



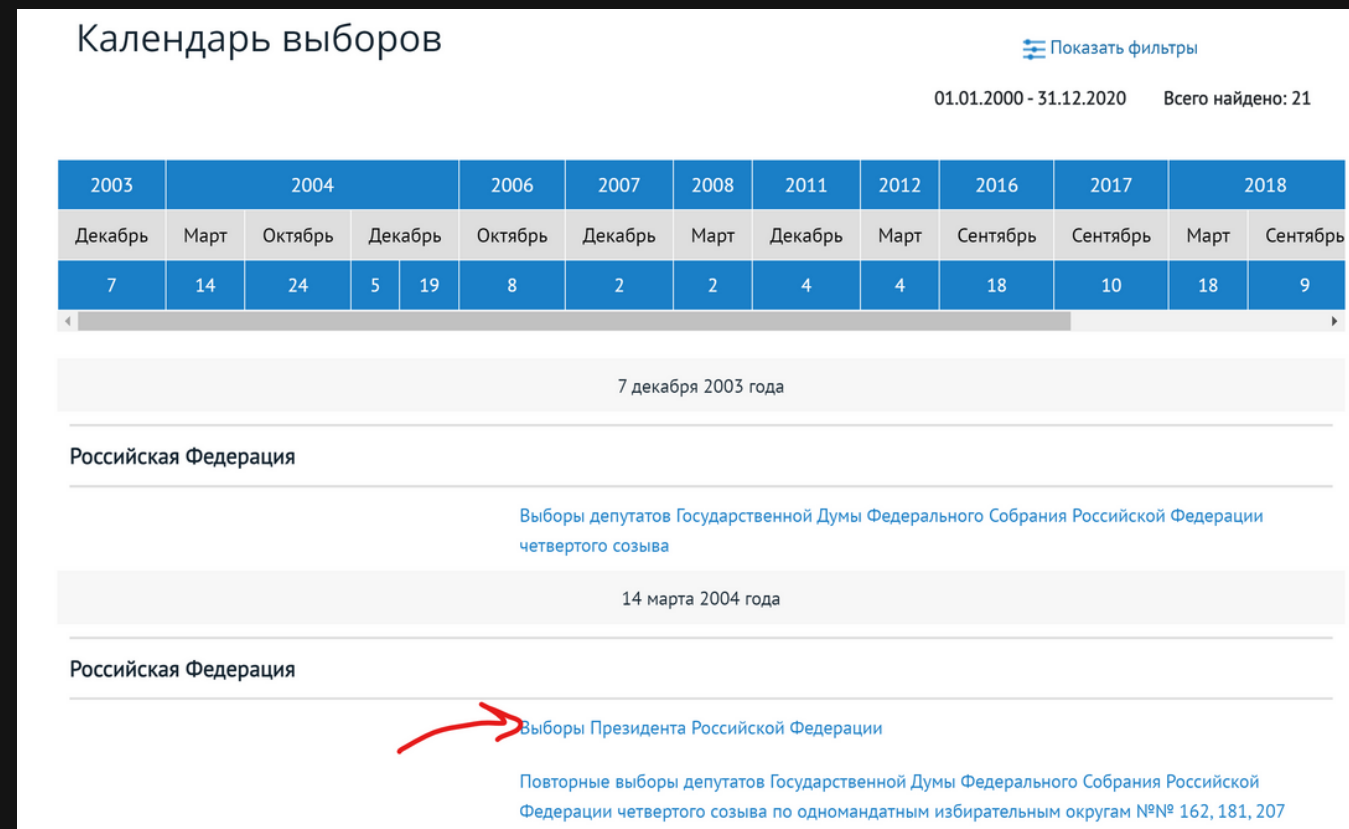
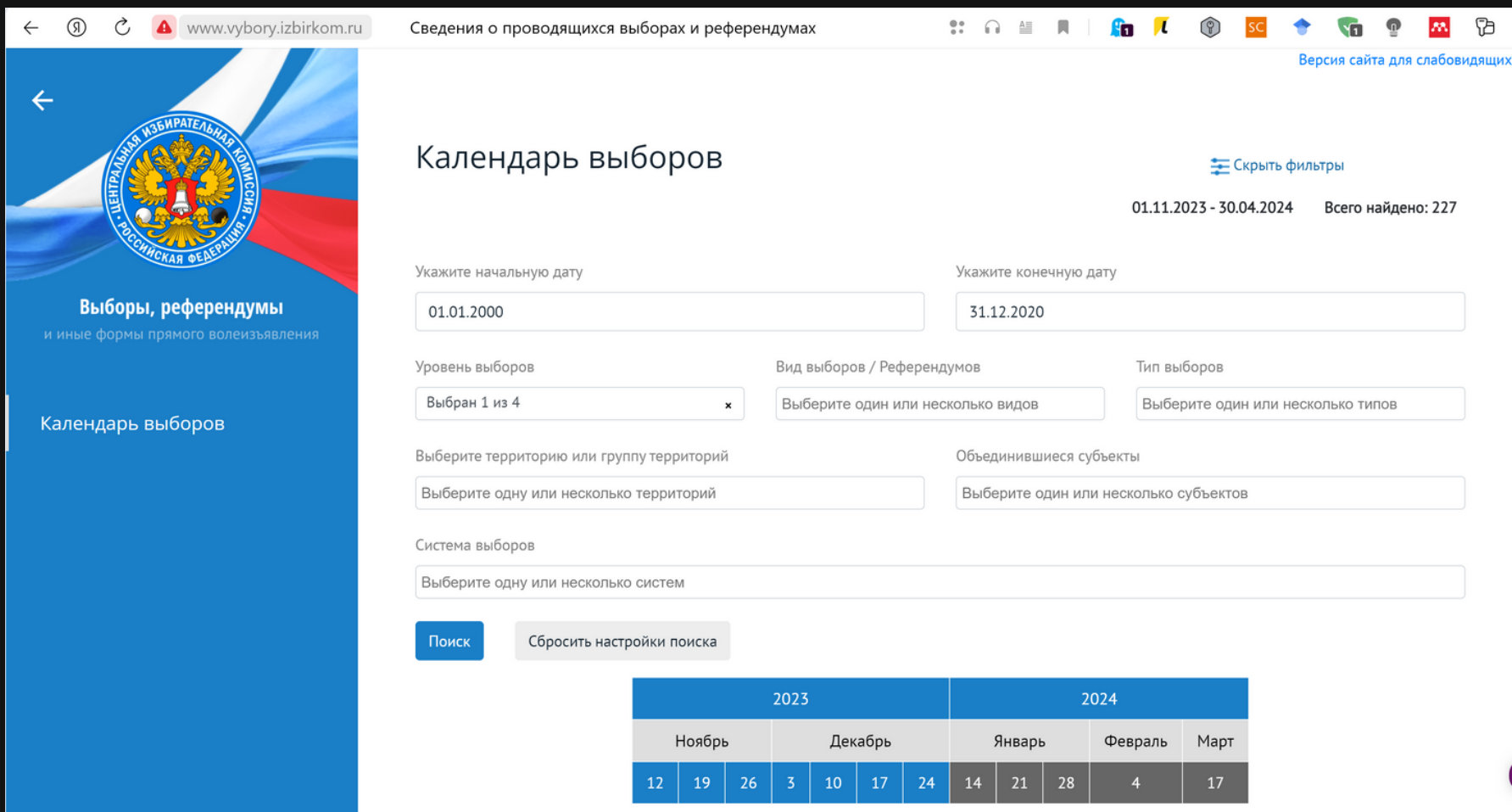
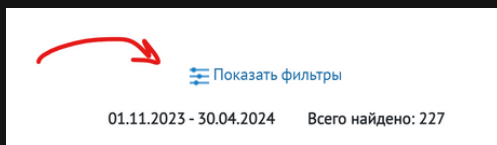
Парсинг ЦИК-а

