTS) (1.25.11)
        Requirement already satisfied: chardet<4,>=3.0.2 in c:\users\srika\anaconda3\lib\site-package
        s (from requests->gTTS) (3.0.4)
        Requirement already satisfied: certifi>=2017.4.17 in c:\users\srika\anaconda3\lib\site-packag
        es (from requests->gTTS) (2020.6.20)
        Installing collected packages: gTTS
        Successfully installed gTTS-2.2.2
```

After installing the required libraries and packages, start to write code as follow, the gTTS is google Text To Speech API uses internet, we can also use python text to speech library called pyttsx3 and also available in offline mode.

```
from selenium import webdriver

from gtts import gTTS

import os

from sys import exit

def main():

    driver = webdriver.Firefox()    #Command to access Firefox web browser

    driver.get('https://inshorts.com/en/read/')        #get() function to
pass the link of the website

    print(driver.title)

    driver.maximize_window()


    for headline in driver.find_elements_by_class_name("card-stack"):
#using Locating Elements method to find the class in the website

        test = headline.text

        f = open("news.txt",'a',encoding='utf-8')  #Encoding UTF-8
characters to text format

        f.write(test + "\n")                    #Writing the text in the file

        f.close()

        print(test)

        driver.quit()              #closing the driver using quit()command



    f = open("news.txt",'r',encoding='utf-8')        #Opening the file in
read mode

    a = f.read()

    tts=gTTS(text=a,lang='en')                #Converting text to audio format

    tts.save('readnews.mp3')            #Saving the file in mp3 audio format

    f.close()

    exit()


    ## Exception handling
if __name__ == "__main__":

    try:
```
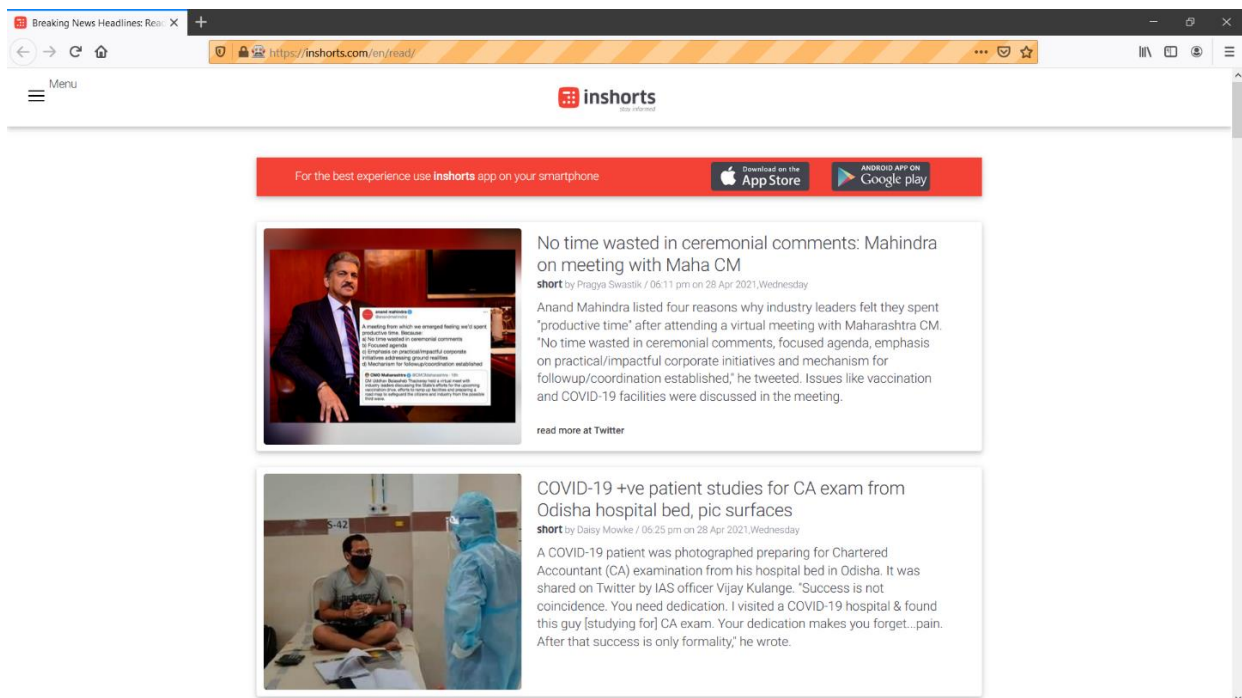
```
    main()
except KeyboardInterrupt:
    # do nothing here
    pass
```
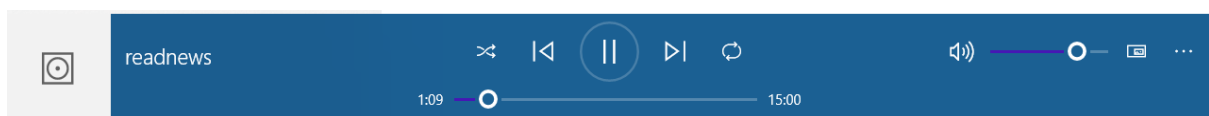
Output:

1. **The website called using get() opened in web browser Firefox:**

2. **The saved audio file named readnews.mp3 is playing in default music player:**

3. *The text file and audio file we saved will be saved in the selected path of the whole project file, the path need to be given if the audio file is not in the saved in the same folder.*

- www.google.com

- https://selenium-python.readthedocs.io/installation.html

- https://github.com/mozilla/geckodriver/releases

- https://gtts.readthedocs.io/en/latest/

- https://pypi.org/project/playsound/

- https://docs.python.org/3/library/

- https://jupyter.org/

- https://docs.python.org/3/library/exceptions.html