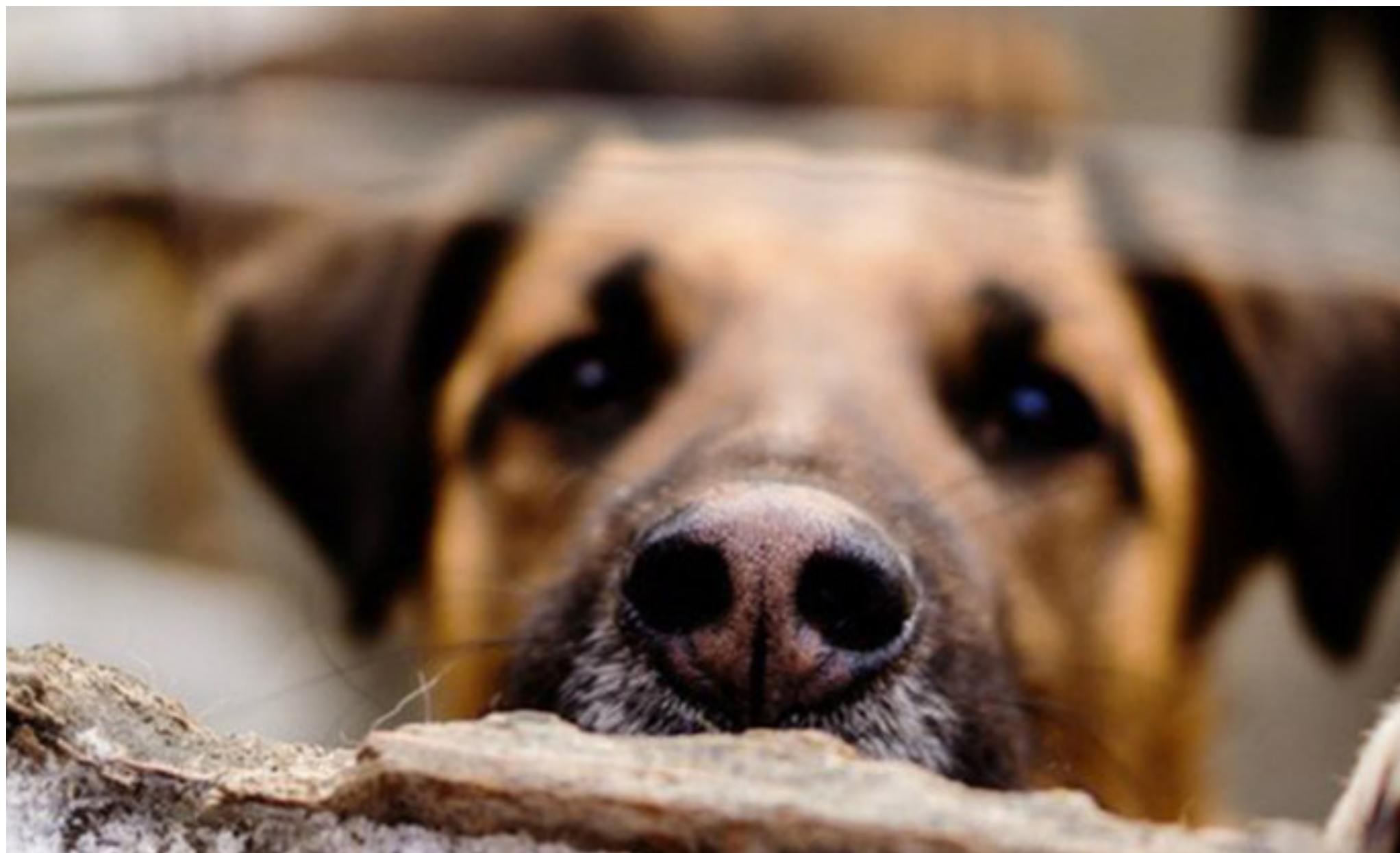


# Исследование краудфандинга в благотворительности

Есть ли связь между особенностями текста и  
успешностью сбора?





