

# Парсер-выгрузка данных по выборам в России из ЦИК

## Общая информация

Репозиторий проекта:

<https://github.com/EmilStasevski/Parsing-CEC-Central-Election-Commission-of-the-Russian-Federation>

Контакты участников (tg):

- Сташевски Эмиль @emilstrider
- Рубанов Владислав @rubanoww
- Жужлев Борис @Zhbzhhb

## Описание проекта

### Функционал

Конечная реализация проекта представляет из себя **веб-приложение**.

По умолчанию пользователь попадает на главную страницу и имеет возможность выбрать нужный ему набор фильтров для получения данных о выборах любого уровня (от федеральных до муниципальных), проведенных в России с 2003 г. на любом уровне (т.е. все среди всех данных, внесенных в информационную систему ГАС «Выборы» и доступных в цифровом формате).

Для навигации и выбора пользователь может использовать интерактивную карту России, а также графики, имеющие сквозную фильтрацию (например, выбрать конкретного кандидата и/или регион, “провалиться” глубже в конкретный УИК).

После выбора нужных фильтров пользователь может ознакомиться с краткой аналитикой результатов (описательные статистики по в целом / в разбивке по регионам и уровням), либо выгрузить данные в формате excel/csv в наименьшем доступном разрезе (по УИК-ам) для дальнейшего анализа.

### Техническая логика работы

- Пользователь на сайте приложения выбирает набор фильтров и отправляет запрос на поиск данных
- Парсер принимает нужные параметры и обращается к сайту <http://www.vybory.izbirkom.ru/>, на котором находится информация про все выборы в РФ с 2003 г. на всех уровнях
- Парсер выбирает нужные фильтры, пробегается по иерархической выдаче в зависимости от нужной глубины разреза (страна / регион / район / муниципалитет / УИК) и сохраняет данные внутри себя
- После окончания сбора данных со страницы, парсер сохраняет их в файл и передает в веб-приложение для отрисовки

- Веб-приложение принимает и читает данные и отрисовывает их в UI-интерфейсе, приводит краткую аналитику на графиках и диаграммах
- При необходимости полученный файл может быть выгружен в формате excel/csv

## Предполагаемые для использования технологии и модули

Технологии:

- Веб: HTML, CSS

Модули python:

- dev-back: flask
- dev-front: dash
- визуализация: matplotlib, seaborn, plotly
- парсинг: selenium, bs4, requests, time
- работа с данными: pandas, numpy, statistics, collections

## Распределение ролей

- Парсинг данных (selenium), dev-back (путь от сбора данных до приложения), работа с данными – Рубанов Владислав
- dev-front, dev-back (путь от приложения до UI), CSS – Сташевски Эмиль
- Парсинг данных (bs4, requests), визуализация, политическая экспертиза – Жужлев Борис