

Dokumentacja Projektowa Systemu Analitycznego GPW

Autor: Jakub Karol Żakowski
Kierunek: Informatyka Ogólna
Uczelnia: Uniwersytet Warmińsko-Mazurski w Olsztynie
Data: 8 czerwca 2025
Spis Treści

Dokumentacja Projektowa Systemu Analitycznego GPW.....	1
Spis Treści.....	1
1. Wprowadzenie i Cel Projektu.....	1
2. Opis Systemu.....	2
2.1. Główne Funkcjonalności.....	2
2.2. Architektura i Technologie.....	3
2.3. Planowany Rozwój Systemu.....	3
3. Analiza Systemu i Przypadki Użycia.....	3
3.1. Aktorzy Systemu.....	3
3.2. Diagram Przypadków Użycia.....	4
3.3. Opis Przypadków Użycia.....	4
4. Scenariusze Użycia.....	5
4.1. Scenariusz Główny: Kompleksowa Analiza Spółki.....	5
4.2. Scenariusz Alternatywny: Brak Raportów dla Wskazanej Spółki.....	5
5. Projekt Bazy Danych.....	5
5.1. Aktualny Opis Tabel (MySQL).....	5
5.2. Opis Relacji i Tabel.....	6
5.3. Proponowane Rozszerzenia Bazy Danych.....	7

1. Wprowadzenie i Cel Projektu

Celem projektu jest stworzenie zaawansowanego narzędzia analitycznego, które zautomatyzuje proces pozyskiwania, przetwarzania i analizy danych z Giełdy Papierów Wartościowych (GPW) w Warszawie. System, nazwany roboczo "GPW Scraping Tool", ma na celu wsparcie analityków finansowych oraz inwestorów indywidualnych w procesie podejmowania decyzji inwestycyjnych.

Głównym problemem, który system ma rozwiązać, jest ogromna ilość danych w formie raportów okresowych i bieżących publikowanych przez spółki giełdowe. Ręczna analiza tych dokumentów jest czasochłonna i nieefektywna. Aplikacja, poprzez automatyczny scraping danych oraz ich inteligentną analizę z wykorzystaniem modeli językowych, dostarcza

użytkownikom skondensowane, kluczowe informacje, znacząco skracając czas potrzebny na ocenę sytuacji finansowej i rynkowej wybranej spółki.

2. Opis Systemu

Aplikacja jest systemem działającym w przeglądarce z interfejsem graficznym, który umożliwia interakcję z danymi giełdowymi w sposób intuicyjny. Użytkownik (analityk) może filtrować i pobierać raporty giełdowe według określonych kryteriów, a następnie poddawać je automatycznej analizie i sumaryzacji.

Company Name
What company report would you like to check?

Report amount
How many reports would you like to take?

☒ Download the CSV report?

Download
Some reports files may be heavy so downloading may take a while
☒ PDF ☐ HTML

Date (optional)
From which day you would like to check reports? (dd-mm-yyyy). If empty all dates will be taken.

Report type
What type of report would you like to take?
☒ current ☒ semi-annual ☒ quarterly ☒ interim
☒ annual

Report category
What category of report would you like to take?
☒ EBI ☒ ESPI

Company Name: not specified
Report amount: 20
Download report types: not specified
Report date: not specified
Report type: current semi-annual quarterly interim annual
Report category: EBI ESPI

Company Name: grupa kęty
Report amount: 8
Download report types: HTML PDF
Report date: not specified
Report type: current quarterly semi-annual annual interim
Report category: EBI ESPI

Scraped data

date	title	report type	report category	exchange rate
12-01-2025 09:56:51	Zawarcie umowy kredytu	Bieżący	ESPI	3.62
10-01-2025 23:29:02	Projekty uchwał na Nadzwyczajne Walne Zgromadzenie "MERCOR" S.A. zwołane na dzień 6 lutego 2025 roku	Bieżący	ESPI	-5.61
10-01-2025 23:18:36	Zwołanie Nadzwyczajnego Walnego Zgromadzenia "MERCOR" S.A. na dzień 6 lutego 2025 roku	Bieżący	ESPI	-5.61
10-01-2025 22:31:42	Oddalenie wniosku o zabezpieczenie roszczenia	Bieżący	ESPI	3.5
10-01-2025 21:49:10	Otrzymanie zawiadomienia w trybie art. 69 Ustawy o ofercie publicznej	Bieżący	ESPI	-0.22

