

Sprawozdanie

**Opracowanie aplikacji do statystycznej i geostatystycznej analizy
danych**

Autorzy: Lidia Teleszko (325708)
Izabella Kaim (319193)
Julia Zapała (325710)

Spis treści:

1. Cel projektu.....	2
2. Wstęp teoretyczny.....	2
3. Opis wykorzystanych danych.....	2
4. Realizacja projektu.....	3
5. Podsumowanie i wnioski.....	7
6. Bibliografia i załączniki.....	7

1. Cel projektu

Celem projektu było stworzenie aplikacji przeznaczonej do analizy danych meteorologicznych IMGW, z uwzględnieniem danych astronomicznych oraz położenia stacji pomiarowych w podziale na województwa i powiaty. W ramach realizacji projektu zaimplementowano metodę operowania na danych, wykorzystując bazę danych MongoDB i bazę danych Redis.

2. Wstęp teoretyczny

Współczesne systemy informacyjne coraz częściej opierają się na nowoczesnych technologiach w zakresie przechowywania i przetwarzania danych. Jedną z tych technologii są bazy NoSQL, które różnią się od tradycyjnych relacyjnych baz danych strukturą, skalowalnością oraz sposobem przetwarzania danych. Można wśród nich wyróżnić bazy danych MongoDB oraz Redis.

MongoDB to baza danych typu dokumentowego, która przechowuje dane w formacie BSON. Charakteryzuje ją duża elastyczność, co umożliwia przechowywanie zróżnicowanych typów danych. Służy głównie do przechowywania danych o znacznej objętości. Dane są przechowywane w dokumentach, które swoją budową przypominają obiekty JSON. Dokumenty składają się z par pól i wartości i mogą zawierać inne dokumenty. Dokumenty są przechowywane w kolekcjach, które nie wymuszają schematu. MongoDB zapewnia bogatą funkcjonalność, jest możliwe indeksowanie czy agregacja. Ten typ baz danych jest szczególnie chętnie wykorzystywany w nowoczesnych aplikacjach internetowych, aplikacjach Big Data oraz podczas analizy danych w czasie rzeczywistym.

Redis to baza danych klucz wartość, która działa w pamięci operacyjnej. Dzięki temu zapewnia wyjątkową szybkość dostępu do danych. Redis obsługuje wiele zaawansowanych struktur danych, co pozwala na szerokie zastosowanie. Dodatkową funkcjonalnością bazy danych jest zarządzanie danymi przestrzennymi w czasie rzeczywistym.

Powyższe bazy danych mają swoje unikalne cechy i zastosowania. Redis jest magazynem danych w pamięci, co zapewnia bardzo szybki dostęp do nich. Natomiast MongoDB jest wydajnym magazynem dokumentów na dysku. Obie bazy danych często są używane razem, aby poprawić szybkość działania aplikacji.

3. Opis wykorzystanych danych

Projekt wykorzystuje różnorodne źródła danych, które umożliwiają kompleksową analizę meteorologiczną z uwzględnieniem danych astronomicznych i administracyjnych. Poniżej przedstawiono opis wykorzystywanych danych:

Dane pogodowe IMGW

Dane meteorologiczne pochodzą z publicznie dostępnych zasobów IMGW: <https://danepubliczne.imgw.pl/datastore>. Dane pomiarowe pochodzą z automatycznych stacji IMGW i są przechowywane w Systemie Hydrologii (SH), który wykonuje ich wstępną weryfikację. Pliki tekstowe (kolumnowe, rozdzielone średnikiem) zawierają: Kod SH (9-cyfrowy identyfikator stacji), Parametr SH (oznaczenie parametru), Datę (czas GMT) oraz Wartość (zaobserwowaną wartość). Dane mają 10-minutową rozdzielczość czasową, są dostępne od 2008 roku, a ich szczegółowość określają pliki kody_stacji.csv i kody_parametrów.csv. Dla potrzeb projektu wykorzystane zostały tylko dane pogodowe z października 2024.

Dane astronomiczne

Dane astronomiczne pozyskiwane są za pomocą biblioteki *Astral*. Umożliwia ona obliczenie kluczowych zdarzeń astronomicznych, takich jak wschód i zachód Słońca za pomocą funkcji. Niezbędne jest ustawienie lokalizacji miejsca, z którego chcemy pozyskać dane astronomiczne oraz strefa czasowa danego miejsca. Dla Polski wybieramy strefę Europe/Warsaw. Funkcja zwraca godzinę wschodu i zachodu Słońca dla konkretnej daty i lokalizacji z dokładnością do sekund.

