

Модуль
Работа с данными
в Python



Домашняя работа № 1. Работа с файлами

Python для продвинутых
специалистов

Домашнее работа № 1

Цель работы: научиться использовать Python для чтения данных из текстовых файлов и загрузки данных в базу.

В качестве примера мы будем использовать открытые данные о российских компаниях, индивидуальных предпринимателях и видах экономической деятельности. Анализ этих данных является важным этапом многих задач в области финансовых технологий.

- Единый государственный реестр юридических лиц, на начало 2022 года - <https://ofdata.ru/open-data/download/egrul.json.zip>
- Единый государственный реестр индивидуальных предпринимателей, на начало 2022 года - <https://ofdata.ru/open-data/download/egrip.json.zip>
- Финансовая отчетность от Росстата и ФНС за 2011–2020 годы – https://ofdata.ru/open-data/download/finances_2020.json.zip
- Общероссийский классификатор видов экономической деятельности – https://ofdata.ru/open-data/download/okved_2.json.zip

Задание 1

1. Скачайте общероссийский классификатор видов экономической деятельности в формате JSON по ссылке – https://ofdata.ru/open-data/download/okved_2.json.zip
2. Установите SQLite command-line tools, скачать можно по ссылке – <https://www.sqlite.org/download.html>
3. Создайте базу данных SQLite hw1.db

4. Разработайте программу на Python, которая выполняет следующие действия:

- Читает файл с общероссийским классификатор видов экономической деятельности в формате JSON.
- Записывает классификатор видов экономической деятельности в базу данных hw1.db, в таблицу okved со столбцами: code, parent_code, section, name, comment (названия столбцов таблицы совпадают с названиями ключей JSON).

Задание 2

1. Скачайте архив Единого государственного реестр юридических лиц в формате JSON по ссылке <https://ofdata.ru/open-data/download/egrul.json.zip>.
2. Разработайте программу на Python, которая выполняет следующие действия:
 - Читает файл Единого государственного реестр юридических лиц в формате JSON.
 - Выбирает из классификатора компании, занимающиеся деятельностью в сфере телекоммуникаций (Группировка ОКВЭД 61).
 - Записывает информацию о компаниях, занимающихся деятельностью в сфере телекоммуникаций, в базу данных hw1.db, таблица telecom_companies. Столбцы таблицы выберите сами на свое усмотрение (минимум 5 столбцов, включая название компании, ИНН, код ОКВЭД).

Для сдачи задания пришлите исходные коды программ на Python.