

Laboratorium 7

Dane przestrzenne wykorzystane w projekcie

Cele

1. Zapoznanie się z danymi, które będą wykorzystywane w projekcie.
2. Uporządkowanie i harmonizacja danych.

Dane

Projekt będzie wykorzystywał następujące dane:

- Dane pogodowe pochodzące z IMGW:
 - <https://danepubliczne.imgw.pl/datastore>
 - <https://dane.imgw.pl/datastore>,
 - <https://danepubliczne.imgw.pl/>
- Dane astronomiczne (<https://pypi.org/project/astral/>)
- Dane administracyjne

Dane IMGW

Opis danych w pliku <https://dane.imgw.pl/datastore/getfiledownload/Arch/Telemetria/Meteo/opis.txt>:

Dane pomiarowe prezentowane w poniższych katalogach, należy traktować jako dane operacyjne.

Dane pomiarowe w plikach pochodzą z automatycznych stacji pomiarowych, stan bieżący stacji pomiarowej jak i położenie, można śledzić na stronie <http://monitor.pogodynka.pl>.

Dane pomiarowe napływają do bazy danych zwanej System Hydrologii (w skrócie SH), podstawowej aplikacji operacyjnej Biur Prognoz Hydrologicznych. Gdzie następuje wstępna, operacyjna weryfikacja danych.

Plik zostały wygenerowane przez System Hydrologii, w standardowym pliku eksportu. Jest to plik tekstowy w formacie kolumnowym, rozdzielony średnikiem. Struktura pliku to:

- Kolumna 1 - KodSH – 9-cyfrowy kod stacji pomiarowej, stosowany w IMGW PIB do identyfikacji stacji pomiarowej,
- Kolumna 2 - ParametrSH – parametr pomiarowy, stosowany w Systemie Hydrologii,
- Kolumna 3 - Data – czas, w którym została zaobserwowana wartość pomiarowa,
- Kolumna 4 - Wartość – wartość pomiarowa.

Każda stacja pomiarowa reprezentowana jest poprzez kod, nazwę oraz rzekę. Kod SH identyfikuje typ stacji pomiarowej:

- 1XXXXXXXX - stacja wodowskazowa,
- 2XXXXXXXX - stacja meteorologiczna, niższego rzędu, o mniejszej liczbie parametrów pomiarowych,
- 3XXXXXXXX – stacja meteorologiczna I rzędu (tzw. synoptyczna), o pełny zakresie pomiarowym.

Lista stacji znajduje się w pliku kody_stacji.csv.

Parametr SH, stosowany w Systemie Hydrologii, identyfikuje jakie to są dane pomiarowe. Początkowa literka oznacza typ danych (dane rzeczywiste, prognoza), np. B – dana rzeczywista. Trzy pierwsze cyfry, oznaczają grupę pomiarową np. 000 – pomiar stanów wody, 003 – pomiar temperatury powietrza. Dwie kolejne,

określają dokładnie jaki parametr np. 01 – dana pomiarowa od obserwatora, 02 – dana pomiarowa z czujnika. Ostatnia litera oznacza czy wartość pomiarowa jest bezpośrednio zaobserwowana, czy przetworzona w SH:

- A - Bezpośrednio zaobserwowana dana pomiarowa,
- B - Wartość pomiarowa przeliczona np. przepływ wody,
- S – Oficjalna wartość operacyjna.

Lista parametrów znajduje się w pliku kody_parametrów.csv.

Czas pomiaru podawany jest w czasie uniwersalnym (GMT), bez uwzględnienia czasu lokalnego oraz zmian czas letni/czas zimowy. W IMGW PIB, stosowana jest 10 minutowa rozdzielczość czasowa danych pomiarowych pochodzących z Systemu Telemetrii (ST). Wartość pomiarowa podawana jest z dokładnością zdefiniowaną dla poszczególnych parametrów, wynikająca z dokładności urządzeń pomiarowych. Braki danych mogą wynikać z wstępnej weryfikacji danych lub awarii urządzenia pomiarowego. Jak również uzależnione jest od daty uruchomienia stacji.

Zakres danych pomiarowych prezentowany jest od roku 2008. Zgodnie z zarządzeniem dyrektora IMGW PIB w tym roku, dane pomiarowe z sieci stacji telemetrycznych, zostały dopuszczone do użytkowania. W pierwszej kolejności pliki podzielono ze względu na rodzaj danych, osobne katalogi:

- Hydro - hydrologiczne,
- Meteo - meteorologiczne.

W nich znajdują się katalogi odpowiadające roku pomiaru oraz pliki z kodami stacji i parametrów. Dane pomiarowe są porcjowane miesięcznie, każdy parametr jako osobny plik. Pliki z danego miesiąca są spakowane w standardzie zip i tar.

