Image_Text-Speech Synthesizer

Vamshi Jaligama , Saikrishna Reddy Kotha , Pavan Kumar Ponugoti

I. Abstract

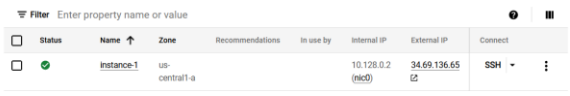
This document gives the detailed walk-through of the Image_Text-Speech Synthesizer project with the help of Google Cloud Platform

II. Introduction

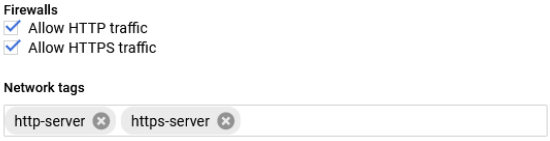
We have implemented an image to text converter and then text to speech using the Cloud Vision API and Text to speech API in VM instances with the help of the google cloud platform .

III . Detailed Walkthrough

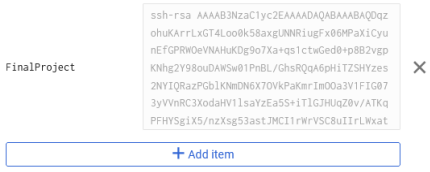
Firstly, we created a Project and then go to compute engine and create a instance with ubuntu OS.



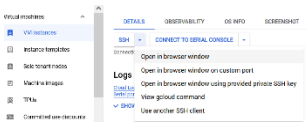
Then open the instance and click on edit and then we need to allow the https / https traffic in the instance firewall section



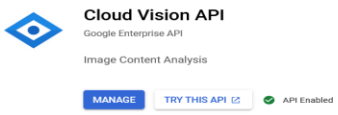
Next, we need to create a public key and then copy that public key into the VM instance by click on add SSH keys then copy and paste the public key into the box and click on save



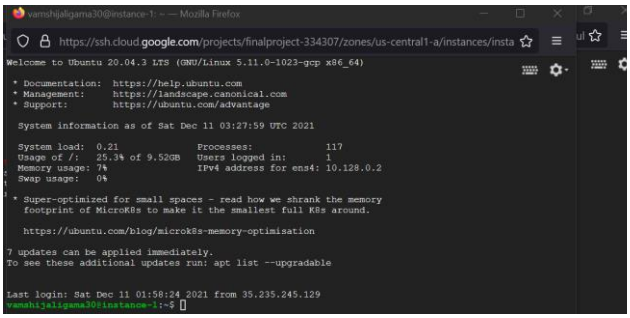
After that open the SSH shell that is available in the VM instances in a new window



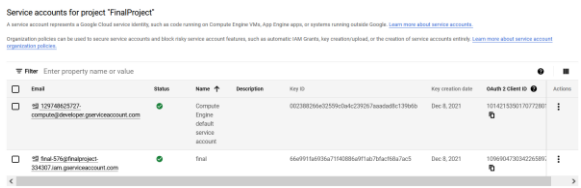
Enable the Cloud vision api for the Final project

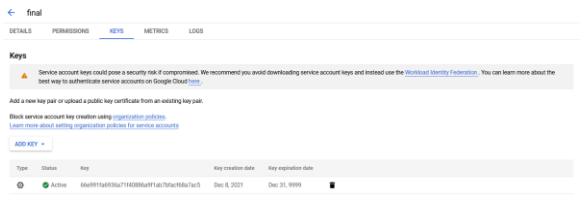


Which will connect to a ssh server and opens a terminal for the further development.

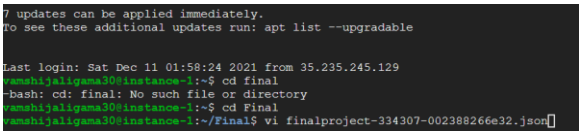


Then go to service accounts and create a new service account and then go to the key section and create a new key and download the JSON file.



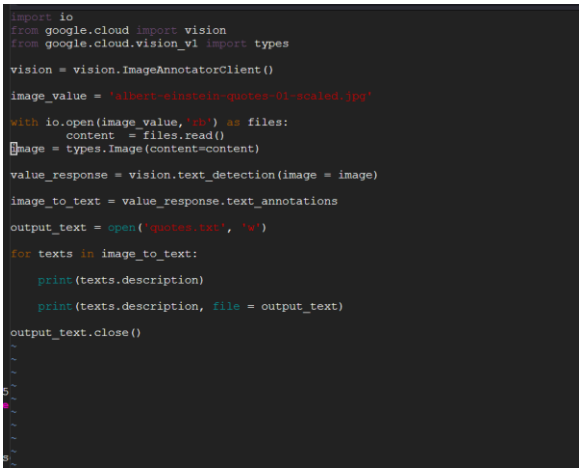


After that create a new directory in the Shell and then save the json file in that directory



