

Горячее

Лучшее

Свежее



Выпить зелье



Войти



empenoso

7 месяцев назад



Программирование на python



## Задолбался руками адреса сверять? Python to the rescue!

Привет, пикабушники! Недавно столкнулся с задачей, от которой у любого глаза на лоб полезут: нужно было сравнить два огромных списка адресов. Проблема в том, что адреса были записаны как попало, без каких-либо нормальных идентификаторов. Один и тот же адрес мог выглядеть так:

- "д. Малое Шилово, ул. Березовая, д. 7" и "Березовая 7\_М Шилово"
- "п. Ласьва, ул. Весенняя, д. 5" и "Весенняя 5\_Ласьва"
- "Луговой пер 5, Краснокамск г" и "г. Краснокамск, пер. Луговой, 5"
- "д. Новая Ивановка, ул. Солнечная, 18" и "д.Новая Ивановка, ул.Солнечная, 18"



#Круги добра



### Войти

Войти

Создать аккаунт

[Забыли пароль?](#)

или продолжите с



Войти с Яндекс ID



Войти через VK ID



Промокоды



Работа



Курсы



Реклама



Игры



Пополнение Steam

sub...	Вход	Файл	Глав	Встав	Рисо	Разм	Форм	Данн	Рецен	Вид
Буфер обмена	Шрифт	Выравнивание	Число							
A378										
322 г. Краснокамск, ул. Раздольная, 11										
323 д. Семичи, ул. Золотая, 2										
324 д. Никитино, ул. Полевая, 20										
325 с. Мысы, ул. Радужная, 8										
326 п. Ласьева, ул. Снежная, 5										
327 п. Ласьева, ул. Снежная, 12										
328 ДНТ Никитино, ул. Крайняя, 9а										
329 ст. Шабуничи, пер. Полевой, 15										
330 ст. Шабуничи, ул. Тракторная, 3										
331 ст. Шабуничи, ул. 3-я Тракторная, 21а										
332 д. Семичи, ул. Вишневая, 10										
333 г. Краснокамск, ул. Камская, 62										
334 с. Мысы, ул. Рублевская, 45										
335 п. Оверята, пер. Сосновый, 8										
336 д. Большое Шилово, ул. Сюзьвенская, 11										
337 д. Новая Ивановна, ул. Тракторная, 15										
338 д. Мошни, ул. Запрудная, 15										
339 д. Хухрята, ул. Изумрудная, 2										
340 г. Краснокамск, ул. Камская, 60										
341 с. Мысы, проезд Рыбачий, 13										
342 г. Краснокамск, пер. Нагорный, 3а										
343 д. Конец-Бор, пер. Технический, 15										
344 с. Мысы, ул. Попова, 6										
345 п. Ласьева, кв-л Восточный, 15а										
346 д. Брагино, ул. Лесная, 5										
Лист1										

Задача стояла так: **"В реестре поданных объектов отметить все согласованные объекты (из общего списка согласованных)".**

Руками это делать - вообще не вариант. Поэтому решил записать скрипт на Python. Какие есть варианты?

1. **Fuzzy matching (нечеткое совпадение):** алгоритмы, которые сравнивают строки, учитывая опечатки, разный порядок слов и сокращения.
2. **Геокодинг:** преобразование текстового адреса в координаты.

**Fuzzy matching** мне показался более подходящим. Он не требует, чтобы адрес был на карте (а вдруг объект еще строится?), и хорошо справляется с разными вариантами написания.


Реклама

Вернуть любой ценс

Военный, Драма  
сериал

СМОТРЕТЬ

**Пикабу Игры**  
+1000 бесплатных  
онлайн игр



**Хитбол**  
Аркады, Казуальные, Для  
мальчиков

Играть

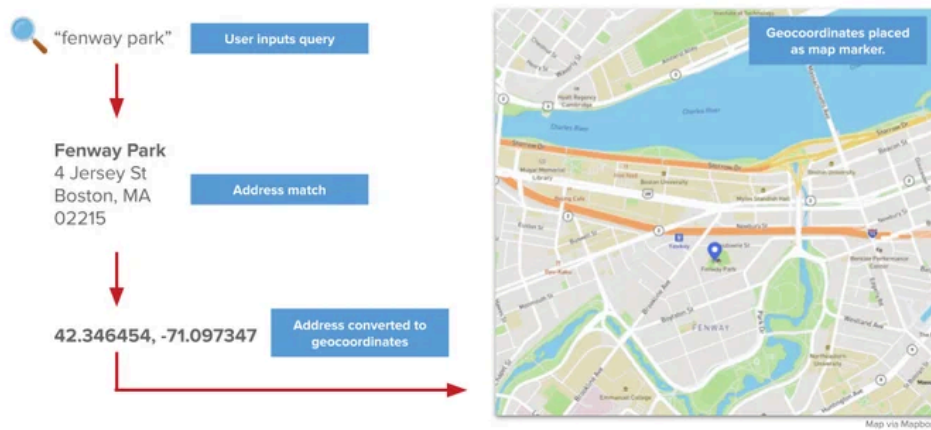
Name	Date of birth
Sophia Gale	22-11-1994
Jane Doe	08-03-1987
Jonathan Durward	02-05-1972
Andrew Bright	30-10-1955
Jan Deo	08-03-1987
Janine Walker	11-05-1986
Janne Doe	08-03-1987
Fred Warden	09-12-1978

