



academyit.ru РЕКЛАМА · 16+

## Обучение MBA от бизнес-лидеров трансформации

Создайте стратегию цифровой трансформации для ва

[Перейти на сайт](#)

Горячее Лучшее Свежее Подписки Сообщества

empenoso 2 месяца 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. Ласьева, ул. Весенняя, д. 5" и "Весенняя 5\_Ласьева"
- "Луговой пер 5, Краснокамск г" и "г. Краснокамск, пер. Луговой, 5"
- "д. Новая Ивановка, ул. Солнечная, 18" и "д.Новая Ивановка, ул.Солнечная, 18"

sub...	Вход
Файл Главная Вставка Рисование Разметка Формы Данные Рецензирование Вид	Файл Главная Вставка Рисование Разметка Формы Данные Рецензирование Вид
Буфер обмена Шрифт Выравнивание Число	Буфер обмена Шрифт Выравнивание Число
A378	D86
A	A
322 г. Краснокамск, ул. Раздольная, 11	118 с. Стряпунята, ул. Набережная, 4а
323 д. Семичи, ул. Золотая, 2	119 г. Краснокамск, пер. Луговой, 5А
324 д. Никитино, ул. Полевая, 20	120 г. Краснокамск, пер. Свободный, 7
325 с. Мысы, ул. Радужная, 8	121 г. Краснокамск, пер. Черный, 3
326 п. Ласьева, ул. Снежная, 5	122 г. Краснокамск, ул. Осинская, 8
327 п. Ласьева, ул. Снежная, 12	123 д. Карабаи, ул. Полевая, 37
328 ДНТ Никитино, ул. Крайняя, 9а	124 д. Большое Шилово, ул. Мирная, 16
329 ст. Шабуничи, пер. Полевой, 15	125 д. Большое Шилово, ул. Сюзьвенская, 19/2
330 ст. Шабуничи, ул. Тракторная, 3	126 д. Большое Шилово, ул. Сюзьвенская, 27
331 ст. Шабуничи, ул. 3-я Тракторная, 21а	127 д. Волеги, ул. Дорожная, 1
332 д. Семичи, ул. Вишневая, 10	128 д. Конец-Бор, ул. Конец-Борская, 2Б
333 г. Краснокамск, ул. Камская, 62	129 д. Конец-Бор, ул. Некрасова, 18А
334 с. Мысы, ул. Рублевская, 45	130 д. Конец-Бор, ул. Трудовая, 51
335 п. Оверята, пер. Сосновый, 8	131 д. Малое Шилово, ул. Дачная, 4
336 д. Большое Шилово, ул. Сюзьвенская, 11	132 д. Нижние Симонята, ул. Набережная, 14а
337 д. Новая Ивановна, ул. Тракторная, 15	133 д. Новая Ивановка ул. Железнодорожная, 5
338 д. Мошни, ул. Запрудная, 15	134 д. Новая Ивановка ул. Зеленая, 27-2
339 д. Хухрята, ул. Изумрудная, 2	135 д. Семичи, ул. 1-я Подгорная, 2
340 г. Краснокамск, ул. Камская, 60	136 д. Семичи, ул. 1-я Подгорная, 22
341 с. Мысы, проезд Рыбацкий, 13	137 д. Семичи, ул. Виноградная, 6
342 г. Краснокамск, пер. Нагорный, 3а	138 д. Семичи, ул. Земляничная, 2
343 д. Конец-Бор, пер. Технический, 15	139 д. Семичи, ул. Молодежная, 11
344 с. Мысы, ул. Попова, 6	140 д. Семичи, ул. Молодежная, 20
345 п. Ласьева, кв-л Восточный, 15а	141 д. Семичи, ул. Светлая, 21
346 д. Брагино, ул. Лесная, 5	142 тер. ДНП Южные Мысы, ул. Лучистая, 37
Лист1	Лист1
Параметры отображения	Параметры отображения

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

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

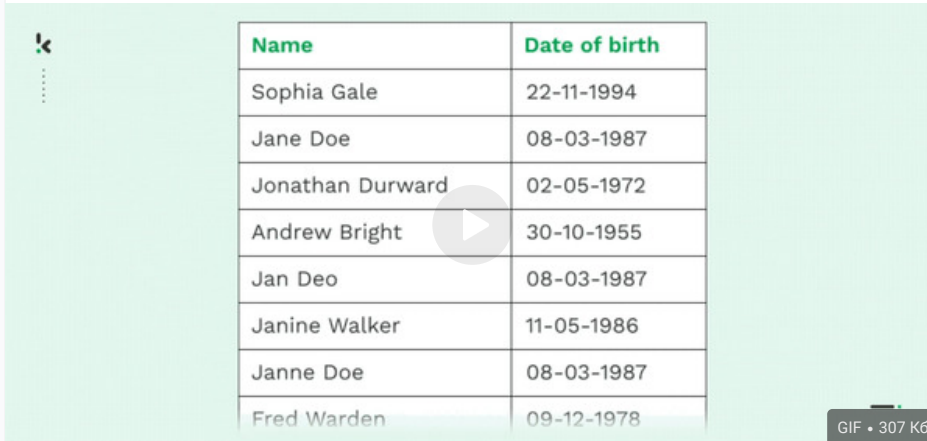
### Используйте аккаунт Яндекса для входа на сервис

Безопасный вход без дополнительной регистрации на сайте

[Войти с Яндекс ID](#)[Войти](#)[Создать аккаунт](#)[Забыли пароль?](#)[или продолжите с](#)[Войти с Яндекс ID](#)[Войти через VK ID](#)[Промокоды](#)[Работа](#)[Курсы](#)[Реклама](#)[Игры](#)[Пополнение Steam](#)

1. **Fuzzy matching (нечеткое совпадение):** алгоритмы, которые сравнивают строки, учитывая опечатки, разный порядок слов и сокращения.
2. **Геокодирование:** 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.

