Практика 2

- воспользоваться руководством по использованию API Wikimedia (<u>API:Учебное</u> пособие MediaWiki, <u>API:Tutorial MediaWiki</u>)
- изучить API Portal
- написать приложение (консольное или в Jupyter Notebook), которое взаимодействует по API со всеми основными проектами Wikimedia: Википедия, Викидата, Викисловарь, Wikimedia Commons
- основная функция приложения автоматизированный сбор данных, на вход которому подается запрос о личности, объекте, событии, животном и т.п., на выход JSON-структура, содержащая основной контент узла (текст статьи, записи; изображение и т.п.) и связанные с ним узлы (т.е. гиперссылки на другие статьи, записи, изображения и т.п.).
- дополнительная функция получить картинку дня (Picture of the Day) и сохранить ее вместе с описанием (сохранить в текстовый файл рядом с картинкой, назвать так же, как файл картинки)

```
import bs4
import requests
from datetime import date
from hashlib import md5
date_iso = date.today().isoformat()
project_url = "https://en.wikipedia.org" # нужно попробовать разные
проекты, например, Wikimedia Commons и Русская Википедия, где тоже
выкладывают картинки дня
title = "Template:POTD_protected/" + date_iso # тут может измениться
название шаблона, его можно посмотреть, открыв файл с картинкой дня в
редакторе Википедии или просто в браузере
parse_query_params = {
    "action": "parse", "format": "json",
    "formatversion": "2", "prop": "text",
    "page": title + '_(en)'
}
potd_caption = requests.get(API_URL, parse_query_params).json()['parse']
['text']
```

```
potd_soup = bs4.BeautifulSoup(potd_caption, 'html.parser')
description = str(potd_soup.find('div', attrs={'class':
'description'}).text).strip(' .')
image_query_params = {
    "action": "query", "format": "json",
    "formatversion": "2", "prop": "images",
    "titles": title
}
potd_image_data = requests.get(API_URL, image_query_params).json()
filename = str(potd_image_data["query"]["pages"][0]["images"][0]
["title"]).replace(' ', '_') # ссылка вида
https://www.mediawiki.org/wiki/File:Wikimedia_Hackathon_2024_-
_Group_photo,_360_cam.jpg вернется, т.е. на файл в движке Вики
# но! сам файл картинки лежит по адресу
https://upload.wikimedia.org/wikipedia/commons/3/35/Wikimedia_Hackathon_
2024_-_Group_photo%2C_360_cam.jpg
md5_hash = md5(filename[5:].encode('utf-8')).hexdigest()
# /3/35 берутся из вычисления md5-хэша (описано в документации движка)
image_src =
f'https://upload.wikimedia.org/wikipedia/commons/{md5_hash[0]}/{md5_hash
[0:2]}/{filename[5:]}'
```