



empenoso 3 дня назад



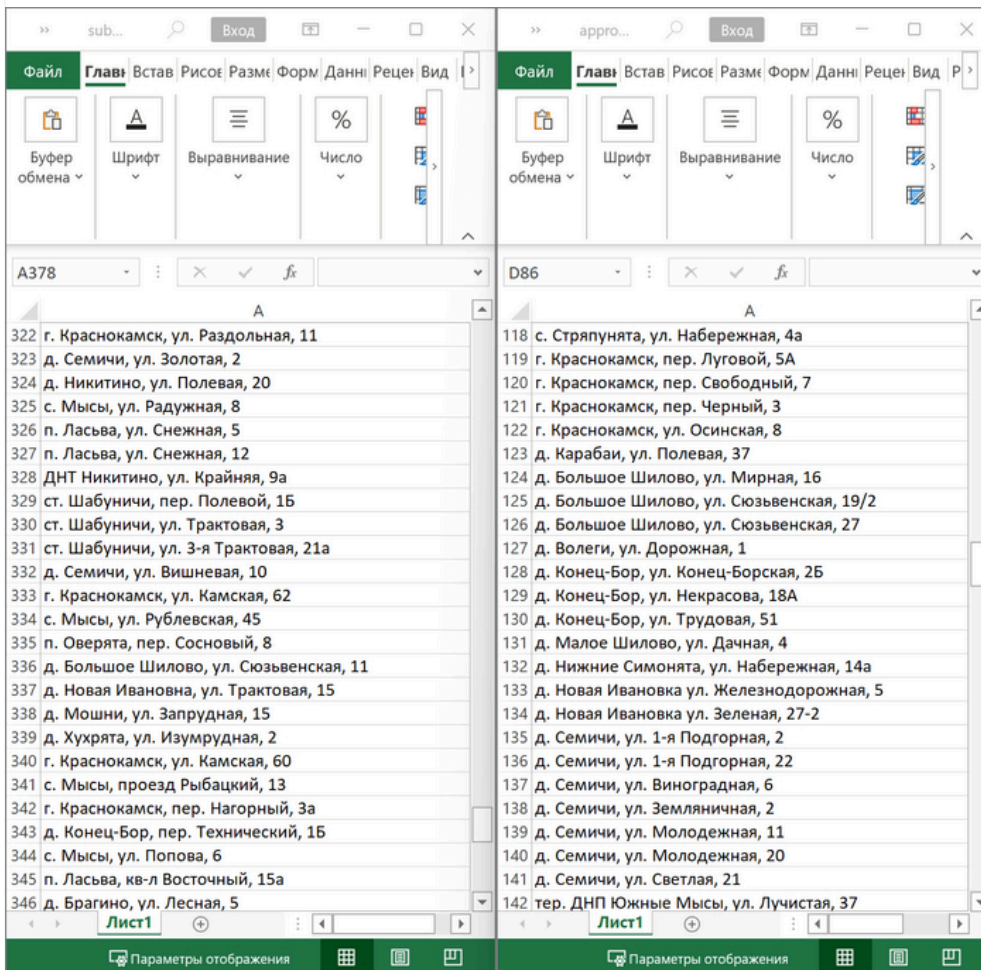
Программирование на python



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"



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?

Войти

Войти

Создать аккаунт

Забыли пароль?

или продолжите с



Войти с Яндекс ID



Войти через VK ID



Промокоды



Работа



Курсы



Реклама



Игры



Пополнение Steam



О сообществе сообщества



Программирование...

749 постов • 11К п...