Run

Output
SUCCESS! downloaded 3 files | CSV file saved
files saved in:
E:\PyCharm Projekty\Motokomando_800_czw\REPORTS\

report 2 file 1 Projekty Uchwał NWZA_2025-02-06.pdf

liczba stron: 18

Oto odpowiedź na Twoje pytanie w języku polskim:

§ 4. Liczba akcji podlegających Programowi oraz cena nabycia akcji

1. Programem objęte będą w pełni opłacone akcje Spółki, będące przedmiotem obrotu i notowane na rynku regulowanym, prowadzonym przez Giełdę Papierów Wartościowych w Warszawie S.A. z siedzibą w Warszawie (dalej: „GPW”), oznaczone kodem ISIN: PLMRCOR00016.

2. Łączna kwota przeznaczona na nabycie akcji własnych wynosi: 10 000 000,00 zł (słownie: dziesięć milionów złotych 00/100).

3. Nabycie akcji własnych będzie finansowane ze środków własnych Spółki, z wcześniej utworzonego kapitału rezerwowego z kwoty, która zgodnie z art. 348 § 1 k.s.h. może być przeznaczona do podziału między Akcjonariuszami.

4. Przedmiotem odpłatnego nabycia w ramach Programu będzie nie więcej niż: 778.463 akcji (siedemset siedemdziesiąt osiem tysięcy czterysta sześćdziesiąt trzy), o wartości nominalnej: 0,25 zł (zero złotych 25/100) każda, stanowiących nie więcej niż 5,00 % kapitału.

[Zrzut ekranu interfejsu użytkownika aplikacji]

2.1. Główne Funkcjonalności

- **Scraping Danych:** Automatyczne pobieranie raportów (bieżących, okresowych, kwartalnych, rocznych) dla wybranych spółek bezpośrednio z oficjalnych źródeł GPW (systemy ESPI i EBI).
- **Filtrowanie Danych:** Możliwość precyzyjnego filtrowania raportów na podstawie nazwy spółki, daty publikacji, typu oraz kategorii raportu.
- **Eksport Danych:** Zapisywanie pobranych metadanych raportów do pliku CSV oraz

pobieranie pełnych raportów w formatach PDF i HTML.

- **Sumaryzacja AI:** Generowanie zwięzłych podsumowań treści raportów PDF przy użyciu dużego modelu językowego (LLM).
- **Zarządzanie Historią:** Zapisywanie historii wyszukiwań i operacji w dedykowanej bazie danych.

2.2. Architektura i Technologie

- **Język Programowania:** Python 3.12
- **Interfejs Użytkownika:** Biblioteka Gradio
- **Scraping:** Biblioteki takie jak requests, BeautifulSoup.
- **Przetwarzanie PDF:** Biblioteki do ekstrakcji tekstu z plików PDF.
- **Analiza Językowa:** Lokalnie uruchomiony model **Llama 3.2** za pośrednictwem platformy **Ollama**.
- **Baza Danych:** MySQL / MariaDB.

2.3. Planowany Rozwój Systemu

Projekt będzie rozwijany o następujące, kluczowe funkcjonalności, aby zwiększyć jego wartość analityczną i dostępność:

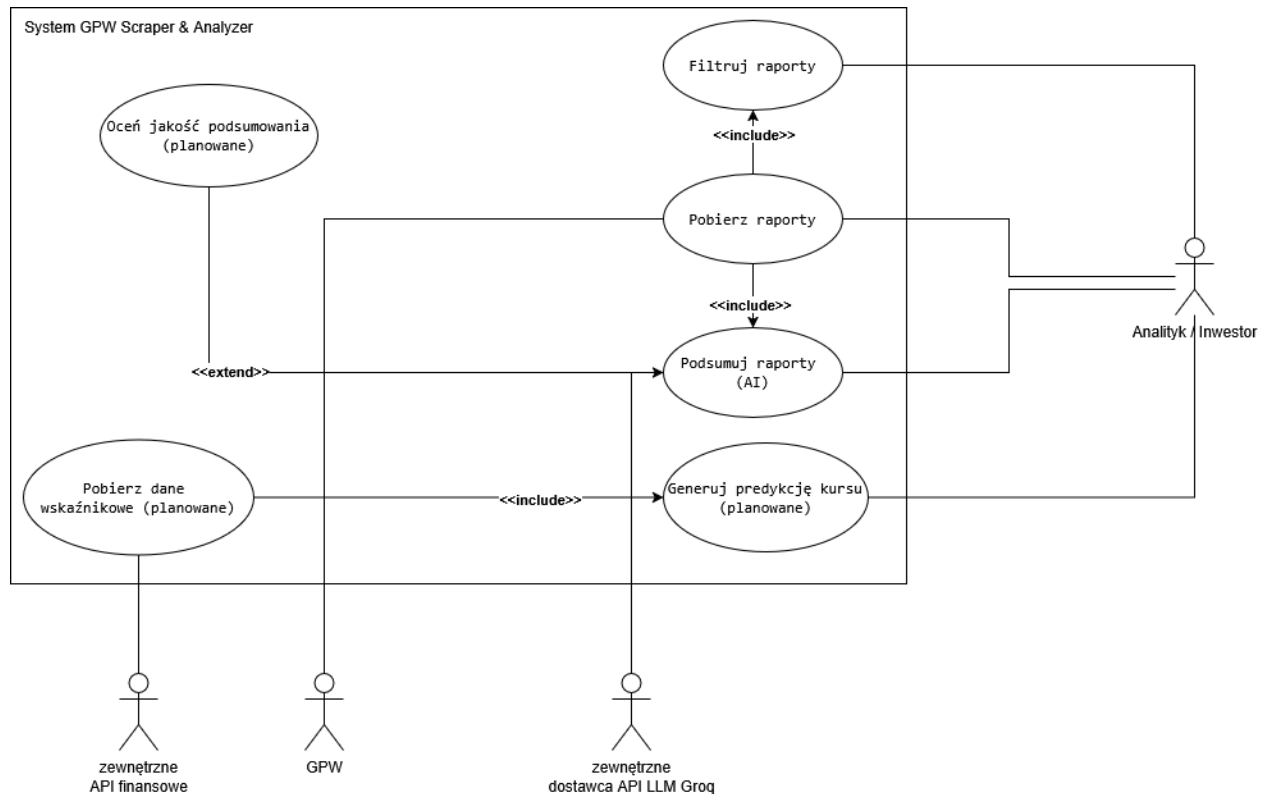
- **Integracja z Groq API:** Dodanie możliwości przetwarzania zapytań przez API Groq, co stanowi alternatywę dla lokalnego modelu AI i pozwoli na korzystanie z aplikacji na komputerach o niższej mocy obliczeniowej (bez dedykowanego GPU).
- **Metryki Oceny Modelu:** Implementacja i wyświetlanie w interfejsie metryk oceny jakości odpowiedzi generowanych przez model LLM (np. ROUGE, BLEU), co pozwoli na ocenę wiarygodności i precyzji podsumowań.
- **Integracja z Zewnętrznym API Finansowym:** Połączenie z API dostarczającym bieżące dane statystyczne i wskaźniki finansowe dla spółek z GPW (np. wskaźnik C/Z, C/WK, rentowność dywidendy).
- **Model Predykcyjny:** Stworzenie i integracja modelu regresji, który na podstawie danych historycznych, treści raportów oraz bieżących wskaźników finansowych będzie prognozował przyszłe trendy kursu akcji danej spółki.
- **Wizualizacja Danych:** Dodanie modułu do graficznej prezentacji predykcji oraz kluczowych wskaźników finansowych.

3. Analiza Systemu i Przypadki Użycia

3.1. Aktorzy Systemu

- **Analitik (Użytkownik Główny):** Osoba (np. pracownik firmy FinTech, inwestor), która bezpośrednio obsługuje system. Inicjuje procesy pobierania, filtrowania i analizy danych.
- **GPW (System Zewnętrzny):** Giełda Papierów Wartościowych, traktowana jako zewnętrzne źródło danych (raportów i informacji o nich).
- **Inwestor (Odbiorca Końcowy):** Osoba, która korzysta z przetworzonych przez system analiz i raportów w celu podejmowania decyzji inwestycyjnych.