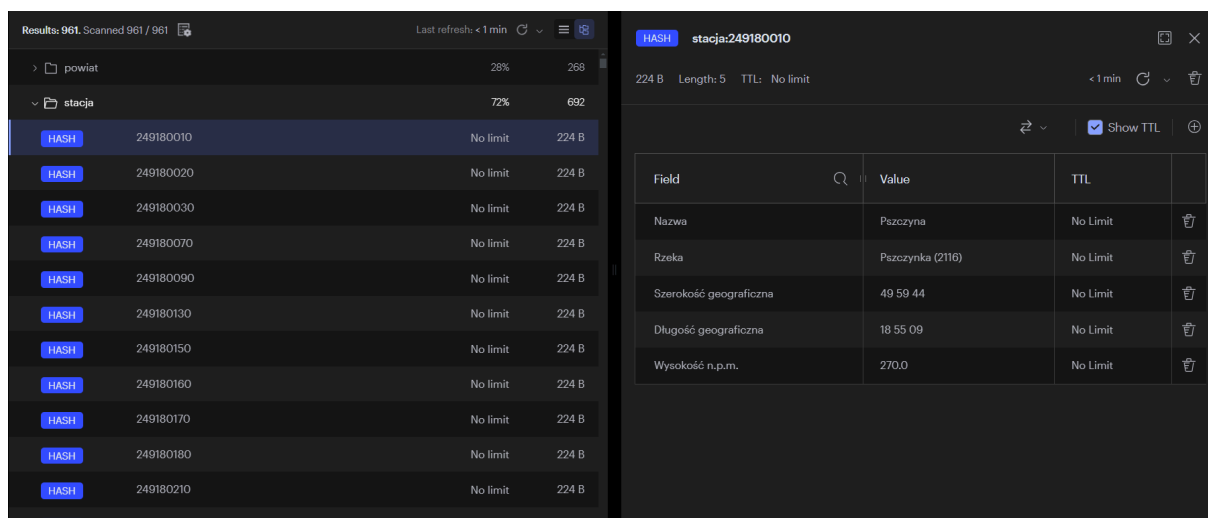
Dane administracyjne

Dane administracyjne są dostępne w formacie SHP, co pozwala na integrację przestrzenną z danymi meteorologicznymi. Na potrzeby aplikacji została stworzona warstwa powiaty_woj.shp. Jest to zmodyfikowana warstwa powiatów pobrana z PRG. W Arcgis Pro dodano do niej kolumnę z nazwą województwa. Następnie na podstawie wartości TERYT powiatu oraz danych o kodach województw,¹ możliwe było przypisanie do każdego powiatu nazwy województwa.

4. Realizacja projektu

4.1 Przechowywanie danych

Aplikacja wykorzystuje bazy danych Redis i MongoDB, które należą do baz NoSQL. Do wersji chmurowej bazy danych Redis zostały zaimportowane dane o stacjach, na podstawie pliku kody_stacji.csv. Dane te są przechowywane w postaci hashy, gdzie kluczem jest ID stacji.



The screenshot displays the Redis Insight interface. On the left, a sidebar shows a tree view with 'powiat' (268 items) and 'stacja' (692 items). The main panel shows a list of station hashes. On the right, a detailed view for 'stacja:249180010' is shown, including a table of metadata.

Field	Value	TTL
Nazwa	Pszczyna	No Limit
Rzeka	Pszczynka (2116)	No Limit
Szerokość geograficzna	49 59 44	No Limit
Długość geograficzna	18 55 09	No Limit
Wysokość n.p.m.	270.0	No Limit

Zdj.1 Widok bazy danych Meteodata w Redis Insight - stacje

Na potrzeby optymalizacji działania aplikacji stworzono plik csv, przechowujący informacje o istniejących stacjach w poszczególnych powiatach. Było to możliwe dzięki wykorzystaniu biblioteki geopandas i metody contains(). Na początku wczytano dane o stacjach z Redis i na podstawie współrzędnych stworzono punkty. Następnie funkcja contains() szuka, czy dany punkt/stacja znajduje się w danym poligonie/powiecie. Tak stworzony plik został zaimportowany do bazy danych Redis, gdzie key stanowi nazwa powiatu.