Подписаться

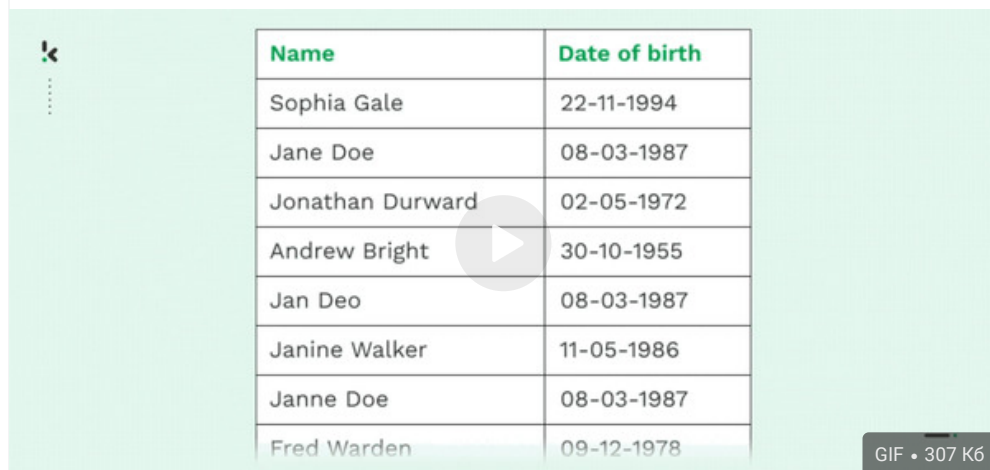


Сообщество любителей языка программирования python и просто неравнодушных к нему. Обучение программированию на языке python. Исследование его интересных возможностей и библиотек.

[Правила сообщества](#)

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

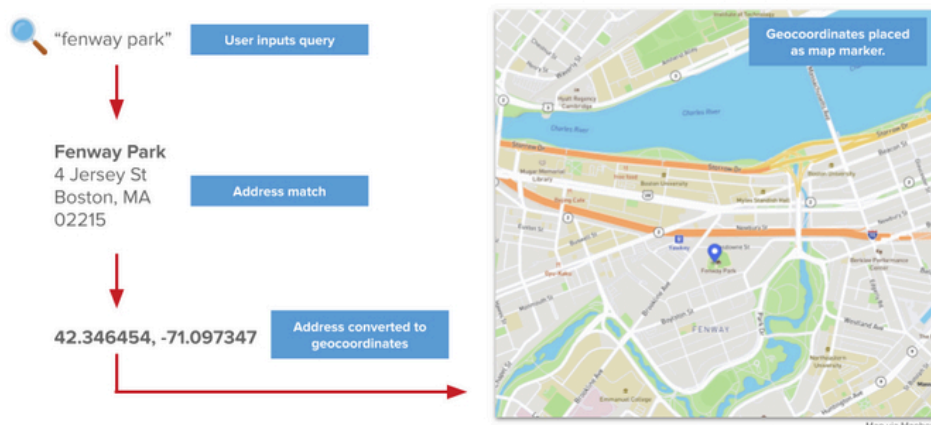
Fuzzy matching мне показался более подходящим. Он не требует, чтобы адрес был на карте (а вдруг объект еще строится?), и хорошо справляется с разными вариантами написания.



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

Геокодинг, 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.

Управление сообществом

rissy Администратор

Популярные авторы

PythonGet

Топ прошлой недели

- Kelebra96
5 постов
- Valaybalalay
22 поста
- Pervuhin
12 постов

[Посмотреть весь топ](#)

**Лучшие посты
недели**

Рассылка Пикабу:
отправляем самые
рейтинговые материалы за 7
дней 🔥

Укажите

Подписаться

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

[Новости Пикабу](#) [Команда Пикабу](#)[Помощь](#) [Награды](#)[Кодекс Пикабу](#) [Контакты](#)[Реклама](#) [О проекте](#)[О компании](#) [Зал славы](#)[Промокоды](#) [Купоны Мегамаркет](#)[Скидки](#) [Купоны Спортмастер](#)[Работа](#) [Купоны М.Видео](#)[Курсы](#) [Купоны Aliexpress](#)[Блоги](#) [Купоны Lamoda](#)

Мобильное приложение



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 9.5K



Поддержать

Эмоции

**Программирование на python**

749 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](#)

.....

.....

.....

.....

.....

.....



.....

.....

.....

.....

.....

.....



.....

.....

.....

.....

.....