на примере [planeta.ru](https://planeta.ru)

# Есть похожие исследования и датасеты, но....

- Не на русском
- Не про тексты
- Не про благотворительность



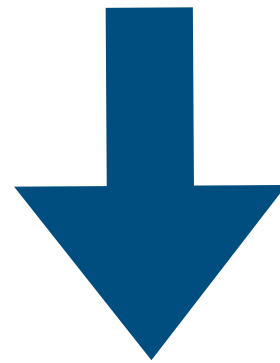


# Обычно в краудфандинге:

- Видео и фото
  - Текст
- + Подарки и бонусы**



В благотворительности - безвозмездные  
вклады  
(пожертвования)



Роль текста особенно важна

# Этапы

1. Собрать свой датасет
2. Проанализировать тексты
3. Сопоставить особенности текстов с показателями успешности сбора



# Этап 1

сбор данных с сайта





# Инструменты



# **Python 3**

## **+**

# **Google Colab**

# Основной инструмент сбора данных

- **Requests** - не работает с динамическими сайтами (javascript, ajax)
- **Requests HTML** - должна работать с динамическими, но не получилось (возможно, библиотека не обновляется)
- **Selenium WebDriver** - да!



# Selenium WebDriver

- инструмент для автоматизации действий веб-браузера (имитирует действия пользователя)
- используется для тестирования Web-приложений, но не только
- Может быть использован для регулярного получения данных с сайтов
- [www.selenium.dev](http://www.selenium.dev)



# Локаторы

- указывают на нужный элемент
- бывают простые: по id или name
- и сложные: css и xpath (все спорят, что лучше)
- вебинар - как их писать:

<http://prostoitblog.ru/kak-sostavlyat-xpath-i-css-selektoryi/>



```
def get_result(url):
    driver.get(url)
    wait = WebDriverWait(driver, 15)
    try:
        x=wait.until(EC.presence_of_element_located((By.XPATH, '//div[@class="project-view-meta_amount-val ng-star-inserted"]'))).text
    except:
        driver.refresh()
    try:
        x=wait.until(EC.presence_of_element_located((By.XPATH, '//div[@class="project-view-meta_amount-val ng-star-inserted"]'))).text
    except:
        x = "NaN"
    return(x)
```

```
def get_goal(url):
    driver.get(url)
    wait = WebDriverWait(driver, 15)
    try:
        x=wait.until(EC.presence_of_element_located((By.XPATH, '//span[@class="project-view-meta_amount-lbl-val"]'))).text
    except:
        driver.refresh()
    try:
        x=wait.until(EC.presence_of_element_located((By.XPATH, '//span[@class="project-view-meta_amount-lbl-val"]'))).text
    except:
        x = "NaN"
    return(x)
```

```
def get_donations(url):
    driver.get(url)
    wait = WebDriverWait(driver, 15)
    try:
        x=wait.until(EC.presence_of_element_located((By.XPATH, '//div[@class="project-view-info_lbl"][text() ="Поддержали"]/preceding-sibling::*'))).text
    except:
        driver.refresh()
    try:
        x=wait.until(EC.presence_of_element_located((By.XPATH, '//div[@class="project-view-info_lbl"][text() ="Поддержали"]/preceding-sibling::*'))).text
    except:
        x = "NaN"
    return(x)
```

# Проблемы парсинга динамического сайта





# Локатор верный, элемент виден «ВЖИВУЮ», но не ловится:

1. еще

2. уже

# Что делать?

- Waits: explicit & implicit
- Expected conditions
- Try... except

Помогает, но проблема остается + появляются дубликаты



```
def get_briefly(url):  
    driver.get(url)  
    wait = WebDriverWait(driver, 15)  
    try:  
        x=wait.until(EC.presence_of_element_located((By.XPATH, '//p[@class="project-view_short-desc ng'  
    except:  
        driver.refresh()  
        try:  
            x=wait.until(EC.presence_of_element_located((By.XPATH, '//p[@class="project-view_short-desc'  
        except:  
            x = "NaN"  
    return(x)
```