Центральная избирательная комиссия

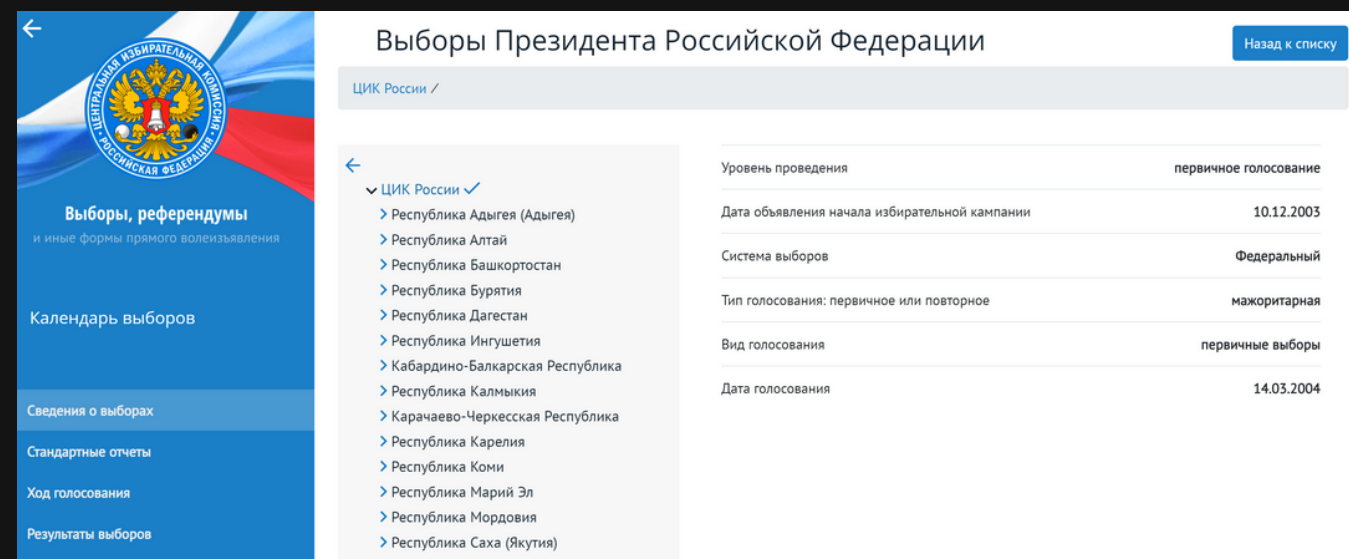


“Путь пользователя”

Или под кого нам нужно было прикинуться



Выбираем набор фильтров,
конкретные выборы,
“проваливаемся” на нужный уровень





Архитектура проекта

И подводные камни, с которыми мы
столкнулись

Почему selenium?

Трудности

1

**Сильная защита
страницы**

**Обычные
парсеры не
работают,
роботам ничего
нельзя + капчи**

2

**Данные хранятся
странно**

**Они хранятся по-
разному в разных
выборах, но
одинаково внутри
вложенных структур
(регионы, ТИК-и, etc.)**

3

**Сайт тупой и
медленный**

**Сайт долго грузится
при открытии и
переходе по
разделам, из-за чего
парсер может
сломаться**

Разный формат 1

2004

```
<table class="table-borderless table-sm" style="width:100%">
  <tbody>
    <tr>
      <td>...</td>
      <td>Сахалинская область</td> == $0
    </tr>
  </tbody>
</table>
<br>
```

2008

```
<table class="table-borderless" width="100%">
  <tbody>
    <tr>
      <td width="30%" style="vertical-align:top;"> Наименование избирательной комиссии
      </td>
      <td class="text-center">
        <b>Республика Бурятия</b> == $0
      </td>
    </tr>
  </tbody>
</table>
```

2012

```
<table class="table-borderless" width="100%">
  <tbody>
    <tr>
      <td width="30%" style="vertical-align:top;"> Наименование избирательной комиссии
      </td>
      <td class="text-center">
        <b>Республика Алтай</b> == $0
      </td>
    </tr>
  </tbody>
</table>
<div>...</div>
<div class="table-responsive">...</div>
<div class="table-wrapper">...</div>
```

2018

```
<table class="table-borderless" width="100%">
  <tbody>
    <tr>
      <td width="30%" style="vertical-align:top;"> Наименование избирательной комиссии
      </td>
      <td class="text-center">
        <b>Республика Ингушетия</b> == $0
      </td>
    </tr>
  </tbody>
</table>
<div>...</div>
```

Уровень проведения	Федеральный
Дата объявления начала избирательной кампании	10.12.2003
Система выборов	мажоритарная
Тип голосования: первичное или повторное	первичные выборы
Вид голосования	первичное голосование
Дата голосования	14.03.2004
Наименование Избирательной комиссии	Сахалинская область