Dodatkowo dane w formacie GeoJSON dostępne w Teams:

- Położenie stacji pomiarowych - Effacility.geojson

Dane astronomiczne

Biblioteka astral (<https://pypi.org/project/astral/>) pozwala wyliczyć dokładny czas następujących zjawisk: świt, wschód słońca, południe, zachód słońca, zmierzch. Na ich podstawie można wyliczyć, kiedy trwa dzień, a kiedy noc (<https://astral.readthedocs.io/en/latest/index.html>).

Dane administracyjne

Dane w formacie SHP dostępne w Teams:

- Granice powiatów – powiaty.shp
- Granice województw – woj.shp

Harmonizacja danych

Proponowane kroki:

1. Pobranie danych źródłowych z portalu IMGW dla wybranego roku, miesiąca i zakresu danych:

Kod	Nazwa
B00300S	Temperatura powietrza (oficjalna)
B00305A	Temperatura gruntu (czujnik)
B00202A	Kierunek wiatru (czujnik)
B00702A	Średnia prędkość wiatru czujnik 10 minut

B00703A	Prędkość maksymalna (czujnik)
B00608S	Suma opadu 10 minutowego
B00604S	Suma opadu dobowego
B00606S	Suma opadu godzinowego
B00802A	Wilgotność względna powietrza (czujnik)
B00714A	Największy poryw w okresie 10min ze stacji Synoptycznej
B00910A	Zapasy wody w śniegu (obserwator)

2. Wczytanie danych meteo do struktur danych modułu Pandas.
3. Pobranie danych przestrzennych (folder w Teams)
4. Wczytanie danych przestrzennych do struktur danych modułu GeoPandas.

Uwagi

Przy opracowaniu rozwiązania warto zwrócić uwagę na:

1. Sposób pobierania plików ZIP:

```
import requests, zipfile, io
file = requests.get("https://.....")
zip = zipfile.ZipFile(io.BytesIO(file.content))
zip.extractall("C:\Dane\")
```

2. Wczytywanie plików CSV za pomocą biblioteki Pandas - https://www.w3schools.com/python/pandas/pandas_csv.asp
3. Typy danych wprowadzone przez bibliotekę Pandas:
 - a. `DataSet` - https://www.w3schools.com/python/pandas/pandas_series.asp
 - b. `DataFrame` - https://www.w3schools.com/python/pandas/pandas_dataframes.asp
4. Sposób liczenia średniej w Pandas - <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.mean.html>
5. Sposób liczenia mediany w Pandas - <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.median.html>
6. Sposób liczenia średniej obcinanej w bibliotece SciPy:

```
from scipy import stats
stats.trim_mean(data, 0.1) # ucinanie na poziomie 10%
```

7. Sposób przetwarzania danych o lokalizacji w bibliotece Astral:

```
from astral import LocationInfo
city = LocationInfo("Warsaw", "Poland", "Europe/Warsaw", 52.232222, 21.008333)
print(city.name, city.latitude, city.longitude)
```

8. Sposób liczenia parametrów słońca w zadanym miejscu:

```
from astral.sun import sun
s = sun(city.observer, date=datetime.date(2021, 02, 14), tzinfo=city.timezone)
print(s["dawn"], s["sunrise"], s["noon"], s["sunset"], s["dusk"])
```

Przydatne linki

Biblioteka Pandas - <https://pandas.pydata.org/>

Podręcznik biblioteki Pandas - <https://pandas.pydata.org/docs/pandas.pdf>

Pandas Tutorial - <https://www.w3schools.com/python/pandas/default.asp>

Odnosińki do ciekawych źródeł wiedzy o GeoPandas

- Instalacja: https://geopandas.org/getting_started/install.html
- https://geopandas.org/getting_started/introduction.html
- GeoSeries https://geopandas.org/docs/user_guide/data_structures.html#geoseries
- GeoDataFrame https://geopandas.org/docs/user_guide/data_structures.html#geodataframe
- Praca z plikami danych https://geopandas.org/docs/user_guide/io.html#reading-spatial-data
- Praca z bazą danych przestrzennych: https://geopandas.org/docs/user_guide/io.html#spatial-databases
- Przykłady: <https://github.com/geopandas/geopandas>
- Introduction to Geospatial Data in Python
<https://www.datacamp.com/community/tutorials/geospatial-data-python>
- <https://towardsdatascience.com/geopandas-hands-on-geospatial-relations-and-operations-a6e7047d7ba1>
- <https://towardsdatascience.com/plotting-maps-with-geopandas-428c97295a73>