1. **Co to jest? Do czego służy?**  
     
   Skrypt napisany w Pythonie, służy do pobrania **danych rejestrowych oraz informacji prasowych** z APIwywiadowni gospodarczej *EMIS* dla wprowadzonej listy identyfikatorów firm i ich nazw. Dozwolone identyfikatory to: NIP, REGON, KRS. Jako wartość domyślną skrypt wykorzystuje NIP. Po pobraniu dane są zapisywane w formacie JSON a następnie parsowane do postaci wymaganej do wykorzystania danych do narzędzia FAIT.
2. **Jak działa?**  
     
   Skrypt pobiera dane w formacie JSON, dla wprowadzonej przez użytkownika listy podmiotów. Pobrane dane są eksportowane do plików, a następnie przetwarzane, zapisywane w postaci tabelarycznej i następnie eksportowane do formatu xlsx w oczyszczonej postaci. Następnie dane są wykorzystywane przez narzędzie FAIT.

Wybór użytkownika polega na:

1. Zapisania jako zmienną *ID\_CLASS* typu identyfikatorów wprowadzonych podmiotów, dla których mają zostać pobrane dane finansowe. Dozwolone wartości to: 'NIP', 'REGON', 'KRS'. Domyślna wartość to ‘NIP’,
2. Zapisania jako zmienną *YEARS* liczby lat wstecz, dla jakiej mają zostać sprawdzone informacji prasowe dla weryfikowanych podmiotów (domyślnie 10).

Zadeklarowaniu w pliku directories.py:

1. 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,
2. Zapisania jako zmienną *JSON\_OUTPUT\_PATH* ścieżki do katalogu, do którego zwrócone mają zostać pobrane pliki w formacie JSON,
3. Zapisania jako zmienną *XLSX\_OUTPUT\_PATH* ścieżki do pliku (łącznie z nazwą pliku),,do którego zwrócone mają zostać sparsowane pliki w formacie .xlsx.
4. Zapisania jako zmienną *PROJECT\_NAME* nazwy projektu, dla którego z bazu danych cz-prgfts008 mają zostać zaczytane identyfikatory i nazwy sprawdzanych firm.

Proces pobierania danych rejestrowych został opisany w dokumencie *EMIS\_REGISTER\_doc*. Do skryptu została dodana część sprawdzająca, czy we wskazanym horyzoncie czasowym na temat podmiotu istnieją informacje w prasie oraz czy pojawiły się negatywne informacje prasowe w kontekście przestępstw, nadużyć i wyłudzeń podatkowych. Jeśli takie informacje pojawiły się w promieniu 25 i 50 słów od nazwy podmiotu, to zostaną one rozpoznane. Przed rozpoczęciem wyszukiwania nazwa podmiotu zostaje oczyszczone ze wszystkich zbędnych znaków i odpowiednio przygotowana do sprawdzenia, następuje też oczyszczenie nazwy z formy prawnej. Prasa sprawdzana jest we wszystkich możliwych formatach.

Za pomocą funkcji **request\_press\_data** w pętli dla wszystkich nazw przedsiębiorstw tworzone jest zapytanie nt informacji prasowych. Dla każdej firmy wysyłane są 3 zapytania:

1. Czy występują jakiekolwiek informacje prasowe nt podmiotu
2. Czy występują negatywne informacje prasowe nt podmiotu w promieniu 25 słów od jego nazwy
3. Czy występują negatywne informacje prasowe nt podmiotu w promieniu 50 słów od jego nazwy

W przypadku wystąpienia bad pressu zapisywany jest artykuł, informacje na jego temat oraz podkreślane są kluczowe frazy, które zostały zmatchowane. Na podstawie sprawdzonych informacji przeprowadzane są 2 testy:

1. Obecność informacji prasowych,
2. Obecność negatywnych informacji prasowych

Po zakończeniu sprawdzania informacji wyniki są przekształcane do postaci tabelarycznej i łączone z wynikami pobrania danych rejestrowych. Informacje nt artykułów prasowych i wyniki testów są eksportowane do oddzielnych zakładek w pliku xlsx w wskazanym katalogu.

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

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

- dodanie funkcjonalności eksportowania przetworzonych danych do bazy danych