# Формирование датасета

- Словарь
- Pandas Dataframe

```
[8] DataPl = {  
    "Title": [],  
    "Briefly": [],  
    "Description": [],  
    "Result": [],  
    "Goal": [],  
    "Donations": [],  
    "Start": [],  
    "Finish": []  
}
```

```
[9] for i in projects:  
    title = get_title(i)  
    briefly = get_briefly(i)  
    description = get_description(i)  
    result = get_result(i)  
    goal = get_goal(i)  
    donations = get_donations(i)  
    start = get_start(i)  
    finish = get_finish(i)  
    DataPl["Title"].append(title)  
    DataPl["Briefly"].append(briefly)  
    DataPl["Description"].append(description)  
    DataPl["Result"].append(result)  
    DataPl["Goal"].append(goal)  
    DataPl["Donations"].append(donations)  
    DataPl["Start"].append(start)  
    DataPl["Finish"].append(finish)  
    print(title)  
  
#print(DataPl)
```

```
▶ PLdataset = pd.DataFrame(DataPl)  
  PLdataset  
  |
```

```
[23] PLdataset.head(10)
PLdataset.tail(10)
```



	Title	Briefly	Description	Result	Goal	Donations	Start	Finish
0	В одном шаге от выздоровления	У Светланы с мужем трое детей. Илюше 13 лет, А...	Фонд борьбы с лейкемией существует с 2014 года...	162 000 Р	150 000 Р	186 раз	14 декабря 2018	15 ноября 2018
0	Новогодняя ночь одиноким бабушкам и дедушкам!	Волонтеры устроят праздник в новогоднюю ночь д...	0:42\n Мы Смоленская региональная группа благ...	43 120 Р	40 000 Р	102 раза	21 декабря 2018	14 декабря 2018
0	Теплый дом для кошек старейшего приюта Сибири	Сбор средств на строительство нового теплого д...	Один из старейших в стране, приют для бездомны...	520 500 Р	480 000 Р	712 раз	22 августа 2019	24 мая 2019
0	Спектакль "Записки сумасшедшего". Н. Гоголь	«Подпольный театр Ш.З.» — театральная студия Ц...	Меня зовут Викентий Екимов.\nЯ режиссёр и пост...	21 303 Р	20 000 Р	30 раз	13 июня 2013	16 мая 2013
0	Установка роллпандусов в доме престарелых	Для лежачих бабушек и дедушек большая радость ...	Дорогие друзья!\n Мы, благотворительный фонд п...	150 133 Р	140 000 Р	170 раз	27 ноября 2013	13 сентября 2013
0	Покупка медикаментов для детей-бабочек	Всем детям нужны игрушки, а детям с буллезным ...	Благотворительный Фонд «Дети БЭЛА» - это орган...	107 400 Р	100 000 Р	153 раза	21 ноября 2013	25 сентября 2013
0	Уникальный Центр для животных «Мокрый нос»	Наша команда строит уникальный и многофункционал...	Как все начиналось?\nВсе началось 3 года назад...	5 367 000 Р	5 000 000 Р	4729 раз	12 октября 2014	11 апреля 2014
0	"Лыжи мечты" - реабилитация детей с ДЦП	"ЛЫЖИ МЕЧТЫ" – программа эффективной реабилита...	«Лыжи мечты» – программа эффективной реабили...	1 823 952 Р	1 710 000 Р	163 раза	25 мая 2015	15 апреля 2015
0	Ветеринарное оборудование для приюта «Ласка»	Приют "Ласка" вновь просит Вас о помощи! В при...	В Барнауле — довольно крупном городе — существ...	53 360 Р	50 100 Р	138 раз	16 июня 2014	2 июня 2014
0	Праздничный ужин для пожилых и одиноких.	Волонтеры организуют угощение, поздравления, о...	Зачем мы проводим Пасхальный Праздничный ужин...	85 500 Р	80 000 Р	46 раз	12 апреля 2015	20 марта 2015



