



Big Data na rynku nieruchomości.

Księgi Wieczyste

GUS

Archiwa sądowe

Operaty szacunkowe

RCiWN

CRCN

Archiwa urzędowe

Geoportal

Rozdział 1: Księga Wieczysta – fantasy, sci-fi czy horror dla inwestora?

- 1.1. Księga, która mówi prawdę (a przynajmniej powinna)
- 1.2. Historia KW, czyli jak pruski urzędnik zrobił nam przysługę
- 1.3. Co się mieści w środku – poznaj działy od I do IV
- 1.4. Dlaczego KW to złoto dla inwestora (i nie tylko dla niego)
- 1.5. Jak nie zginąć w prawnym gąszczu – słownik po ludzku

Rozdział 2: Co w KW piszczy – czyli jak czytać księgę i nie oszaleć

- 2.1. Dział I – Co tu jest i dlaczego każdy go chce?
- 2.2. Dział II – Kto jest właścicielem? I czy na pewno?
- 2.3. Dział III – Kto zajął nieruchomość, a kto ją obciążył?
- 2.4. Dział IV – Hipoteki, czyli prawne bomby zegarowe
- 2.5. Jak łączyć dane z działów? Logika i pułapki

Rozdział 3: Geoportal – mapy i dane geodezyjne w praktyce

- 3.1. Scrapowanie danych z Geoportalu – poradnik krok po kroku
- 3.2. Automatyzacja przetwarzania danych geodezyjnych – od surowych danych do wartościowych analiz
- 3.3. Integracja danych geodezyjnych z systemami GIS i platformami inwestycyjnymi

Rozdział 4: Rejestr Cen i Wartości Nieruchomości – ile naprawdę jest warte złoto?

- 4.1.1. Rejestr Cen i Wartości Nieruchomości (RCiWN) – oficjalny rejestr prowadzony przez starostwa powiatowe. Pobranie wszystkich danych
- 4.1.2. Centralny Rejestr Cen Nieruchomości (CRCN) – prowadzony przez Główny Urząd Geodezji i Kartografii (GUGiK). Pobranie wszystkich danych

Rozdział 5: Archiwa sądowe i papierowe cuda – kiedy internet nie wystarcza

- 5.1. Wydziały Ksiąg Wieczystych offline – jak odzyskać dane, których nie da się kliknąć
- 5.2. Operaty techniczne i mapy analogowe – kiedy działka ma więcej warstw niż cebula
 - 5.2.1. Źródła danych i ich formaty
 - 5.2.2. ETL (Extract, Transform, Load)
 - 5.2.3. Georeferencja map analogowych
 - 5.2.4. Integracja danych operatu technicznego
 - 5.2.5. Warstwa wyceny – agregacja operatów szacunkowych
 - 5.2.6. Interfejs użytkownika dla inwestora i prawnika
 - 5.2.7. Gdzie i jak uzyskać dostęp do operatów technicznych i map w Polsce
 - 5.2.8. Przykłady fragmentów kodu dla integracji operatów i map
- 5.3. Archiwa urzędów gminy i miast – nieoczywiste źródła oczywistych danych
 - 5.3.1. Pozyskiwanie dokumentów z archiwów urzędowych
 - 5.3.2. Ekstrakcja tekstu z dokumentów PDF i skanów (OCR)
 - 5.3.3. Normalizacja i indeksacja danych

