

**Andrzej KRZYŻANOWSKI, Bartosz SADOCHA, Bartosz SOKÓŁ,  
Piotr WERESZCZYŃSKI**

Studenckie Koło Naukowe Pasjonatów Elektroniki,  
Wydział Elektroniki i Informatyki, Politechnika Koszalińska  
shaidel@wp.pl, marlokk99@gmail.com, tomanycreeps@gmail.com

## **PODSTAWY EKSPLOKACJI DANYCH W PYTHONIE Z WYKORZYSTANIEM BIBLIOTEKI PANDAS NA PRZYKŁADZIE APLIKACJI MONITORUJĄCEJ POSTĘP COVID-19**

### **Streszczenie**

W artykule przedstawiono projekt aplikacji monitorującej postęp CoViD-19 z podziałem na kraje i regiony. Program pobierając dane w czasie rzeczywistym przedstawia je w formie wykresu prezentując liczbę zachorowań, wyzdrowień oraz zgonów.

Aplikacja została napisana w języku Python z wykorzystaniem bibliotek: *Pandas* (pobieranie i obróbka danych), *matplotlib* (tworzenie wykresów), *PIL* (zapis wykresu do pliku graficznego) oraz *tkinter* (odpowiedzialnej za graficzny interfejs użytkownika).

Słowa kluczowe: Python, Pandas, eksploracja danych

### **1. Wstęp**

Pojawienie się epidemii Covid-19 postawiło przed społeczeństwem wiele wyzwań. Jednym z nich jest zapewnienie szybkiego dostępu do wiarygodnych danych o postępie choroby. Jednym z rozwiązań są aplikacje eksplorujące dane.

Eksploracja danych zyskuje na popularności wraz ze wzrostem szybkości pracy komputerów, która pozwala na pozyskanie i przetworzenie dużych zbiorów danych w rozsądnym czasie.

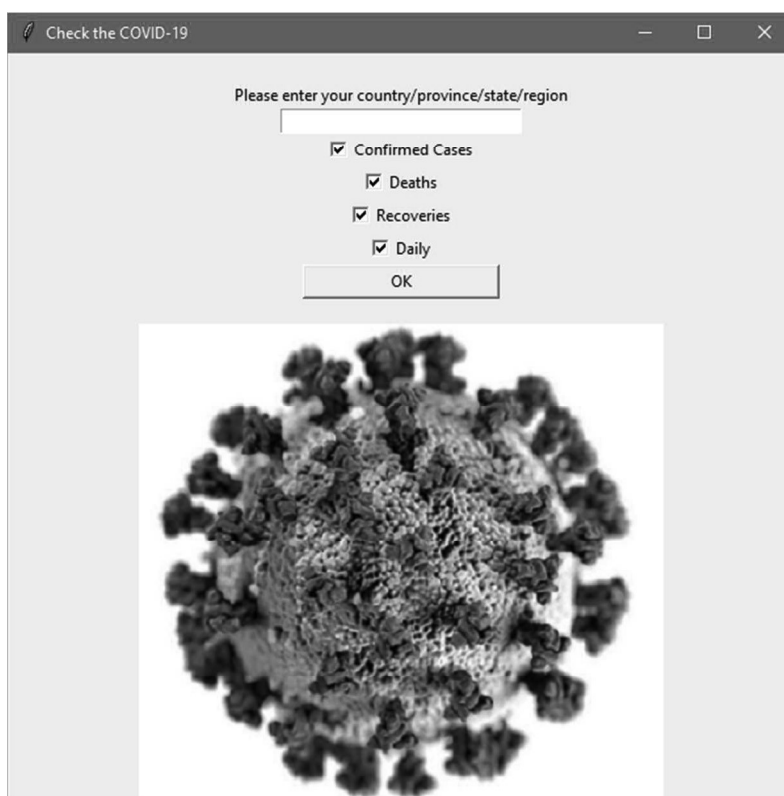
Celem projektu było napisanie programu, który w szybki sposób, poprzez pobranie i przetworzenie wiarygodnych danych, przedstawi wykresy postępu choroby (zachorowania, ozdrowienia, zgony) w ujęciu narastającym oraz dziennym, z podziałem na kraje. Aplikacja powstała podczas zajęć Koła Pasjonatów

Elektroniki, które działa przy Katedrze Systemów Cyfrowego Przetwarzania Sygnałów na Wydziale Elektroniki i Informatyki Politechniki Koszalińskiej.