¹ <https://pl.wikipedia.org/wiki/Wojew%C3%B3dztwo>

The screenshot shows the Redis Insight interface. On the left, a list of counties (powiat) is displayed with columns for HASH, name, and size. On the right, the details for 'powiat:augustowski' are shown, including a table with fields like 'Stacje' and their values.

Field	Value	TTL
Stacje	253220080, 253220360	No Limit

Zdj.2 Widok bazy danych Meteodata w Redis Insight- powiaty

Zamieszczono również dane w bazie MongoDB reprezentujące położenie stacji pomiarowych wraz z dodatkowymi informacjami o każdej stacji, takimi jak link do stacji czy data powstania.

Dane przechowywane w bazie MongoDB mają postać dokumentów w formacie GeoJSON.

```
{
  "_id": {
    "$oid": "6791791cae406fe5cddc3e8d"
  },
  "type": "Feature",
  "properties": {
    "ifcid": 149180010,
    "gmlid": "ID.149180010",
    "gmlidentif": "http://iip.imgw.pl/so/EF/EnvironmentalMonitoringFacility/PL.ZIPGW.2144/SZS/149180010",
    "id_localid": "149180010",
    "id_namespa": "PL.ZIPGW.2144/SZS",
    "name": "149180010",
    "name1": "Krzyżanowice",
    "additional": "Stacja wodowskazowa rzeczna II rzędu",
    "mediamonit": "http://inspire.ec.europa.eu/codelist/MediaValue/water",
    "mediamon_1": null,
    "responsibl": "Instytut Meteorologii i Gospodarki Wodnej Państwowy Instytut Badawczy",
    "mobile": 0,
    "activitytype": "1927-05-09",
    "activity_1": null
  },
  "geometry": {
    "type": "Point",
    "coordinates": [
      51.844343444343814,
      15.173175695409498
    ]
  }
}
```

Zdj.3 Przykład danych przechowywanych w MongoDB

4.2 Wykonanie statystyk

Statystyki dla poszczególnych stacji są obliczane za pomocą dedykowanych funkcji, które umożliwiają przetwarzanie danych według określonych kryteriów. Funkcje te pozwalają na wybór danych dla konkretnej daty lub przedziału czasowego, a także na podział wyników na porę dnia i noc. Za pomocą tych metod można obliczyć kluczowe parametry statystyczne, takie jak:

- Średnia – wartość przeciętna mierzonych danych,
- Mediana – środkowa wartość zbioru danych,

- Średnia obciąża – średnia z pominięciem skrajnych wartości, co zwiększa odporność na wartości odstające.

Dodatkowo, funkcje zawierają mechanizmy obsługi sytuacji, w których brak jest dostępnych danych, co pozwala na uniknięcie błędów w analizie. Statystyki te umożliwiają analizę różnych parametrów meteorologicznych, takich jak : temperatura powietrza (oficjalna), temperatura gruntu (czujnik), kierunek wiatru (czujnik), średnia prędkość wiatru (czujnik, 10 minut), prędkość maksymalna wiatru (czujnik), suma opadu 10-minutowego, suma opadu dobowego, suma opadu godzinowego, wilgotność względna powietrza (czujnik), największy poryw wiatru w okresie 10 minut (stacja synoptyczna) oraz zapas wody w śniegu (obserwator).

Analiza może być przeprowadzona zarówno dla pojedynczej daty, jak i określonego przedziału czasowego, co umożliwia elastyczne dostosowanie zakresu badania do potrzeb użytkownika.