цинско	Мы собираем деньги на строительство	<p>Дорогие друзья! Спасибо за интерес к нашим проектам и вашу помощь, за отзывчивость и неравнодушие. Благотворительный фонд «Шередарь» уже 4 года помогает детям и подросткам, пережившим онкологическое заболевание. Мы являемся единственной в России некоммерческой организацией, занимающейся развитием детского онкологического центра.</p> <p>Так как безопасность жизни и здоровья детей в реабилитационном центре — главное условие проведения реабилитационных мероприятий. Согласно проекту, в новом медицинском блоке будет приемный кабинет, два инфекционных отделения.</p> <p>В данный момент мы собираем деньги на монтаж охранно-пожарной сигнализации в здание медпункта.</p> <p>С технической точки зрения автоматическая пожарная сигнализация – это довольно сложная система, состоящая из:</p> <ul style="list-style-type: none"> <li>- аппаратура для обнаружения источника возгорания;</li> <li>- устройство речевого оповещения;</li> <li>- системы пожаротушения и дымоудаления;</li> <li>- устройство подачи сигнала на СКУД (система контроля и управления доступом);</li> <li>- сенсорные устройства (датчики, извещатели) – обнаруживают возгорание и передают сигнал на приемно-контрольное устройство;</li> <li>- приемно-контрольное устройство – осуществляет сбор и обработку информации от датчиков;</li> <li>- контрольная панель – обеспечивает управление, установку и настройку всей системы;</li> <li>- исполнительные устройства – элементы пожаротушения;</li> <li>- системы оповещения – светозвуковые или речевые устройства, подающие сигнал тревоги.</li> </ul> <p>Стоимость установки такой системы составляет 331 695 рублей.</p> <p>Прямо сейчас фонду очень нужна ваша поддержка в этом большом деле! Даже самая малая помощь</p>	1140 ₽	331695 ₽	16 раз	9 мая 2017
цинско	Мы собираем деньги на строительство	<p>Дорогие друзья! Спасибо за интерес к нашим проектам и вашу помощь, за отзывчивость и неравнодушие. Благотворительный фонд «Шередарь» уже 4 года помогает детям и подросткам, пережившим онкологическое заболевание. Мы являемся единственной в России некоммерческой организацией, занимающейся развитием детского онкологического центра.</p> <p>Так как безопасность жизни и здоровья детей в реабилитационном центре — главное условие проведения реабилитационных мероприятий. Согласно проекту, в новом медицинском блоке будет приемный кабинет, два инфекционных отделения.</p> <p>В данный момент мы собираем деньги на монтаж охранно-пожарной сигнализации в здание медпункта.</p> <p>С технической точки зрения автоматическая пожарная сигнализация – это довольно сложная система, состоящая из:</p> <ul style="list-style-type: none"> <li>- аппаратура для обнаружения источника возгорания;</li> <li>- устройство речевого оповещения;</li> <li>- системы пожаротушения и дымоудаления;</li> <li>- устройство подачи сигнала на СКУД (система контроля и управления доступом);</li> <li>- сенсорные устройства (датчики, извещатели) – обнаруживают возгорание и передают сигнал на приемно-контрольное устройство;</li> <li>- приемно-контрольное устройство – осуществляет сбор и обработку информации от датчиков;</li> <li>- контрольная панель – обеспечивает управление, установку и настройку всей системы;</li> <li>- исполнительные устройства – элементы пожаротушения;</li> <li>- системы оповещения – светозвуковые или речевые устройства, подающие сигнал тревоги.</li> </ul> <p>Стоимость установки такой системы составляет 331 695 рублей.</p> <p>Прямо сейчас фонду очень нужна ваша поддержка в этом большом деле! Даже самая малая помощь</p>	1140 ₽	331695 ₽	16 раз	9 мая 2017
