

Основы программирования в Python – 2021/22 уч. г. // Проект // ОП «Политология»

Проект представляет собой написание корректно работающей программы прикладного назначения с последующей устной защитой в виде презентации.

Проект включает в себя следующие части:

- описание работы (описание идеи, функциональности, источников данных для будущей программы)
- программа Python (файл Jupyter Notebook с расширением .ipynb или файл с расширением .py)
- документация (pdf-файл с описанием назначения и принципов работы программы)
- устная презентация, включающая публичную демонстрацию работы программы, ответы на вопросы по коду

Проект обязательно должен включать использование навыков, полученных в результате прохождения тем второго модуля, а именно выгрузку данных с веб-страницы или с помощью API, сохранение данных в датафрейм pandas с последующей их обработкой, по возможности - визуализацию данных.

Проект выполняется индивидуально или в группе из 2 человек. При коллективной защите проекта все члены группы должны быть подготовлены и готовы ответить на вопросы по проекту, каждому участнику группа ставится индивидуальная оценка за защиту.

Оценка за проект определяется по следующей формуле (без округления):

$$0.15 * \text{План} + 0.4 * \text{Программа} + 0.15 * \text{Документация} + 0.3 * \text{Презентация}$$

Подбирайте сайт для сбора данных так, чтобы

- было возможно обработать не одну страницу сайта, а больше (хотя бы 3 – 5)
- было возможно собрать хотя бы две количественные переменные и хотя бы две качественные переменные

Например, <http://www.biblio-globus.ru/search/catalog/products?query=Pythonpage=1> – здесь мы можем «полистать» страницы, собрать ссылки на каждую книжку, вытащить оттуда, например, цену и тираж, возрастную категорию и статус (в наличии/не в наличии). Подробно разбор этого сайта мы посмотрим на парах.

1. Описание

– (2 балла) Опишите, что именно планируете сделать:

- Укажите сайт, с которого планируете собрать данные, объясняете, почему вас интересует именно этот сайт
- Укажите, какие именно данные вас интересуют, для чего они могут пригодиться

– (3 балла) Опишите, как именно вы планируете собрать данные (какие именно средства, библиотеки Python вам потребуются и для чего они нужны)

Уже после создания программы и сбора данных:

– (2 балла) Проинтерпретируйте собранные данные (посмотрите основные статистики по количественным и качественным переменным). Если обращаться к примеру выше, то можно посмотреть среднее значение по цене или самую часто встречающуюся возрастную категорию.

- (2 балла) Опишите, с какими трудностями столкнулись, как их решали (от чего пришлось отказаться в работе или что нового было необходимо изучить).
- (1 балл) Опишите, что еще можно было бы сделать в вашем проекте в будущем.

2. Программа (код)

- (1 балл) Соблюдены правила написания кода ([PEP8](#))
- (1 балл) Корректно и уместно использованы функции и методы (например, строка кода `str(input())` не несет в себе никакого смысла – применение функции `str()` тут лишнее т.к. `input()` и так нам выдает строку).
- (1 балл) Написанный код собирает информацию с одной страницы.
- (2 балла) Написанный код ходит по страницам и собирает информацию с каждой страницы.
- (2 балла) Корректно обрабатывается код веб-страниц, производится правильный поиск необходимых элементов по тегам/значениям атрибутов соответствующих тегов.
- (2 балла) Наличие минимум трех собственных функций (для обработки информации на веб-странице, сбора данных со страниц в зависимости от пользовательского запроса, описания собранных данных).
- (1 балл) Использование в коде списковых включений/`lambda`-функций/(грамотное использование) конструкций `try-except`.

Дополнительные баллы

- (1 балл) За использование Selenium
- (1 балл) За визуализацию данных

3. Документация

Примеры документаций можно посмотреть в описании любого модуля Python. Например, для модуля `math` документация такая: <https://docs.python.org/3/library/math.html>

Документация должна содержать два раздела: 1) с описанием функций, которые вы используете в коде; 2) с описанием своих собственных функций

Для первого раздела:

- (5 баллов) Указано, из каких модулей импортированы функции. Также указано, какие аргументы они принимают на вход и что выдают. Важно: не копируйте все из документации к этим модулям, опишите, что функции делают именно у вас в коде.

Для второго раздела:

- (5 баллов) Указано, что функция принимает на вход. Расписано, что происходит с входными параметрами внутри функции, а также, что используется для их обработки (циклы, методы, другие функции). Указано, что должно ожидаться в результате корректной работы функции.

Описание, код и документация сдаются в день дедлайна, который будет назначен в апреле-мае.

4. Презентация/устная защита

В презентации кратко рассказываете о чем ваша работа. Также будьте готовы открыть код и прокомментировать его.