[temples/googlengine.py at main · NATURE-LABS/temples (github.com)](https://github.com/NATURE-LABS/temples/blob/main/googlengine.py)

|  |
| --- |
| """  Purpose : Google Connect is a utility to connect to the Google Search Engine.  Generate the csv data based on the search query given in the argument.  The demonstrated program generate the datasheet in a format in CSV.  Design and developed by :  Kyndryl Solutions Private Limited  Project Team : Google Cloud Platform - Guild.  Lab : Nature Labs @ GCP  How to use  ------------  python googledata.py <https://www.googlengine.com/tnt>  Contact  --------  Kyndryl GCP Guild Moderator : Ramamurthy V  Email : ramamurthy.valavandan@kyndryl.com  GCP Contact : gcpguild@gmail.com  Date : June 8 2022.  Contributors : 42 key members from GCP Guild.  """  import requests, re, csv, sys  import pandas as pd  from bs4 import BeautifulSoup  import bs4 as bs  import urllib.request  import numpy as np  basepath = "C:"  codepath="python"  function = "temples"  N="\\"  namefile = "googledata"  def prt(p):  width = len(p) + 4  print('┏' + "━"\*width + "┓")  print('┃' + p.center(width) + '┃')  print('┗' + "━"\*width + "┛")  arglen = len(sys.argv)  exampleurl = "https://www.google.com/search?q='popular hindu temples tamil nadu'"  if (arglen != 2):  pi="\'Input website is not given .. ! \' :"  p = ("{} {}".format(pi,exampleurl))  prt(p)  pi="python "  p = ("{} {}".format(pi,exampleurl))  prt(p)  exit(1)  searchpg = str(sys.argv[1])  gourl = re.sub(r'^.+/([^/]+)$', r'\1', searchpg)  csvfile = ("{}{}{}{}".format(namefile,"\_",gourl,".csv"))  genfile = ("{}{}{}{}{}{}{}{}{}".format(basepath,N,codepath,N,function,N,"data",N,csvfile))  website = requests.get(searchpg)  soup = BeautifulSoup(website.content, 'html.parser')  if (soup.title is not None):  title = soup.title.string  bese= soup.prettify()  #print(bese)  lt = "&lt;"  gt = "&gt;"  bese = re.sub(lt, '<', bese)  bese = re.sub(gt, '>', bese)  patent = ("{}{}{}".format('<h3>', "([^$]\*)", '</h3>'))  bese = re.findall(patent, str(bese))  bese = re.sub('\,', '|', str(bese))  bese = re.sub('<h3>', '', str(bese))  bese = re.sub('</h3>', ',', str(bese))  bese = re.sub('[^A-Za-z|,]+', ' ' , str(bese))  mano = []  mano = [ x.strip() for x in bese.strip('[]').split(',') ]  gencsv = pd.DataFrame(mano)  gencsv.to\_csv(genfile, index=False, header = False, na\_rep='Unknown')  pi="\'Temple Data sheet is generated .. ! \' :"  p = ("{} {}".format(pi,genfile))  prt(p)  pi="\'Now you can validate the data .. ! \' :"  p = ("{} {}".format(pi,genfile))  prt(p) |