	Помогать ближнему, кто нуждается в помощи	<p>Помощь людям с распятым детством - пожилым людям, ветеранам, бывшим малолетним узникам фашизма.</p> <p>На видео - скромный домик бывшего малолетнего узника. Этот домик также служит офисом для Благотворительного фонда "Счастье".</p> <p>С 1988 года мы работали за свои средства.</p> <p>Теперь мы просим Вас о помощи.</p> <p>Когда молча просят о помощи, иногда прорываясь рассказать о своих трудностях не в силах сдержаться. И именно в этом малом мы можем помочь ВМЕСТЕ. Для каждого из них непосильная ноша, а если мы объединимся, то сможем помочь.</p> <p>Сейчас мы даже просим обеспечить комфортное нахождение в доме с теплыми стенами и функционирование бытовой техники.</p> <p>А еще все вместе мы можем подарить такие нужные подарки к Новому Году и Рождеству, 11 апреля и 19 декабря.</p> <p>Мы намерены поздравить как можно большее количество бывших узников с праздниками, помочь ветеранам.</p> <p>Для приобретения продуктов питания, теплых вещей и необходимых подарков и организация праздников.</p> <p>Приглашаем присоединиться к сбору и раздаче подарков.</p>	30 ₽	995000 ₽	3 раза	27 января 2017
ore	Современные технологии в помощь детям	<p>Благотворительный фонд "Счастливый мир" был создан для помощи детям с тяжелыми заболеваниями. Во время школьной перемены Денис не спешит выбегать из класса вместе со своими одноклассниками. Сейчас, имея специальный сенсор, Денис справляется с этой задачей за пару минут и без прокола папки.</p> <p>Заболевание обнаружили в 2014 году во время прохождения обычной школьной медкомиссии. С того времени Денис растет в многодетной семье в Кемеровской области. У мальчика два младших братика. Работа фонда направлена на помощь Денису.</p> <p>Мальчику необходимо 26 сенсоров, которых хватит на целый год. Мы открываем сбор для Дениса -</p>	130 ₽	124289 ₽	4 раза	5 декабря 2016
Морозу	Волшебное путешествие детей с ограниченными возможностями	<p>Здравствуйтесь друзья! Меня зовут Ильнар. Я руководитель Благотворительного Фонда " Радуга" г. Казань. Мы помогаем детям с ограниченными возможностями.</p> <p>Так же мы сотрудничаем с казанским детским хосписом. Работа ведется по всей территории Казани и Республики Татарстан.</p> <p>Работа ведется по всей территории Казани и близлежащим районам. За последние полтора года вышло много детей из хосписа.</p> <p>В начале года, в Москве фонд удостоился премии добрых дел "Сможем вместе"- это нематериальный</p>	950 ₽	450000 ₽	8 раз	11 января 2017

# Характеристики

- Время работы кода: 1,5 - 4 часа :(
- 1.730 строк (из 1.975 ожидаемых) без дубликатов
- 772.000 слов в основных текстах

The screenshot shows a web application interface. In the background, there is a text editor with a dark theme and a line number bar at the top (1-17). The text in the editor is partially visible and appears to be a mix of Russian and English. Overlaid on the text editor is a white modal window titled "Статистика" (Statistics) with a close button (X) in the top right corner. The modal window contains a table with the following data:

Страницы	1845
Слов	772207
Количество знаков	5517188
Количество знаков (без пробелов)	4698168

Below the table, there is a checkbox labeled "Показывать статистику слов при вводе текста" (Show word statistics when entering text), which is currently unchecked. At the bottom of the modal window, there are two buttons: "Отмена" (Cancel) and "ОК" (OK).



# Следующие шаги

1. Повысить качество и скорость сбора данных
2. Анализировать тексты