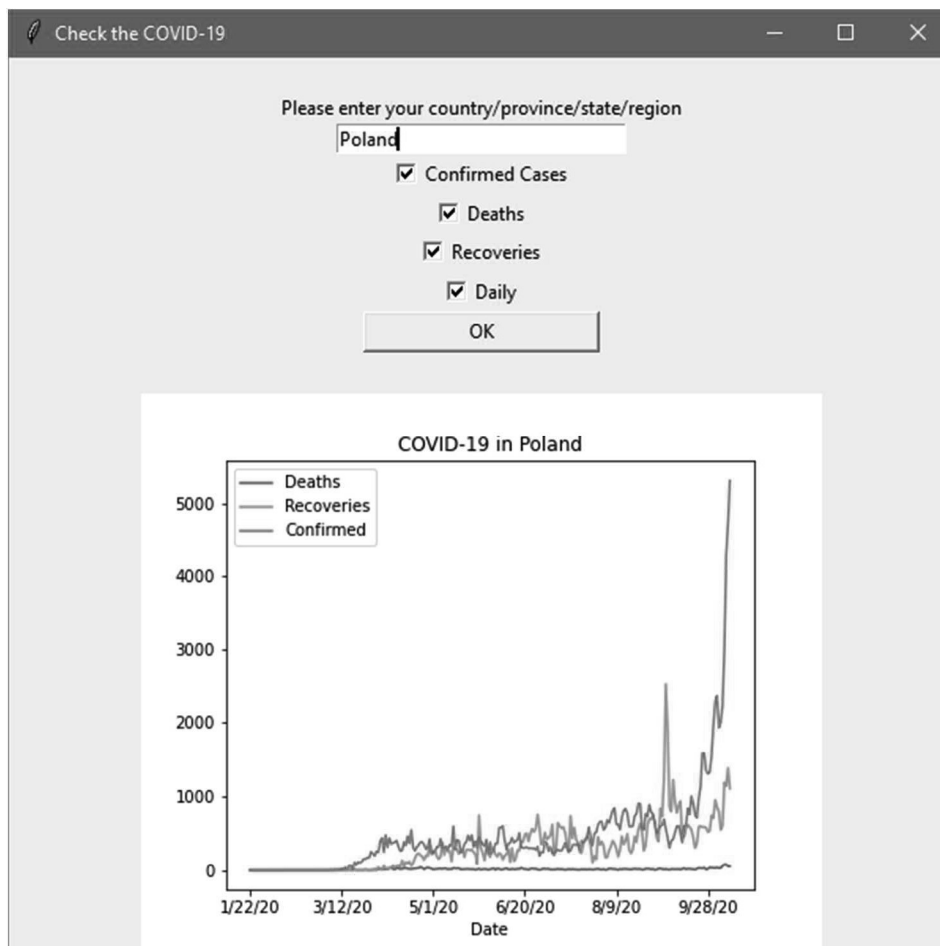
W kolejnych rozdziałach omówiono występujące w programie funkcjonalności wykorzystanych bibliotek, format plików ze źródłami danych oraz samą aplikację.

## 2. Projekt aplikacji

Aplikacja jest bardzo łatwa w obsłudze. Użytkownik wpisuje nazwę kraju w języku angielskim. Po wyborze danych, które go interesują (np. w ujęciu narastającym, czy dziennym) potwierdza swój wybór przyciskiem OK. W oknie aplikacji prezentowany wtedy jest wykres z poszukiwanymi danymi.



Rys. 1. Startowy wygląd aplikacji



Rys. 2. Przykładowy wykres

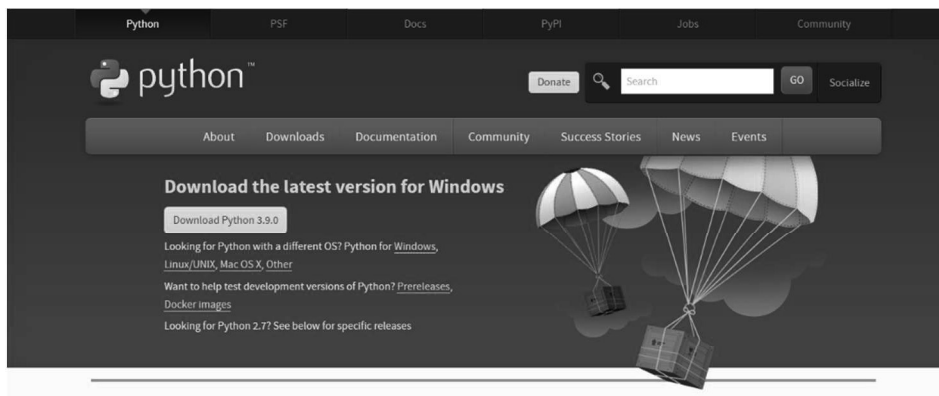
Program jest też zabezpieczony przed wprowadzeniem błędnych danych.



