1. **Co to jest? Do czego służy?**

Skrypt napisany w Pythonie, służy do pobrania i procesowania danych z *API Polskiej Wywiadowni Gospodarczej (PWG)* w celu pozyskania informacji nt obecnych i przeszłych powiązań pomiędzy firmami a osobami, które zajmowały stanowiska w zarządzie, radzie nadzorczej, prokurze, reprezentacji podmiotów. Dane są zapisane w formacie JSON, przetwarzane oraz eksportowane do pliku w formacie xlsx. Dane mogą zostać wykorzystane do realizacji projektów śledczych, w szczególności do utworzenia map powiązań typu BTT / „Pająk”.

Skrypt wykorzystuje metodę getFull, która została opisana w dokumentacji udostępnionej przez PWG:

<https://api.pwginfo.pl/doc/Api_Pwc.html>

1. **Jak działa?**

Skrypt pobiera JSONy zawierające informacje nt. danych rejestrowanych dla wprowadzonych przez użytkownika identyfikatorów podmiotów (mogą to być NIP, KRS, REGON, PWG\_ID lub PESEL). Wykorzystywana metoda to getFull. Dane mogą zostać pobrane dla podmiotu gospodarczego (identyfikator firmy) lub osoby fizycznej, która pracuje/pracowała na stanowisku kierowniczym/reprezentacyjnym w firmie zarejestrowanej w KRS.

Działanie użytkownika polega na:

1. Podaniu danych wymaganych do uruchomienia skryptu:
   1. Zapisania jako zmienną *INPUT\_FILE* ścieżki do pliku (łącznie z nazwą i formatem) w formacie .xlsx, w którym znajduje się lista NIPów sprawdzanych podmiotów oraz podania w pliku NIPów pierwszej kolumnie w pierwszej zakładce, na przykład: *r'C:\Users\mmandziej001\Desktop\Projects\PWG\INPUT\_API\NIP\_list.xlsx'*
   2. Podania jako zmienną *JSONS\_OUTPUT\_PATH* ścieżki do katalogu, do którego mają zostać zwrócone pobrane JSONy, na przykład:

*r'C:\Users\mmandziej001\Desktop\Projects\PWG\OUTPUT\_API\GET\_FULL\\'*

* 1. Podania jako zmienną *OUTPUT\_PATH* ścieżki do katalogu, do którego mają zostać wyeksportowane sparsowane dane rejestrowe w formacie .xslx, na przykład:

*r'C:\Users\mmandziej001\Desktop\Projects\PWG\OUTPUT\_PARSE\\'*

* 1. Opcjonalnie podania wartości parametru *DOWNLOAD\_TYPE (person* lub *company*). Skrypt został napisany w taki sposób, żeby sprawdzić długość i poprawność wprowadzonych ID tak i na tej podstawie zweryfikować typ wprowadzonych danych.

1. Uruchomieniu skryptu:

W skrypcie-bibliotece **pwg\_ordinary** znajdują się obiekty oraz funkcje wykorzystywane do parsowania i eksportowania danych. Skrypt został udokumentowany w pliku *pwg\_doc*. Proces pobierania danych opiera się na skrypcie *get\_data.py*, który został udokumentowany w pliku *get\_data\_doc*. Zaczytanie danych z pliku wejściowego odbywa się przy wykorzystaniu funkcji *get\_nip\_list*.

Skrypt działa w następujący sposób:

Identyfikatory firm/osób wprowadzone w pliku wejściowym są zaczytywane jako lista i w pętli dla każdego z nich pobierane są dane, które są eksportowane w formacie JSON do wskazanego katalogu. Po pobraniu dane są przekazywane do obiektów *company\_ids\_jsons* (słownik identyfikator: json dla podmiotów, gdzie pobranie danych zakończyło się sukcesem)*, company\_ids\_errors* (słownik identyfikator: json dla podmiotów, gdzie podczas pobrania danych wystąpił błąd). Następnie dane, które zostały pobrane prawidłowo są przetwarzane. W zależności od typu identyfikatora przy wykorzystaniu funkcji *parsed\_linked\_entities* z JSONów wyciągane są niezbędne informacje. Istnieją dwa streamy parsowania danych:

1. Dane pobrane dla podmiotów gospodarczych – klucz -> identifkator podmiotu:

imię, nazwisko, PESEL, stanowisko, rola w firmie, data końca i początku wykonywania obowiązków, aktualność powiązania

1. Dane pobrane dla osób indywidualnych – klucz -> numer PESEL:

PESEL, imię, nazwisko, NIP, REGON, KRS i nazwa podmiotu, rola w firmie, data końca i początku wykonywania obowiązków, aktualność powiązania.

Po zakończeniu parsowania danych wyniki są eksportowane w formie tabelarycznej do wskazanego katalogu. Dane są gotowe do wykorzystania do generowania mapy powiązań typu BTT / Pająk.

1. **Uwagi. Możliwości poprawy/rozwoju.**

- Error handling, skrypt jest napisany w bardzo prosty sposób, nie przewidziano obsługi ewentualnych błędów, np. nt typu danych, struktury danych