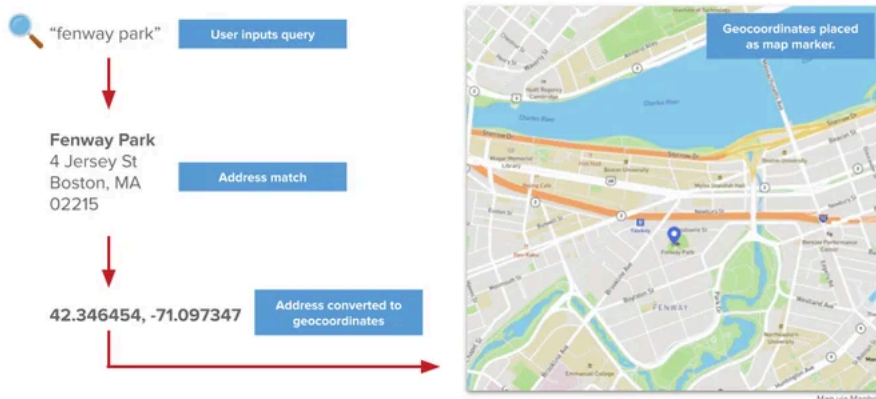


Name	Date of birth
Sophia Gale	22-11-1994
Jane Doe	08-03-1987
Jonathan Durward	02-05-1972
Andrew Bright	30-10-1955
Jan Deo	08-03-1987
Janine Walker	11-05-1986
Janne Doe	08-03-1987
Fred Warden	09-12-1978

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

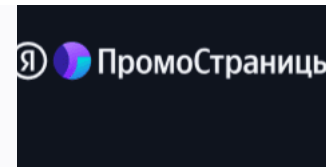
**Геокодирование**, конечно, тоже можно использовать, но он может быть неточным, особенно в деревнях. Да и не все адреса на картах есть.

## Geocoding



Как я это сделал:

1. **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`)
2. **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.



### Топ прошлой недели

- Kerbis  
2 поста
- MaximVeter  
6 постов
- alex.carrier  
16 постов

[Посмотреть весь топ](#)



### Маркетплейс

для реального сектора экономики

industrial.market

**Платформа для закупок  
без тендеров  
и согласований**

[Узнать больше](#)

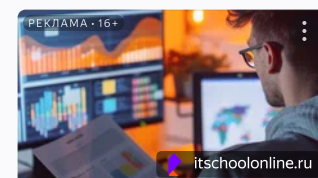
### Лучшие посты недели

Рассылка Пикабу:  
отправляем самые  
рейтинговые материалы за 7  
дней 🔥

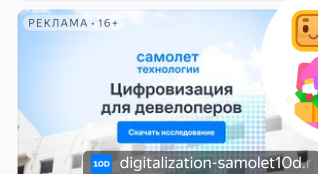
Укажи

[Подписаться](#)

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



**Курс - специалист  
по анализу данных  
от Ростелеком**



Скачайте новое исследование о Цифровизации



ции стройки

Новости Пикабу Команда Пикабу  
Помощь Награды  
Кодекс Пикабу Контакты  
Реклама О проекте  
О компании Зал славы

Промокоды Купоны Мегамаркет  
Скидки Купоны Тефаль  
Работа Купоны М.Видео  
Курсы Купоны YandexTravel  
Блоги Купоны Lamoda

Мобильное приложение



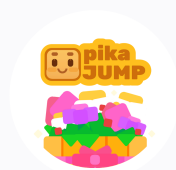
РЕКЛАМА promo.dav-changanauto.ru



**2 719 900 ₽**  
UNI-V Changan купить



**1 949 900 ₽**  
CS35PLUS NEW Changan ку



**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 комментарии!**

Автор: [Михаил Шардин](#),

20 января 2025 г.



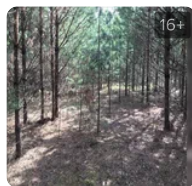
15 10K

₽ Поддержать

Эмоции

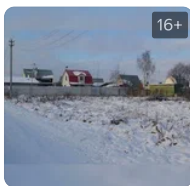
РЕКЛАМА · 16+

portal-da.ru



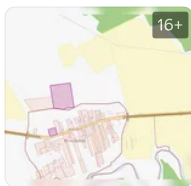
298 000 ₽

Земельный участок,  
10 сот



813 000 ₽

Земельный участок,  
6 сот



1 500 000 ₽

Земельный участок  
1/2 доля в вашем  
доме или офисе



4 200 000 ₽

Земля с/х  
назначения, 379 га  
Омская область



Программирование на python

822 поста · 11.9К подписчика

Добавить пост

Подписаться



**Правила сообщества**

Публиковать могут пользователи с любым рейтингом. Однако!

Приветствуется...

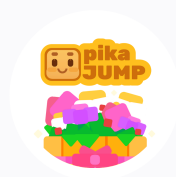
[Подробнее](#) ▾

[Все комментарии](#)

[Автора](#)

Раскрыть 15 комментариев

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



\_\_\_\_\_

● \_\_\_\_\_ ■ \_\_\_\_\_

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

● \_\_\_\_\_ ■ \_\_\_\_\_

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