Сахалинская область	
- 20	
- 20	

Итоги голосования							
	Сумма	Александровск-Сахалинская	Анивская	Долинская	Корсаковская	Курильская	Макаровский
1	Число избирателей, внесенных в список	411142	13247	11684	22074	31253	5590
2	Число полученных избирательных бюллетеней	412028	12810	11161	20889	30253	5440
3	Число избирательных бюллетеней, выданных досрочно	1021	0	34	0	38	54
4	Число избирательных бюллетеней, выданных в день голосования	211702	6355	7644	10831	15644	3685
5	Число избирательных бюллетеней, выданных вне помещения	12838	776	592	770	798	194
6	Число погашенных избирательных бюллетеней	186467	5679	2891	9288	13773	1507
7	Число избирательных бюллетеней в переносных ящиках	13856	776	626	770	836	248
8	Число бюллетеней в стационарных ящиках для голосования	211551	6354	7643	10819	15643	3684
9	Число недействительных избирательных бюллетеней	1844	48	79	125	167	27
10	Число действительных избирательных бюллетеней	223563	7082	8190	11464	16312	3905
11	Число полученных открепительных удостоверений	6750	213	190	380	532	83
12	Число выданных открепительных удостоверений	3266	127	156	161	142	40
13	Число избирательных бюллетеней, полученных в день голосования	211702	6355	7644	10831	15644	3685

Разный формат 2

2004

2008

Дата голосования: 02.03.2008		Республика Бурятия							
Наименование избирательной комиссии									
	Сумма	Баргузинская	Баунтовская	Бичурская	Джидинская	Еравнинская	Заиграевская	Закаменская	
15	удостоверений, выданных избирателям ТИК	341	5	0	13	2	1	5	21
16	Число утраченных избирательных бюллетеней	4	0	0	0	1	0	0	0
17	Число избирательных бюллетеней, не учтенных при получении	0	0	0	0	0	0	0	0
17а	Число утраченных открепительных удостоверений	0	0	0	0	0	0	0	0
17б	Число открепительных удостоверений, не учтенных при получении	0	0	0	0	0	0	0	0
18	Богданов Андрей Владимирович	5007 1.03%	132 0.95%	45 0.79%	92 0.73%	98 0.61%	58 0.66%	228 0.99%	102 0.67%
19	Жириновский Владимир Вольфович	40110 8.29%	1077 7.79%	528 9.32%	1417 11.20%	1079 6.68%	454 5.20%	2101 9.10%	617 4.05%
20	Зюганов Геннадий Андреевич	89315 18.46%	2091 15.12%	867 15.30%	3301 26.08%	3347 20.73%	1270 14.55%	4427 19.18%	2316 15.20%
21	Медведев Дмитрий Анатольевич	342736 70.84%	10361 74.93%	4156 73.32%	7681 60.69%	11343 70.25%	6843 78.42%	16023 69.43%	11975 78.60%

Дата голосования: 04.03.2012		Республика Алтай											
Наименование избирательной комиссии													
	Сумма	Горно-Алтайская городская	Кош-Агачская	Майминская	Онгудайская	Турочакская	Улаганская	Усть-Канская	Усть-Коксинская	Чемальская	Чойская	Шебалинская	
4	Число избирательных бюллетеней, выданных в помещении для голосования в день голосования	94827	26219	7608	12998	7070	5394	4671	6841	9025	4795	4086	6120
5	Число избирательных бюллетеней, выданных вне помещения для голосования в день голосования	5889	896	407	1017	521	349	330	417	480	375	517	580
6	Число погашенных избирательных бюллетеней	46680	12876	5110	6960	3909	3311	2752	2661	1649	1755	1897	3800
7	Число избирательных бюллетеней в переносных ящиках для голосования	7050	893	1282	1017	521	395	575	416	480	375	516	580
8	Число бюллетеней в стационарных ящиках для голосования	94805	26206	7608	12995	7070	5391	4669	6841	9025	4794	4086	6120
9	Число недействительных избирательных бюллетеней	1141	415	51	144	94	51	41	50	150	52	50	43
10	Число действительных избирательных бюллетеней	100714	26684	8839	13868	7497	5735	5203	7207	9355	5117	4552	6657
11	Число полученных открепительных удостоверений	3763	668	188	388	371	183	130	188	180	94	187	334
Данные территориальной избирательной комиссии о числе открепительных удостоверений													
	Сумма	Горно-Алтайская городская	Кош-Агачская	Майминская	Онгудайская	Турочакская	Улаганская	Усть-Канская	Усть-Коксинская	Чемальская	Чойская	Шебалинская	
а	Число открепительных удостоверений, полученных территориальной избирательной комиссией	2960	700	300	400	230	230	150	200	200	100	200	250
б	Число открепительных удостоверений, выданных нижестоящим участковым избирательным комиссиям	2762	658	285	383	221	182	139	188	189	94	187	236
в	Число неиспользованных открепительных удостоверений, оставшихся в территориальной избирательной комиссии	34	0	2	0	0	26	0	1	1	2	2	0
г	Число утраченных в территориальной избирательной комиссии открепительных удостоверений	0	0	0	0	0	0	0	0	0	0	0	0

