**Парсер даних по API**

<https://rtfm.co.ua/python-modul-argparse-opcii-komandnoj-stroki-v-primerax/> - argparce

<https://python-scripts.com/logging-python> - logging

Ресурс даних:  
<https://github.com/HackerNews/API>  
  
Приклад отримання списку записів:  
<https://hacker-news.firebaseio.com/v0/topstories.json?print=pretty>

де [topstories](https://hacker-news.firebaseio.com/v0/topstories.json?print=pretty) фільтр по категоріях.

Приклад отримання детальної інформації по ID запису:  
<https://hacker-news.firebaseio.com/v0/item/192327.json?print=pretty>  
де 192327 - ID

список доступних категорій: **askstories, showstories, newstories, jobstories**

**Завдання:**

1. З допомогою **argparse** модуля отримати назву категорії дані по якій будуть парситись.

1.1 У випадку якщо пустий ввід, тобто натиснуто ентер і більш нічого повинні братись параметри за замовчуванням, повинна вказуватись категорія **newstories.**1.2 У випадку якщо вказана неіснуюча категорія повиненно виводитись повідомлення про помилку вводу. Список категорій повинен братись з конфіг файлу

(choices – контейнер допустимых значений для аргумента).

2. Створити лог файл hn\_parser.log з допомогою модуля logging. Повинні логуватись всі дії, наприклад: Прийнято параметри, надіслано запит, отримано запит, створено папку, створено файл… тільки основні дії!!! і декілька загальних прінтів щоб було видно прогрес.

2.1 Файл повинен дописувати в кінець при наступному запуску  
2.2 При першому запуску має створюватись

2.3 Формат логу повинен бути наcтупний:

<2012-08-02 15:37:40,592 - exampleApp - INFO - Program started>

3. Створити папку **results** якщо її немає, в цю папку зберігати файл репорт і лог файл.

4. Створити конфіг файл **config.py** де будуть зберігатись імена файлів, папки(**results**), значення дефолтного параметра(параметра по замовчуванню **newstories**) і філтрів **from\_date, score.**

5. Всі вузькі місця потрібно оброблювати з допомогою Виключень, наприклад якщо надсилається запит то потрібно врахувати що може обірватись конект чи ще щось і відповідь може не прийти чи прийти з помилкою, тоді обробка наступних ітерацій не повинна постраждати.  
  
6.Потрібно спочатку отримати список всіх ID по певник категорії, а вже потім ці ID використати для отримання детальної інформації.

список доступних категорій(повинен міститись в конфіг файлі): **askstories, showstories, newstories, jobstories**

Отримання списку  
url = **'https://hacker-news.firebaseio.com/v0/newstories.json?print=pretty'**

response = urllib2.urlopen(url, timeout=5)

data = json.loads(response.read())

Отримання запису  
url = **'https://hacker-news.firebaseio.com/v0/item/192327.json?print=pretty'**

response = urllib2.urlopen(url, timeout=5)

data = json.loads(response.read())

Значення посилань(url) також потрібно вказати в конфіг файлу.

7. Додати фільтри які будуть регулюватися змінними з конфігу:  
з дати (**from\_date**) тобто в звіті повинні потрапити записи де дата публікації >=   
рейтинг (**score**) в звіт повинні потрапити записи де рейтинг >=

8. Результуючі дані потрібно зберегти в файл report.csv назва файлу також повинна бути в конфіг файлі.

9. Код повинен притримуватися стандарту [pep8](http://defpython.ru/pep8) <http://defpython.ru/pep8>  
Перевірити свій код можна з допомогою ресурсу <http://pep8online.com/checkresult>

**Завдання для бажаючих(хто на 5)**

10. Якщо в полі текст зустрічається [тег](https://ru.wikipedia.org/wiki/%D0%A2%D0%B5%D0%B3_(%D1%8F%D0%B7%D1%8B%D0%BA%D0%B8_%D1%80%D0%B0%D0%B7%D0%BC%D0%B5%D1%82%D0%BA%D0%B8)), значення яке взято <sone tag></some tag> його потрібно вирізати, прибрати разом з внутрішнім текстом якщо такий є   
<http://joxi.ru/nAyKNlwuX53DLm> виділений блок повинен зникнути і інші якщо такі є, зробити це регуляркою.