Name: Urinov Azizbek 20162672
Class: Internet and Information Communication
Date: 10/12/2019

Face_shop: chromedriver edition

from selenium import webdriver from bs4 import BeautifulSoup import requests from selenium.webdriver.common.desired_capabili ties import DesiredCapabilities import time from selenium.webdriver.common.keys import Keys import datetime as dt import pandas as pd import matplotlib.pyplot as plt import numpy as np import gensim from gensim.models import CoherenceModel from gensim.models.word2vec import Word2Vec from gensim import corpora, models from nltk.corpus import stopwords from nltk.stem.porter import PorterStemmer from nltk.tokenize import RegexpTokenizer from selenium.webdriver.support.ui import WebDriverWait import re import os	Import necessary files like webdriver, time,os, gensim models more important tools. I just downloaded all files to escape bugs and fails later.
from selenium import webdriver driver = webdriver.Chrome('/Users/name/Downloads/ chromedriver') driver.get("http://jolse.com/category/sheet-ma	First I imported webdriver from selenium. Secondly, I showed a PATH of chromedriver to my computer which i downloaded chromedriver website. Then, I gave a link to a website which i want

sks/1027/") products = driver.find_elements_by_xpath('//p[@class="name"]') prices = driver.find_elements_by_xpath('//li[@item-title=""]') prices_original = driver.find_elements_by_xpath('//li[@item-title="Price"]') items = len(products) for i in range(items): print("Product Name: " + products[i].text + " // Discount Price: " + prices[i].text + " // Orgiginal Price: " + prices_original[i].text) driver.close()	to crawl and show it on the google chrome window. Then, I inserted XPATHs of products, prices and price_originals. Lastly,I printed results as a text by using “for” And close driver.

from selenium import webdriver
 	
max_num = 11
max_dig = 1
driver = webdriver.Chrome('/Users/name/Downloads/chromedriver')
 	
with open('results.csv', 'w') as f:
  f.write("Product Names // , Discount Prices // , Original Prices\n")
 	
for i in range(1, max_num + 1):
  page_num = (max_dig - len(str(i))) * "0" + str(i)
  url = "http://jolse.com/category/sheet-masks/1027/?page=" + page_num
 	
  driver.get(url)
 	
  products = driver.find_elements_by_xpath('//p[@class="name"]')
	I imported selenium driver one more time to avoid bugs. Then I defined number of pages and digits with max_num as well as max_dig.
	
	Then define chromedriver one more time.
	
	Then I opened one csv file to save result and divided it into 3 columns "Product Names // , Discount Prices // , Original Prices\n".
	
	(1, max_num +1) is used to add 1 to link and it goes to the second page and again add 1 and it goes to the 3rd page. It goes like that and finish at 11. "http://jolse.com/category/sheet-masks/1027/?page=1" is an original page add 1 and result ?page=2 which is the second page of the website (+1) (+1)(+1)(+1) like that until 11.
	
	Then define XPATH of products, prices and prices_original.

prices = driver.find_elements_by_xpath('//li[@item-title=""]') prices_original = driver.find_elements_by_xpath('//li[@item-title="Price"]') num_page_items = len(prices) with open('results.csv', 'a') as f: for j in range(num_page_items): f.write(products[j].text + ", " + prices[j].text + ", " + prices_original[j].text +"\n") driver.close()	When computer calculation reach to the last page, it automatically write to the csv file and name it “result”. And close driver.