5.3.4. Budowanie bazy danych dokumentów archiwalnych (ETL)

5.3.5. Agregacja i analiza danych historycznych

5.3.6. Wizualizacja dokumentów i powiązań

5.3.7. Automatyczne pobieranie i harmonogram pracy

5.4. Detektyw terenowy – jak ręcznie pozyskać dane, których nikt nie wpisał

5.4.1. Wstęp: Co robi inwestor z lupą?

5.4.2. Czym są dane „niewpisane”?

5.4.3. Detektyw terenowy – narzędzia analogowe i cyfrowe

5.4.4. Proces ręcznego pozyskiwania danych – krok po kroku

5.4.5. Dokumentacja terenowa – jak ją gromadzić

5.4.6. Jak przekształcić dane terenowe w twarde argumenty?

5.4.7. Zastosowanie praktyczne – case study

5.4.8. Czego unikać? Czego nie wolno?

5.4.9. Humor z pola

5.4.10. Podsumowanie

5.5. Scrapowanie operatów szacunkowych online

5.5.1. Wprowadzenie do scrapowania danych z operatów szacunkowych

5.5.2. Przygotowanie środowiska i narzędzi do scrapowania

5.5.3. Techniki scrapowania operatów szacunkowych

5.5.3.1. Pobieranie operatów – requests i Selenium w akcji

5.5.3.2. Przetwarzanie PDF – tekst vs skan, czyli Tesseract i pdfplumber na scenie

5.5.3.3. Parsowanie tekstu i wyciąganie kluczowych danych z operatów

5.5.3.4. Ekstrakcja tabel i organizacja danych w Pandas

5.5.3.5. Kompletny pipeline – od pobrania do pliku CSV

5.5.3.6. Bonus: Praca z CAPTCHA i automatyczne logowanie (krótki wstęp)

5.5.4. Przetwarzanie i strukturyzacja pozyskanych danych

5.5.4.1. Ekstrakcja kluczowych informacji z operatów

5.5.4.2. Normalizacja i walidacja danych

5.5.4.3. Integracja z bazami danych i systemami analitycznymi

5.5.4.4. Prosty przykład zapisu do SQLite

5.5.4.5. Automatyzacja całego procesu

5.5.4.6. Podsumowanie

Rozdział 6: Techniczna rozgrzewka – narzędzia, które sprawią, że komputer zacznie pracować dla Ciebie

6.1. Narzędzia pierwszej potrzeby – podstawowy zestaw poszukiwacza danych

6.1.1. Python – język magii i pragmatyzmu

6.1.2. Biblioteka Requests – rozmowy z internetem

6.1.3. BeautifulSoup – czarodziej do wyciągania danych z HTML

6.1.4. Selenium – kiedy strony są kapryśne i dynamiczne

6.1.5. Pandas – do porządkowania i analizy danych

6.2. Łatwe pobieranie i czytanie stron – jak odczytać treść bez zbędnego szumu

6.2.1. Proste żądania HTTP

6.2.2. Analiza struktury HTML

6.2.3. Filtrowanie niepotrzebnych elementów

- 6.2.4. Obsługa formatów tekstowych i plików
- 6.2.5. Zapisywanie danych w formacie przyjaznym do analizy

6.3. Radzenie sobie ze stronami, które się nie chcą pokazać

- 6.3.1. Strony z dynamiczną zawartością (JavaScript)
- 6.3.2. Problemy z CAPTCHA
- 6.3.3. Blokady IP i limitacje żądań
- 6.3.4. Proxy i rotacja adresów IP
- 6.3.5. Analiza i odczyt plików cookies oraz sesji

6.4. Kamuflaż w sieci – jak nie dać się złapać i nie zostać zbanowanym

- 6.4.1. Zmiana user-agentów
- 6.4.2. Symulacja zachowań użytkownika
- 6.4.3. Odpowiednie opóźnienia między żądaniami
- 6.4.4. Obsługa ciasteczek i sesji
- 6.4.5. Monitorowanie i reagowanie na błędy serwera

6.5. Pierwsze polecenia dla Twojego cyfrowego pomocnika

- 6.5.1. Proste pobieranie strony z Requests
- 6.5.2. Wyciąganie danych z HTML za pomocą BeautifulSoup
- 6.5.3. Obsługa błędów i wyjątków
- 6.5.4. Zapis do pliku i prosta analiza
- 6.5.5. Uruchomienie prostego pętlowego scrapera

Rozdział 7: Jak nie dać się zbanować – etyczne i techniczne triki scraperów

- 7.1. Limity zapytań – dlaczego warto nie przesadzać i jak to kontrolować
- 7.2. User-Agent i inne maski – jak się stroić, żeby nie wyglądać jak bot
- 7.3. Sleep – czyli dlaczego czasem warto zrobić sobie przerwę
- 7.4. Rotacja IP – zmieniaj adres jak kameleon
- 7.5. CAPTCHA i jak jej (legalnie) unikać lub pokonać

