Beautifulsoup

**Disini saya download dulu web nya dan dijadikan sebuah file karena kalau di scrap secara online websitenya diblokir oleh SSL Telkom**

from bs4 import BeautifulSoup as bs

from requests.api import get

import sys

import os

sys.stdout = open("link.txt", "w")

folder = "C:\\Users\\Moch Faizal Nazili\\Documents\\GitHub\\Tugas-Besar-Bs4-1914311022-Pagi-Ubhara-Surabaya"

for filename in os.listdir(folder):

    if filename.endswith(".html"):

        fname = os.path.join(folder, filename)

        # print("Filename: {}".format(fname))

        with open(fname, "r") as f:

            soup = bs(f.read(), "html.parser")

            info = soup.find\_all("a", href=True)

        for i in info:

            new = i.get("href")

            if "https://www.wavsource.com/snds\_2020-10-01" in new:

                print(new)

sys.stdout.close()

**setelah itu linknya saya masukkan ke file link.txt lalu saya download menggunakan powershell**

cd D:\Tugas-Besar-Bs4-1914311022-Pagi-Ubhara-Surabaya\soundfile

$list = Get-Content D:\Tugas-Besar-Bs4-1914311022-Pagi-Ubhara-Surabaya\link.txt

$clnt = New-Object System.Net.WebClient

foreach($url in $list)

{

    #Get the filename

    $filename = [System.IO.Path]::GetFileName($url)

    #Create the output path

    $file = [System.IO.Path]::Combine($pwd.Path, $filename)

    Write-Host -NoNewline "Mengunduh ""$url""... "

    #Download the file using the WebClient

    $clnt.DownloadFile($url, $file)

    Write-Host "Selesai!."

}

**Setelah semua file terdownload. Saya recognition menggunakan IBM Cloud dan hasilnya saya masukkan ke file txt**

pip install ibm\_watson

from ibm\_watson import SpeechToTextV1

from ibm\_watson.websocket import RecognizeCallback, AudioSource

from ibm\_cloud\_sdk\_core.authenticators import IAMAuthenticator

import pandas as pd

import sys

import os

apikey = 'dPG1OcuucjbYNK3CFDEqX\_7YKJNPo3nvFiprPfszsjHP'

url = 'https://api.us-south.speech-to-text.watson.cloud.ibm.com/instances/06599e43-94ba-4e51-8dd8-679d3fd01c83'

authenticator = IAMAuthenticator(apikey)

stt = SpeechToTextV1(authenticator=authenticator)

stt.set\_service\_url(url)

for i in os.listdir("/content/Tugas-Besar-Bs4-1914311022-Pagi-Ubhara-Surabaya/soundfile/"):

  path = "/content/Tugas-Besar-Bs4-1914311022-Pagi-Ubhara-Surabaya/soundfile/" + i

  with open(path, 'rb') as file:

    try:

      print("Sedang mengenali audio")

      print("-"\*100)

      res = stt.recognize(audio=file, content\_type='audio/wav', model='en-US\_NarrowbandModel', continuous=True).get\_result()

      text = res['results'][0]['alternatives'][0]['transcript']

      # confidence = res['results'][0]['alternatives'][0]['confidence']

      nama\_file = file.name

      print("Nama File        : "+ file.name)

      print("Hasil pengenalan : "+ text)

      # print("Tingkat Confindence  : "+confidence)

      print(" "\*100)

    except Exception as ex:

      print(ex)