3.2. Diagram Przypadków Użycia



[Diagram Przypadków Użycia]

Główne przypadki użycia obejmują:

1. Filtruj raporty
2. Pobierz raporty
3. Podsumuj raporty (AI)
4. Pobierz dane wskaźnikowe (planowane)
5. Generuj predykcję kursu (planowane)
6. Oceń jakość podsumowania (planowane)

3.3. Opis Przypadków Użycia

PU1: Filtrowanie i Pobieranie Raportów

- **Aktorzy:** Analityk, GPW
- **Opis:** Analityk specyfikuje kryteria wyszukiwania (nazwa spółki, zakres dat, typ raportu itp.). System wysyła zapytanie do serwisu GPW, scrapuje wyniki i prezentuje je w formie tabeli. Analityk może następnie pobrać wybrane raporty (PDF/HTML) oraz listę wyników (CSV).
- **Warunki wstępne:** Działające połączenie internetowe, dostępność serwisu GPW.
- **Warunki końcowe:** Raporty i/lub plik CSV zostają zapisane lokalnie na dysku użytkownika. Dane operacji zostają zapisane w bazie danych.

PU2: Generowanie Podsumowania AI

- **Aktorzy:** Analityk

- **Opis:** Po pobraniu raportów w formacie PDF, analityk może zlecić ich podsumowanie. System przekazuje treść raportów do modelu językowego (lokalnego Llama 3.2 lub zdalnego przez Groq API), który generuje zwięzłe streszczenie najważniejszych informacji.
- **Warunki wstępne:** Co najmniej jeden raport został pobrany. Model AI jest uruchomiony i dostępny.
- **Warunki końcowe:** W interfejsie użytkownika wyświetlone zostaje tekstowe podsumowanie raportu.

PU3: Generowanie Predykcji Kursu (Planowane)

- **Aktorzy:** Analityk
- **Opis:** Analityk wybiera spółkę. System pobiera dla niej aktualne dane wskaźnikowe z zewnętrznego API oraz dane historyczne. Na podstawie tych danych oraz ewentualnie sentymentu z ostatnich raportów, zintegrowany model regresji generuje prognozę przyszłego zachowania kursu.
- **Warunki wstępne:** Dostęp do zewnętrznego API finansowego, wytrenowany i zintegrowany model regresji.
- **Warunki końcowe:** W interfejsie użytkownika zostaje wyświetlona predykcja kursu (np. w formie tekstowej i/lub wykresu).

4. Scenariusze Użycia

4.1. Scenariusz Główny: Kompleksowa Analiza Spółki

1. **Analityk** uruchamia aplikację.
2. W polu "Company Name" wpisuje nazwę interesującej go spółki, np. "Mercor".
3. Wybiera liczbę raportów do pobrania, np. 20.
4. Zaznacza opcję "Download the CSV report".
5. Zaznacza typy raportów, które go interesują (np. current, quarterly).
6. Klika przycisk "Submit".
7. System łączy się z GPW, pobiera listę raportów zgodnych z kryteriami, wyświetla je w tabeli i zapisuje plik CSV. Równocześnie pobiera pełne raporty w formacie PDF do lokalnego katalogu.
8. **Analityk** przegląda listę i decyduje się na analizę ostatniego raportu kwartalnego.
9. Uruchamia funkcję podsumowania dla wybranego pliku.
10. System wysyła treść raportu do modelu LLM, który zwraca podsumowanie kluczowych wyników finansowych i zdarzeń. Podsumowanie pojawia się w polu "Output".
11. **(W przyszłości)** Analityk klika przycisk "Generuj Predykcję".
12. **(W przyszłości)** System pobiera najnowsze wskaźniki dla "Mercor" z API, łączy je z danymi historycznymi i wynikiem analizy raportu, a następnie model regresji oblicza prognozę, która zostaje wyświetlona użytkownikowi.

4.2. Scenariusz Alternatywny: Brak Raportów dla Wskazanej Spółki

1. **Analityk** uruchamia aplikację.
2. W polu "Company Name" wpisuje nazwę spółki, która nie istnieje lub nie opublikowała

żadnych raportów.