GIF • 100 K6

картинка не совсем в тему, но прикольная, про fuzzy logic

**Геокодирование**, конечно, тоже можно использовать, но он может быть неточным, особенно в деревнях. Да и не все адреса на картах есть.

## Geocoding



## Как я это сделал:

- Подготовка данных:** Сначала нужно привести адреса к единому формату. Убрать лишние пробелы, точки, запятые, сокращения типа "д.", "ул.", "г.". Для этого использовал Python с библиотеками pandas, openpyxl и fuzzywuzzy. (pip install pandas openpyxl fuzzywuzzy)
- Fuzzywuzzy magic:** Библиотека fuzzywuzzy использует алгоритм Левенштейна, чтобы определить, насколько строки похожи. Я использовал fuzz.token\_sort\_ratio, которая сортирует слова по алфавиту перед сравнением, чтобы порядок слов не мешал. Еще добавил фильтрацию по цифрам в адресе, чтобы ускорить процесс.

Реклама

Я тебя любл

Музыкальный

фильм

СМОТРЕ

## Топ прошлой недели

- SpongeGod  
1 пост
- Uncleyogurt007  
9 постов
- ZaTaS  
3 поста

[Посмотреть весь топ](#)

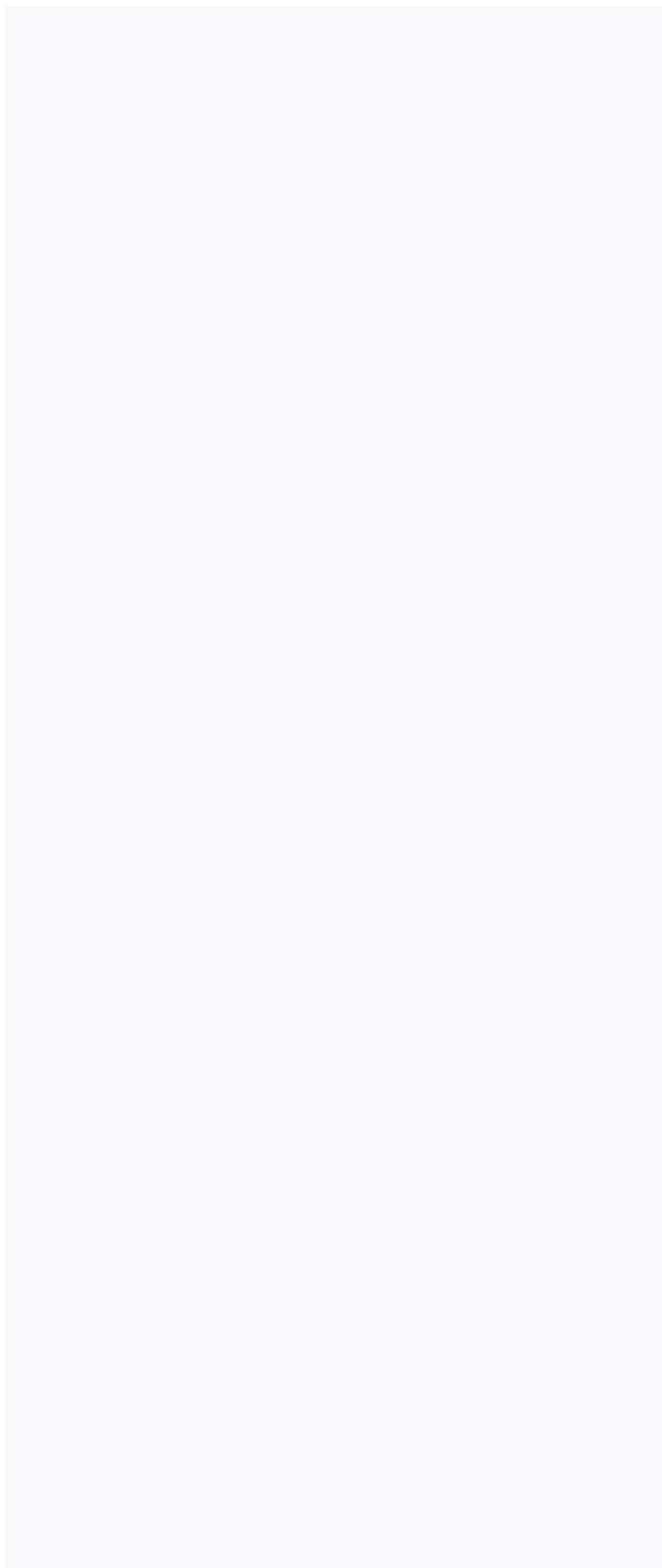
## Лучшие посты недели

Рассылка Пикабу:  
отправляем самые  
рейтинговые материалы за 7  
дней 🔥

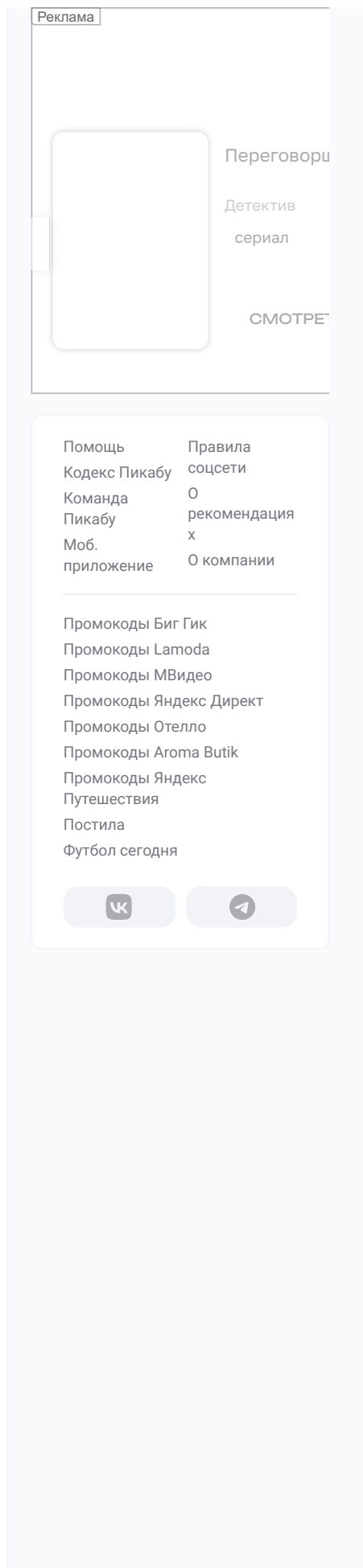
Укажи

Подписаться

Нажимая кнопку  
«Подписаться на рассылку»,  
я соглашаюсь с [Правилами](#)  
[Пикабу](#) и даю согласие на  
обработку персональных  
данных.