4.3 Interfejs graficzny

Interfejs graficzny aplikacji został zaprojektowany przy użyciu biblioteki **Tkinter**. Aplikacja została podzielona na kilka sekcji, które umożliwiają użytkownikowi wykonywanie operacji związanych z analizą danych meteorologicznych. W pierwszej sekcji użytkownik wybiera województwo z rozwijanej listy, co automatycznie aktualizuje dostępne powiaty w kolejnej sekcji. Następnie, po wybraniu powiatu, możliwe jest wybranie konkretnej stacji meteorologicznej. Każda sekcja interfejsu zawiera etykiety oraz rozwijane listy , które pomagają w szybkim wyborze odpowiednich wartości. Dodatkowo, aplikacja posiada funkcję wyświetlania szczegółowych informacji o wybranej stacji meteorologicznej. Po dokonaniu wyboru stacji, użytkownik może skorzystać z przycisku „Info o stacji”, który otwiera nowe okno z danymi pobranymi z bazy MongoDB. Te dane są przedstawiane w postaci tabelarycznej. Kolejnym krokiem jest wybór sensora, z którego dane mają być analizowane. Lista dostępnych sensorów została wstępnie zdefiniowana i zawiera kluczowe parametry meteorologiczne, takie jak temperatura powietrza, wilgotność czy prędkość wiatru. Po wyborze sensora użytkownik podaje datę, korzystając z pola tekstowego lub wygodnego kalendarza umożliwiającego łatwe wskazanie interesującego dnia. Interfejs pozwala również na precyzyjne określenie zakresu czasowego analizy poprzez wybór godzin, minut i sekund dla początku oraz końca przedziału czasowego. Po wybraniu wszystkich wymaganych parametrów użytkownik może nacisnąć przycisk „Pokaż statystyki”, który uruchamia proces przetwarzania danych. Wyniki są następnie prezentowane w dolnej części interfejsu w formie czytelnego podsumowania, obejmującego statystyki dzienne, godzinowe oraz dane dotyczące dnia i nocy w odniesieniu do wschodu i zachodu słońca.


Statystyki Meteorologiczne

Wybierz województwo:

Wybierz powiat:

Wybierz stację:

Wybierz sensor:

Data (YYYY-MM-DD): 

Godzina startowa:

Godzina końcowa:


Statystyki Meteorologiczne

Wybierz województwo:

Wybierz powiat:

Wybierz stację:

Wybierz sensor:

Data (YYYY-MM-DD): 

Godzina startowa:

Godzina końcowa:

Statystyki na dzień 2024-10-10

	Wartość
średnia	16.2
mediana	15.3
średnia_obc	16.0

dla pory dnia: dzień

	Wartość
średnia	18.9
mediana	19.4
średnia_obc	19.1

dla godziny 03:00:00

	Wartość
średnia	15.2
mediana	15.2
średnia_obc	15.2

dla pory dnia: noc

	Wartość
średnia	12.0
mediana	12.0
średnia_obc	12.1

dla przedziału godzin 03:00:00-09:00:00

	Wartość
średnia	16.8
mediana	15.7
średnia_obc	16.7

Wschód słońca: 06:59:04
Zachód słońca: 18:02:28

Zdj.4 Interfejs graficzny- okno główne

Informacje o stacji: Pszczyna

Attribute	Value
mediamonit	None
ifcid	249180010
additional	Stacja klimatologiczna IV rzędu
activitytype	1954-02-25
id_localid	249180010
id_namespa	PL.ZIPGW.2144/SZS
activity_1	None
gmliid	ID.249180010
gmliidentif	http://iip.imgw.pl/so/EF/EnvironmentalMonitoringFacility/PL.ZIPGW.2144/SZS/249180010
name1	Pszczyna
mediamon_1	http://inspire.ec.europa.eu/codelist/MediaValue/water
name	249180010
responsibl	Instytut Meteorologii i Gospodarki Wodnej Państwowy Instytut Badawczy
mobile	0

Zdj.5 Interfejs graficzny- okno z informacją o stacji

5. Podsumowanie i wnioski

Aplikacja umożliwia efektywne zarządzanie danymi meteorologicznymi. Dzięki połączeniu z bazami MongoDB i Redis możliwe jest przechowywanie wszystkich danych w jednym miejscu, szybkie pobieranie i ich przetwarzanie. Wadą takiego rozwiązania jest fakt, że wersje bezpłatne umożliwiają przechowywanie bardzo małych zbiorów danych (Redis do 3MB), dlatego nie można było dodać wszystkich danych meteorologicznych. Interfejs aplikacji jest intuicyjny i posiada wiele możliwości obliczania statystyk meteorologicznych.

6. Bibliografia i załączniki

- [Dokumentacja MongoDB](#)
- [Dokumentacja Redis](#)
- [Wykaz BDOT10k](#)
- [Dokumentacja biblioteki tkinter](#)
- [Dokumentacja biblioteki Astral](#)
- [Wikipedia - województwo](#)