3. Klika "Submit".
4. System łączy się z GPW, ale nie znajduje żadnych pasujących wyników.
5. W interfejsie użytkownika (np. w tabeli lub w polu "Output") pojawia się komunikat: "Nie znaleziono raportów dla podanych kryteriów". Proces kończy się.

5. Projekt Bazy Danych

Baza danych została zaprojektowana w celu przechowywania danych pozyskanych w procesie scrapingu oraz historii operacji wykonywanych przez użytkownika.

5.1. Aktualny Opis Tabel (MySQL)

Tabela firma

Przechowuje unikalne nazwy spółek giełdowych.

Nazwa kolumny	Typ danych	Opis
id_firmy	int(11)	Klucz główny, autoinkrementowany identyfikator spółki.
nazwa	text	Unikalna nazwa spółki.

Tabela dane

Główna tabela przechowująca szczegółowe informacje o każdym zescrapowanym raporcie.

Nazwa kolumny	Typ danych	Opis
data	text	Data i godzina publikacji raportu.
tytul_raportu	text	Pełny tytuł raportu.
typ_raportu	enum(...)	Typ raportu (Bieżący, Półroczny, Kwartalny, etc.).
kategoria_raportu	enum('ESPI','EBI')	Kategoria raportu (ESPI lub EBI).
zmiana	double	Procentowa zmiana kursu.
kurs	double	Kurs zamknięcia w dniu publikacji raportu.
link	text	Bezpośredni link do strony z raportem na GPW.
Id_firmy	int(11)	Klucz obcy, łączący raport ze spółką w tabeli firma.

Tabela historia

Loguje parametry zapytań wysłanych przez użytkownika w celu umożliwienia przeglądania historii.

Nazwa kolumny	Typ danych	Opis
id	int(11)	Klucz główny, autoinkrementowany identyfikator wpisu.
company_name	text	Nazwa spółki użyta w zapytaniu.
report_amount	int(11)	Liczba raportów do pobrania.
download_type	text	Wybrane formaty pobierania (np. "PDF, HTML").
report_date	date	Data użyta w filtrze (jeśli podano).
report_type	text	Wybrane typy raportów (np. "current, annual").
report_category	text	Wybrane kategorie raportów (np. "EBI, ESPI").

5.2. Opis Relacji i Tabel

- **firma:** Tabela słownikowa przechowująca unikalne nazwy spółek giełdowych oraz przypisane im identyfikatory (id_firmy).
- **dane:** Główna tabela z danymi, zawiera szczegółowe informacje o każdym zescrapowanym raporcie. Pole Id_firmy jest kluczem obcym łączącym dany raport z konkretną spółką w tabeli firma.
- **historia:** Tabela logów, zapisująca parametry każdego zapytania wykonanego przez użytkownika, co pozwala na śledzenie historii jego działań.

Relacja: firma (1) --- (*) dane (Jeden do wielu). Jedna firma może mieć wiele raportów.

5.3. Proponowane Rozszerzenia Bazy Danych

W celu obsługi planowanych funkcjonalności, schemat bazy danych powinien zostać rozszerzony o następujące tabele:

- **Tabela podsumowania:**
 - id_podsumowania (PK)
 - id_raporu (FK do tabeli dane, jeśli dodamy tam PK)
 - tresc_podsumowania (TEXT)
 - model_uzyty (np. 'Llama3.2', 'Groq - Llama 4')
 - data_utworzenia (TIMESTAMP)
- **Tabela metryki_oceny:**
 - id_metryki (PK)
 - id_podsumowania (FK do tabeli podsumowania)
 - typ_metryki (np. 'ROUGE-1', 'BLEU')

- wartosc (FLOAT)
- **Tabela predykcje:**
 - id_predykcji (PK)
 - id_firmy (FK do tabeli firma)
 - data_predykcji (TIMESTAMP)
 - prognozowany_kurs (DOUBLE)
 - horyzont_czasowy (np. '1D', '7D', '1M')
 - dane_wejscowe (JSON z użytymi wskaźnikami, które wykorzystał do predykcji)