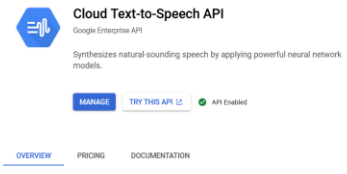
Json file :

Now we have the python file for the convertion of the image to text with the help of the cloud vision API :



```
import io
from google.cloud import vision
from google.cloud.vision_v1 import types

vision = vision.ImageAnnotatorClient()

image_value = 'albert-einstein-quotes-01-scaled.jpg'

with io.open(image_value,'rb') as files:
        content  = files.read()
image = types.Image(content=content)

value_response = vision.text_detection(image = image)

image_to_text = value_response.text_annotations

output_text = open('quotes.txt', 'w')

for texts in image_to_text:

    print(texts.description)

    print(texts.description, file = output_text)

output_text.close()
```

Next we need to enable text to speech api for our project :

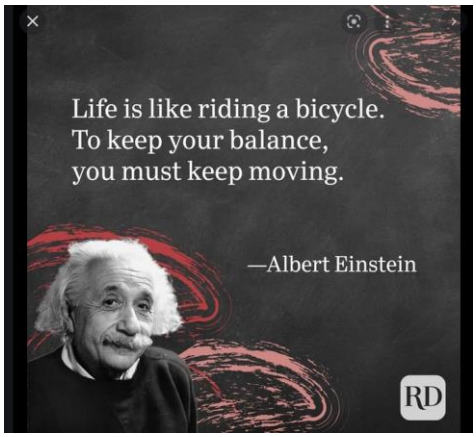


Here is python code for the conversion of the text into the speech:

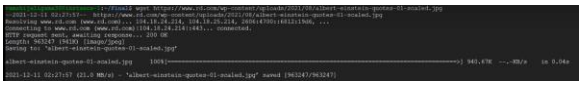
```
from google.cloud import texttospeech
import os
text_to_speech = texttospeech.TextToSpeechClient()

image_text = 'quotes.txt'

with open(image_text, 'r') as files:
    text = files.read()
    value = texttospeech.types.SynthesisInput(text=text)

speech= texttospeech.types.VoiceSelectionParams(language_code='en-US',
                                                ssml_gender=texttospeech.enums.SsmlVoiceGender.MALE)
configuration = texttospeech.types.AudioConfig(audio_encoding=texttospeech.enums.AudioEncoding.MP3)
result  = text_to_speech.synthesize_speech(value, speech , configuration)

with open('file_output_speech1.mp3', 'wb') as final_speech:
    final_speech.write(result.audio_content)
```

Then we need to search for a image of our choice after that we need to copy that image link and go to the shell there, by using the wget command and download it to the instance.

Google search image



Wget command for download:



After downloading the image file then we need to look for the name of the image file and then copy that file name in the converter python code.

```
vision = vision.ImageAnnotatorClient()

image_value = 'albert-einstein-quotes-01-scaled.jpg'
```

Next we need to run the python code for converting the image to text :

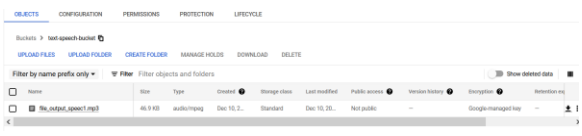
```
vamshijaligama30@instance-1:~/Final$ python3 converter2.py
Life is like riding a bicycle.
To keep your balance,
you must keep moving.
-Albert Einstein
RD

Life
is
like
import io
riding
a
bicycle.
To
keep
your
balance,
you
must
keep
moving.
-Albert
Einstein
```

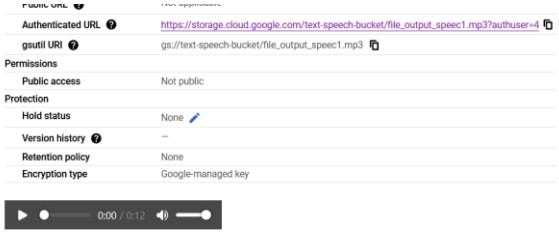
After running the text to speech python code a new mp3 file is created in the instance we can see that a new file created called file_output_speech1.mp3

Finally we need to upload this mp3 into the bucket so that we can access the mp3 file in gcp using the command “gsutil cp file_output_speec1.mp3 gs://text-speech-bucket”

```
vamshijaligama30@instance-1:~/Final$ gsutil cp finalproject-334307-fdd01833fe95.json gs://text-speech-bucket
Copying file://finalproject-334307-fdd01833fe95.json [Content-Type=application/json]...
/ [1 files][  2.3 KiB/  2.3 KiB]
Operation completed over 1 objects/2.3 KiB.
```

We can see that a mp3 is uploaded in the bucket





References

[1] GCP, "Cloud Vision Quick start," [Online]. Available: https://cloud.google.com/vision/docs/samples.

[2] GCP, "Cloud Text to Speech Quickstart," [Online]. Available: https://cloud.google.com/text-to-speech/docs/quickstart-protocol?hl=en_US.