1. **Co to jest? Do czego służy?**  
     
   Skrypt napisany w Pythonie, służy do sprawdzenia **informacji prasowych** **danych rejestrowych** za pomocą APIwywiadowni gospodarczej *EMIS* dla wprowadzonej listy identyfikatorów firm oraz ich nazw. Skrypt umożliwia sprawdzenie informacji prasowych po nazwach firm, jak i filtrowanie poszczególnych słów/fraz występujących w obrębie zdefiniowanego promienia znaków od nazwy firmy. Po pobraniu dane są następnie parsowane do postaci wymaganej do wykorzystania danych do narzędzia FAIT. Dane są eksportowane do pliku xlsx we wskazanej przez użytkownika lokalizacji.
2. **Jak działa?**  
     
   Skrypt sprawdza informacje prasowe dla wprowadzonej przez użytkownika listy identyfikatorów podmiotów, np. NIP/KRS/REGON oraz ich nazw. Jest możliwe także wyszukiwanie poszczególnych słów/fraz występujących w pobliżu nazw weryfikowanych kontrahentów. Po sprawdzeniu dane są następnie przetwarzane, zapisywane jak dataframe i następnie eksportowane do formatu xlsx w oczyszczonej postaci. Wyniki prasówki, oraz znalezione artykuły (badpress) zapisywane są w oddzielnych zakładach. Następnie dane są wykorzystywane przez narzędzie FAIT.

Wybór użytkownika polega na:

1. Zapisania jako zmienną boolean MERGE\_OUTPUTS parametru przyjmującego wartości False/True, który wskazuje na to, czy dane z prasówki mają zostać złączone z pozostałymi danymi pobieranymi do narzędzia FAITCLOUD,
2. Zapisania jako zmienną YEARS liczby lat, dla których wstecz mają zostać sprawdzone informacje prasowe,
3. Zapisania jako zmienną SEARCH25 listy wyszukiwanych słów/fraz, które mają zostać zweryfikowane, czy występują w promieniu 25 słów od nazwy weryfikowanych kontrahentów,
4. Zapisania jako zmienną SEARCH50 listy wyszukiwanych słów/fraz, które mają zostać zweryfikowane, czy występują w promieniu 50 słów od nazwy weryfikowanych kontrahentów,
5. Zapisania jako zmienną *INPUT\_FILE* ścieżki do pliku w formacie .xlsx (łącznie z nazwą pliku), w którym znajduje się lista NIPów podmiotów, dla których pobrane mają zostać dane podmiotów,
6. Zapisania jako zmienną *XLSX\_OUTPUT\_PATH* ścieżki do pliku (łącznie z nazwą pliku), do którego zwrócone mają zostać sparsowane pliki w formacie .xlsx.

W dalszej kolejności skrypt odczytuje wprowadzoną przez użytkownika listę NIP sprawdzanych podmiotów (za pomocą funkcji **get\_nip\_name**) i zapisuje je jako listę. Lista wprowadzonych NIP jest walidowana oraz deduplikowana. Dodatkowo tworzony jest log, przechowujący informacje na temat poprawnych i niepoprawnych NIP.

Następnie skrypt za pomocą funkcji **get\_documents\_links** pobiera dla każdego z poprawnych NIP oraz nazw zwróconych przez funkcję informacje prasowe oraz informacje prasowe sprawdzające występowanie wyszukiwanych fraz.

Następnie za pomocą funkcji **merge\_outputs** dane są łączone, zapisywane w postaci tabeli dataframe i eksportowane w formacie .xslx do wskazanego folderu outputowego. Tworzone są zakładki zawierające odpowiednio:

1. PressResults – całkowity output, zawierający informacje na temat podmiotu, liczby wystąpień nazwy podmiotu w prasie oraz liczby wystąpień wyszukiwanych kluczowych słów/fraz w prasie we wskazanym horyzoncie czasowym,
2. ArticlesResults – informacje na temat podmiotów oraz artykułów wraz z zaznaczonymi znalezionymi słowami kluczowymi, które zostały odnalezione.

Z pobrania tworzony jest log, który w konsoli wyświetla informacje na temat pobierania danych.

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

- rozszerzenie skryptu o sprawdzanie informacji prasowych dla właścicieli i podmiotów zależnych weryfikowanych kontrahentów,

- rozszerzenie skryptu o sprawdzanie informacji prasowych dla poprzednich nazw weryfikowanych kontrahentów