Rys. 3. Komunikat błędu

### 3. Instalacja i uruchomienie aplikacji

Program do prawidłowego działania wymaga zainstalowania Pythona w wersji 3.x oraz bibliotek pandas, matplotlib i Pillow, które można zainstalować poleceniem `pip install`



Rys. 4. Strona startowa języka Python

```

Wiersz polecenia
C:\Users\> pip install pandas
Requirement already satisfied: pandas in c:\users\andrzej krzyżanowski\appdata\local\programs\python\python36\lib\site-packages (1.1.3)
Requirement already satisfied: pytz>=2017.2 in c:\users\andrzej krzyżanowski\appdata\local\programs\python\python36\lib\site-packages (from pandas) (2020.1)
Requirement already satisfied: numpy>=1.15.4 in c:\users\andrzej krzyżanowski\appdata\local\programs\python\python36\lib\site-packages (from pandas) (1.19.2)
Requirement already satisfied: python-dateutil>=2.7.3 in c:\users\andrzej krzyżanowski\appdata\local\programs\python\python36\lib\site-packages (from pandas) (2.8.1)
Requirement already satisfied: six>=1.5 in c:\users\andrzej krzyżanowski\appdata\local\programs\python\python36\lib\site-packages (from python-dateutil>=2.7.3->pandas) (1.15.0)
You are using pip version 18.1, however version 20.2.3 is available.
You should consider upgrading via the 'python -m pip install --upgrade pip' command.

C:\Users\>
C:\Users\>

```

Rys. 5. Przykład instalacji biblioteki z pomocą instalatora pip

### 4. Format plików CSV

Dane, na podstawie których generowane są wykresy, pobierane są ze strony <https://github.com/CSSEGISandData> w formacie csv.



Kolejnym problemem było ograniczenie biblioteki Matplotlib, która nie pozwala w łatwy sposób na wybór opcji wykresów. Rozwiązano to poprzez wykorzystanie pozostałych bibliotek. Biblioteka PIL pozwala na zapisanie wykresu jako plik png, który z pomocą biblioteki Tkinter osadzamy w oknie prezentacji danych.

Sam Tkinter pozwolił też na dodanie przycisków umożliwiających przełączanie opcji prezentowanego wykresu.

## 7. Możliwości rozwoju aplikacji

Aplikacja jest we wstępnej fazie. Wskazane jest przepisanie interfejsu użytkownika, które umożliwiłoby dodanie takich funkcjonalności jak wysłanie wykresu (w formie pliku png) mailem. Kolejnym pomysłem jest możliwość prezentowania wykresów dla dwóch krajów celem porównania.

Innym obszarem rozbudowy programu są dane statystyczne. Dostęp do przetworzonych danych pozwala na wprowadzenie metod predykcji. Zarówno przez zamodelowanie podobieństw postępu CoViD-19 w danych krajach, jak i wprowadzenie zmiennych dla estymacji przyszłych danych.

## 8. Podsumowanie

Program spełnia swoją podstawową funkcję i przy dostępie do Internetu pozwala na szybkie pozyskanie danych o postępie pandemii w wybranym kraju. Zaprezentowane możliwości rozwoju oraz udostępnienie go na licencji MIT pozwala na zwiększenie jego możliwości w najbliższym czasie.

Link do kodu źródłowego aktualnej wersji aplikacji:

<https://github.com/Shaidel/Check-the-COVID-19>

## Bibliografia

1. Eric Matthes: „Python. Instrukcje dla programisty”, Wydawnictwo HELION, Gliwice (2016), ISBN 978-83-283-2595-1
2. Alberto Boschetti, Luca Massaron: „Python. Podstawy nauki o danych”, Wydawnictwo HELION, Gliwice (2017), ISBN 978-83-283-3423-6
3. Online: Pandas – Home <https://pandas.pydata.org/>
4. Online: matplotlib – Home <https://matplotlib.org/>
5. Online: PIL – Home <https://pillow.readthedocs.io/en/stable/>