**Скрипт:** Скрипт загружает данные из Excel, чистит адреса, ищет совпадения с помощью fuzzywuzzy, помечает согласованные объекты плюсиком "+", а несогласованные - крестиком "X", и сохраняет результат в новый файл.



**Profit!** Автоматизация сэкономила кучу времени и нервов. Скрипт легко адаптировать под другие задачи, где нужно сравнивать текст.

#### Что можно улучшить:

- Комбинировать fuzzy matching с геокодингом для большей точности.
- Добавить обработку большего количества сокращений и вариантов написания.

**Вопросы? Предложения? Пишите в комментариях!**

Автор: [Михаил Шардин](#),

20 января 2025 г.



15

11K



РЕКЛАМА



vodohod-perm.ru

**Модный тренд — речные прогулки по Каме на «Экоходъ»**

4,2 ★ Рейтинг организации ⓘ

[Подробнее](#)



РЕКЛАМА



русуралстрой59.рф

**Строительство теплого дома под ключ в Перми**

[Подробнее](#)



**Программирование на python**

885 постов • 11.9K подписчиков

[Добавить пост](#)

[Подписаться](#)



#### Правила сообщества

Публиковать могут пользователи с любым рейтингом. Однако!

Приветствуется:...

[Подробнее](#) ✓

[Все комментарии](#)   [Автора](#)

Раскрыть 15 комментариев

Чтобы оставить комментарий, необходимо [зарегистрироваться](#) или [войти](#)

Реклама

Вернуть любой ценой

Военный, Драма  
сериал

СМОТРЕТЬ

18+

Реклама

Жизнь по вызову. Док

Документальный  
фильм

СМОТРЕТЬ

18+

РЕКЛАМА

Будущее начинается

Построй карьеру  
в лучшей ОЭЗ  
в России

РЕКЛАМА - 16+

Поддай заявку в IT-аг  
селератор Сбера  
и запусти стартап!

● — ■ —

—

—

—

—

—

—

● — ■ —

—

—

—

—

—

—

● — ■ —

—

—

—

—

—

—

