# Ministerul Educației al Republicii Moldova Universitatea Tehnică a Moldovei

# Дисциплина: PD

# Отчёт

# по лабораторной работе №2

тема:

# Автоматизированный сбор данных. Работа со строками.

# Выполнил: учащийся гр. TI-217

1. Virticus

# Проверил: преподаватель по специальности

# Leah

Chișinău, 2023

**Вариант 7.**

С использованием страницы <https://yandex.ru/images/> сформировать запросы для поиска изображений, контент на которых соответствует классам \***zebra**\* и \***bay horse**. Для каждого класса должно быть загружено не менее 1000 изображений. Изображения для каждого класса должны находиться в подпапке папки **dataset**\* с соответсвующим названием.

**Не допускается**:

1. Создание папок вручную. В коде должен быть отражен процесс создания папок и перемещения/загрузки в них файлов.
2. Дублирование изображений для класса.

**Решение:**

import os

import time

from selenium import webdriver

from selenium.webdriver.common.keys import Keys

from selenium.webdriver.common.by import By

from bs4 import BeautifulSoup

from urllib.request import urlretrieve

def DownloadImageFromYandex(query, folder, num\_images):

    os.makedirs(folder, exist\_ok=True)

    driver = webdriver.Chrome()

    driver.get(f"https://yandex.ru/images/search?text={query}")

    time.sleep(1)

    scroll\_count = 0

    while scroll\_count < num\_images // 25:

        element = driver.find\_element(By.TAG\_NAME, 'body')

        element.send\_keys(Keys.END)

        time.sleep(1)

        scroll\_count += 1

        try:

            button = driver.find\_element(By.CLASS\_NAME, 'button2')

            button.click()

        except:

            pass

    html = driver.page\_source

    soup = BeautifulSoup(html, "html.parser")

    thumbnails = soup.find\_all("img", class\_="serp-item\_\_thumb")

    count = 0

    for thumbnail in thumbnails[:num\_images]:

        try:

            full\_image\_url = thumbnail["src"]

            filename = f"{str(count).zfill(4)}.jpg"

            urlretrieve("https:" + full\_image\_url, os.path.join(folder, filename))

            count += 1

            if count >= num\_images:

                break

        except Exception as e:

            print(f"Error : {e}")

    driver.quit()

DownloadImageFromYandex("zebra", "dataset/zebra", 1000)

DownloadImageFromYandex("bay horse", "dataset/bay horse", 1000)

**Результат:**