Rozdział 8: Historia działek – jak czytać przeszłość z map i ksiąg

- 8.1. Mapy historyczne i ich znaczenie w badaniu działek
- 8.2. Jak czytać zmiany w numeracji działek na przestrzeni lat
- 8.3. Powiązania między działką a księgą wieczystą – praktyczne przykłady
- 8.4. Narzędzia cyfrowe wspomagające śledzenie historii działek
- 8.5. Przypadki z życia – jak działki zdradzają swoją przeszłość

Rozdział 9: Po nitce do kłębka – wyciąganie danych z GUS

- 9.1. Fundamenty danych GUS – od czego zacząć i czego szukać?
- 9.2. Wzbogacanie księgi wieczystej – integracja danych z GUS i Geoportalu
- 9.3. Ceny transakcyjne i statystyki rynkowe – co mówi GUS?
- 9.4. MPZP i planowanie przestrzenne w danych GUS i Geoportalu
- 9.5. Automatyzacja pobierania i analizy danych GUS – narzędzia i praktyczne przykłady

Rozdział 10: Wyciąganie danych z RCIWN – co zapłacono naprawdę?

- 10.1. Co to jest RCiWN i dlaczego warto mu zaufać?
- 10.2. Techniki pobierania i interpretacji danych z RCiWN
- 10.3. Weryfikacja cen z KW i RCiWN – gdzie leży prawda?
- 10.4. Trendy rynkowe i analizy statystyczne na podstawie danych RCiWN
- 10.5. Integracja danych RCiWN z innymi rejestrami – pełny obraz rynku nieruchomości

Rozdział 11. MPZP i planowanie przestrzenne – co miasto planuje na jutro?

- 11.1. MPZP – Biblia planisty, horror spekulanta
- 11.2. Jak czytać MPZP bez zasypiania – legenda, symbole i tajemnice oznaczeń
- 11.3. Łączenie MPZP z KW – czyli czy Twoja działka ma prawo marzyć o wieżowcu
- 11.4. Gdzie znaleźć MPZP i jak sprawdzić, co planuje urząd? (i kiedy powiedzą Ci „za wcześnie”)
- 11.5. Planowane drogi, linie kolejowe, strefy zalewowe i... cmentarze

Rozdział 10. Księgi wieczyste

- 10.1. Ministerstwo Sprawiedliwości kontra Skrypty – gdzie klikamy, a gdzie kopiemy
- 10.2. HTML, tagi i ukryte pola – kiedy księga mówi szeptem
- 10.3. Cookies, sesje i inne przysmaki – jak nie zgubić ciasteczka po drodze
- 10.4. Captcha – elektroniczna straż graniczna
- 10.5. Jak przechytrzyć JavaScript, nie łamiąc prawa (ani moralności)
- 10.6. Czy scraper potrzebuje snu? – o harmonogramowaniu i limitach
- 10.7. Python mówi: BeautifulSoup czy Scrapy? – bitwa bibliotek
- 10.8. Regexy i inne zaklęcia – parsowanie treści jak czarodziej danych
- 10.9. „Brak danych” to też informacja – jak nie wpaść w panikę
- 10.10. Scraping vs API, czyli walka Dawida z... brakiem API
- 10.11. Kiedy numer księgi to za mało – jak znaleźć, nie znając adresu
- 10.12. Gdzie mieszkają hipotekarze? – analiza wzorców obciążeń
- 10.13. Ziemia niczyja? – wykrywanie nieaktualizowanych KW
- 10.14. Skrapowanie baz danych dział po dziale
- 10.15. Scraper detektywem – budowanie historii właścicieli na podstawie Księgi Wieczystej
- 10.16. Zbiory, bazy, relacje – jak nie utopić się w danych
- 10.17. Scraper na sterydach – wielowątkowość i proxy w akcji
- 10.18. Jak łączyć dane z różnych źródeł? Sztuka detektywa
- 10.19. Bezpieczeństwo przede wszystkim – jak nie zostać banowanym (ani pozwany)
- 10.20. Czy to legalne, panie władzo? – podstawy prawa dla web-researchera
- 10.21. Zbuduj swój własny serwis poza UE – jak ten, z którego ostatnio kupiłeś numery KW. Koszt? Mniej niż myślisz.