2012

2018

Дата голосования: 18.03.2018		Республика Ингушетия					
Наименование избирательной комиссии							
	Сумма	Джейрахская	Карабулакская городская	Магасская городская	Малгобекская городская	Малгобекская	
1	Число избирателей, включенных в список избирателей	219305	2770	15091	4053	25100	
2	Число избирательных бюллетеней, полученных участковой избирательной комиссией	209693	2528	13400	3862	24215	
3	Число избирательных бюллетеней, выданных избирателям, проголосовавшим досрочно	544	544	0	0	0	
4	Число избирательных бюллетеней, выданных в помещении для голосования в день голосования	177335	1900	10684	2700	20109	
5	Число избирательных бюллетеней, выданных вне помещения для голосования в день голосования	1862	47	246	327	87	
6	Число погашенных избирательных бюллетеней	29952	37	2470	835	4019	
Число избирательных							

Архитектура проекта



Распределение обязанностей

Анализ сайта ЦИК-а (фильтры) и экспертная поддержка в области выборов в РФ -- **Жужлев Б.**

dev-backend: парсер на python (selenium, bs4, lxml) -- **Рубанов В.**

Парсинг (тестирование, написание скриптов, выгрузки), обход блокировок -- **Рубанов В.**

Поднятие **БД Firebase** и передача данных от back на front с помощью python -- **Сташевски Э.**

dev-front: написание **приложения** и **сайта** проекта на **React (JS)** -- **Сташевски Э.**

Создание графиков и аналитика результатов выборов -- **Жужлев Б.**

Стилистическое оформление сайта проекта -- **совместно**

Организация и поддержка репозитория на **GitHub** -- **Рубанов В.**

Итог

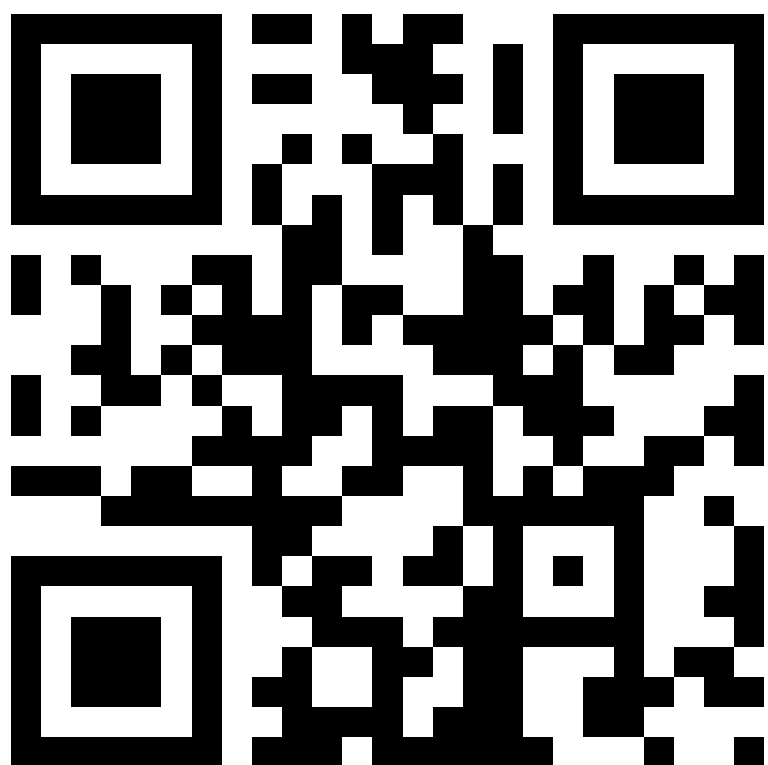
Чего мы добились?

- проанализировали, как хранятся данные по результатам выборов в РФ
- обошли блокировки
- написали полноценный парсер, добились нулевого числа пропусков и высокого качества данных
- создали приложение с возможностью просмотра и анализа результатов выборов на разных уровнях с визуализацией
- сделали возможность выгрузки данных

Что бы мы улучшили?

- увеличили бы число доступных сценариев (региональные выборы)
- изучили бы, как более умно обходить капчи, чтобы ускорить процесс парсинга
- потратили бы больше времени на сайт и UI :)

DEMO



или [ссылка](#) на приложение



Спасибо за внимание!

Будем рады ответить на вопросы!