

## Российская общественная инициатива <https://www.roi.ru/>

Описание, что какая колонка обозначает (ну мало ли): <https://static.roi.ru/content/doc/roi-api-v1.0.pdf>

В том списке они забыли поле “decision”, поле с текстом о предлагаемом решении. Я забираю все поля, кроме внутренних id разных статусов, documents и authorship. Почти все тексты написаны в несколько абзацев – я всё это сохранил.

Этот сайт имеет очень статичное api. Здесь принудительная разбивка по статусам, и нет никакой возможности фильтровать по дате.

Самые сырые данные находятся в файлах:

- poll.csv, poll.xlsx, 2477 записей
- advisement.csv, advisement.xls, 2 записи
- complete.csv, complete.xlsx, 21 запись
- archive.csv, archive.xlsx, 4947 записей

То есть всего  $2477 + 2 + 21 + 4947 = 7447$ , что совпадает с числом на главной сайта. Это хорошо.

Две петиции вылезли за ограничение длины ячейки в Excel (32767 символов) – это петиции с id 14332 и 24800. И они обе об обеспечении информационно-психологической безопасности, что намекает. Я положил эти пасты отдельно в description\_14332.txt и description\_24800.txt – потому что в таблицу они не влезают. В таблице – обрезанные варианты.

Дальше я смержил все петиции в один файл – all.xlsx, отсортировал по date.poll.begin и отфильтровал по [1.02.2015:29.02.2016] – final\_roi.xlsx. Осталось больше, чем было указано: тут 2625, а ты говорила про 2443. Может быть ты учитывала 12 месяцев, а тут сейчас охватывается 13 (февраль дважды).

## Демократор.py <https://democrator.ru/>

Здесь есть несколько проблем:

- Сайт не показывает точную дату, есть только обозначения “n дней назад”, “n месяцев назад”, “n лет назад”. И даже в таком виде этой информации нет в каком-то определенном поле а-ля “Петиция создана:”, а лежит в истории. Даты, которые я взял – первая запись в этой истории, обычно это событие создания петиции. Я засунул эту информацию в колонку “raw\_start\_date”, а в следующей колонке руками посчитал дату создания. Для дат меньше месяца – с точностью до дня, для дат меньше года – с точностью до месяца, для дат больше года – не стал писать “2015”, но, я думаю, при желании это будет не сложно сделать.
- Предполагается, что этих петиций за выбранный промежуток там 64. Видимо, если отбросить петиции с датой “2 года назад” (их 12) и петиции за март (3), то получится как раз  $80 - 12 - 3 = 65$ . Может, на момент подсчетов еще не настало 29 февраля, например, и их было 64. Ок, тогда в файле final\_democrator.xlsx будут только петиции нужного периода времени.

## Change.org <https://change.org/>

У них есть api для того, чтобы доставать петицию, но нет api для поиска. Так что я открыл вкладку последние и забрал все 744 страницы id, после чего запросил все 7432 петиции в api. api отдает какие-то странные заголовки – например, “ернатор : Сохранить на посту главного тренера ХК “Водник”Батова Олега Валерьевича. ” вместо “Сохранить на посту главного тренера ХК “Водник”Батова Олега Валерьевича.”, но иногда совсем промашки, например “?????????? ???? ???? ? ???? ????-?????????” или “femojis-ru” вместо “Пришло время для #фемоджи! (Нам нужна ваша помощь!)”. Так что я обошел все страницы и распарсил все заголовки отдельно (new\_title\_change.org.xlsx).

Снова есть длинные петиции: это 3121266, 4261744, 4308608 – снова приложил их отдельно.

Отфильтрованные по дате данные лежат в final\_change.org.xlsx.