### ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

### К курсовому проекту по теме

# "Методы сбора и обработки данных из сети Интернет"

### Задача

Написать программу, которая собирает динамически загружаемые товары и складывает данные в базу данных (БД)

### 2 Исходные данные

сайт сети "Пятерочка" (раздел "Спецпредложения" - '<https://5ka.ru/special_offers'>)

### 3 Принципы решения задачи

Ввиду того, что в качестве исходных данных используются *динамически загружаемые товары* для решения поставленной задачи используется **Selenium WebDriver.**

Это программная библиотека для управления браузерами. Библиотека используется для отправки HTTP запросов драйверу, в которых указано действие, которое должен совершить браузер в рамках текущей сессии. Примерами таких команд могут быть команды нахождения элементов по локатору, переход по ссылкам, парсинг текста страницы/элемента, нажатие кнопок или переход по ссылкам на странице веб-сайта.

Программный код решения задачи приведен в файле «T7-Kursovik.ipynb». Основные решения сопровождаются подробным комментарием.

### 4 Выходные данные

В качестве выходных данных используются:

'*name*' – наименование товара;

'*price*' – цена товара;

‘*discont*’ – размер скидки в процентах.

### 5. Сохранение полученных данных

Полученные в результате выполнения программного кода данные сохраняются в следующих форматах:

- БД типа MongoDB (БД “*goods\_5ka*”, коллекция “*actions\_db*”

- файл типа .csv (goods\_5ka.csv). Для корректной записи кириллицы используем кодировку 'utf-16'.

### 6. Состав репозитория

«T7-Kursovik.ipynb» - программный код.

«goods\_5ka.csv» - выходные данные.

Файлы, подтверждающие работоспособность кода:

«Скан БД после занесения акционных товаров.jpg»

«Скан работоспособности кода.jpg»