**1º Nota MBA – BI com Big Data**

**Italo Igor Gomes Nascimento**

**Cláudio Falcão**

**Questão 1.** Implemente um programa que entre no site do UOL e imprima apenas a seguinte mensagem: A cotação atual do dólar é: <cotação>, onde <cotação> vai ser o valor capturado do site no momento. Procure uma forma de omitir as mensagens de log na execução do seu programa para aparecer apenas essa mensagem como saída.

import scrapy

class UOLSpider(scrapy.Spider):

    name='uol\_spider'

    start\_urls=['https://www.uol.com.br/']

    def parse(self,response):

        titulos = response.css(".HU\_currency\_\_quote::text")

        conteudo = titulos[0].get().strip()

        if conteudo != "":

        #yield {'titulo': conteudo}

            print('A cotação do dólar é: ',conteudo)

**Questão 2.** Implemente um programa que receba um produto como parâmetro e liste o nome e o preço de todos esses produtos no mercado livre, com paginação incluída. Busque uma forma de passar um parâmetro para o seu programa.

from selenium import webdriver

from selenium.webdriver.common.by import By

from selenium.common.exceptions import NoSuchElementException

import json

def finds(browser, by, expression):

    try:

        return browser.find\_elements(by, expression)

    except NoSuchElementException as nse:

        return None

def find(browser, by, expression):

    try:

        return browser.find\_element(by, expression)

    except NoSuchElementException as nse:

        return None

a=input("Selecione o item\n")

output  = open('output.json', 'w', encoding='utf-8')

driver = webdriver.Firefox()

driver.get("https://lista.mercadolivre.com.br/" + a)

items = finds(driver, By.CLASS\_NAME, 'ui-search-result\_\_content-wrapper')

for item in items:

    descricao = find(item, By.CLASS\_NAME, 'ui-search-item\_\_title').get\_property("textContent")

    preco = find(item, By.CLASS\_NAME, 'ui-search-price').get\_property("textContent")

    print(descricao)

    print(preco)

    item = {'descricao': descricao, 'preco': preco}

    output.write(json.dumps(item))

    output.write('\n')

print("Tem a quantidade de ", len(descricao), a)

**Questão 3.** Quais cuidados devem ser observados ao capturar dados de um site?

O conteúdo capturado não tem direitos autorais, o scraping não onera os serviços do site, o scraper não viola os termos de uso do site, o scraper não coleta informações sensíveis.

**Questão 4.** Quais ameaças capturas automáticas proporcionam para sistemas web?

Geralmente o processamento é distribuído em uma grande quantidade de máquinas físicas ou virtuais com diversas formas de arquitetura existentes. Por isso, é muito fácil um sistema de Crawling ou Scraping derrubar um servidor, causando negação de serviço DDoS Distributed Denial of Service)

**Questão 5.** Você diria que bots ou crawlers são programas facilmente paralelizáveis? Se sim, explique como isso seria implementado dando um exemplo.

Sim. Daria para criar uma tarefa que pode ser dividida em subtarefas, onde não há necessidade de ordem na execução assim como a troca de informações entre essas subtarefas.