

#SF_CRacker

wyszukiwanie oraz analiza treści

sprawozdań na temat wypłacalności i kondycji finansowej (SFCR)

KONTEKST ROZWIĄZANIA

Celem nadzoru nad rynkiem finansowym jest zapewnienie jego prawidłowego funkcjonowania, stabilności, bezpieczeństwa oraz przejrzystości, a także zapewnienie zaufania do rynku finansowego i ochrony interesów uczestników tego rynku poprzez rzetelną informację na jego temat. UKNF obecnie rozwija nadzór oparty na danych, automatyzując procesy oraz zwiększając wykorzystanie metod data science w analizie dużych wolumenów danych oraz dokumentów nieustrukturyzowanych.

Jednym z rodzajów podmiotów nadzorowanych przez KNF są zakłady ubezpieczeń/reasekuracji (dalej: zakłady). Zakład ubezpieczeń oznacza zakład ubezpieczeń bezpośrednich na życie lub innych niż na życie, który uzyskał zezwolenie na prowadzenie działalności ubezpieczeniowej, natomiast zakład reasekuracji oznacza zakład, który otrzymał zezwolenie na prowadzenie działalności reasekuracyjnej, tj. m.in. działalności polegającej na przyjmowaniu ryzyk cedowanych przez zakład ubezpieczeń lub przez inny zakład reasekuracji.

Przepisy regulujące działalność ubezpieczeniową nakładają na zakłady wiele obowiązków informacyjnych zarówno wobec organu nadzoru, jak i ubezpieczających. Terminowe i kompletne wykonywanie obowiązków informacyjnych jest jednym z podstawowych zadań zakładów i umożliwia uzyskanie powszechnego dostępu do części danych związanych z działalnością zakładów ubezpieczeń/reasekuracji nie tylko organowi nadzoru, ale również nabywcom usług ubezpieczeniowych oraz innym interesariuszom.

Każdy zakład jest zobowiązany do ujawnienia swojego sprawozdania na temat wypłacalności i kondycji finansowej nie później niż w ciągu 14 tygodni po zakończeniu swojego roku obrachunkowego (w przypadku podmiotów działających na terenie Polski rok obrachunkowy pokrywa się z rokiem kalendarzowym, tj. od 1 stycznia do 31 grudnia). Najpopularniejszym sposobem ujawnienia tych informacji jest zamieszczenie SFCR na stronie internetowej.

OPIS I CHARAKTERYSTYKA ZADANIA

Waszym zadaniem jest stworzenie narzędzia (preferowany język programowania - Python) pozwalającego na pozyskiwanie sprawozdań SFCR publikowanych przez zakłady w formie cyfrowej. Narzędzie to powinno być w stanie przekształcić dane z SFCR do formatu umożliwiającego sprawną analizę tych sprawozdań (pobranie i analizę zarówno tekstu, jak i obrazu (wykresy, obrazki) z treści SFCR oraz danych ilościowych z tabel będących załącznikiem SFCR), w tym weryfikację poprawności opublikowanego na stronie SFCR zakładu na podstawie klucza odpowiedzi (zawartości).

Projekt powinien obejmować dostarczenie gotowego skryptu – istotne jest zamieszczanie jak największej ilości komentarzy w kodzie oraz przygotowanie dokumentacji projektowej obejmującej wykorzystane techniki oraz parametry techniczne.

Należy pamiętać, że narzędzie będzie w przyszłości rozwijane przez organ nadzoru, dlatego też powinno być łatwo modyfikowalne do zmieniających się przepisów/wymagań.

OCZEKIWANE REZULTATY

Oczekiwanym rezultatem jest przygotowanie następujących elementów dla krajowych zakładów (kolejność punktów odzwierciedla ich priorytet):

- 1) przygotowanie plików CSV (rozdzielony średnikami, kodowanie CP-1250) zawierających dane opisowe oraz ilościowe z SFCR,
- 2) porównanie treści SFCR pomiędzy wybranymi latami - raport w pliku CSV (rozdzielony średnikami, kodowanie CP-1250),
- 3) przygotowanie plików CSV z weryfikacją kompletności SFCR (na podstawie obecności obligatoryjnych wyrażen w sprawozdaniu SFCR) oraz stworzenie wskaźnika kompletności na podstawie wystąpień słów kluczowych (dla danej sekcji, sekcji nadrzędnej i całego SFCR),
- 4) analiza tekstu SFCR na podstawie kryteriów – wyniki w formie tabeli osadzonej w pliku CSV (rozdzielony średnikami, kodowanie CP-1250).

Frontend powinien zawierać wszystkie funkcjonalności określone w specyfikacji technicznej zadania.

Panel użytkownika powinien umożliwiać m.in.:

- 1) pobranie plików SFCR za poszczególne okresy (możliwość zaznaczenia wszystkich zakładów, kilku zakładów lub tylko jednego zakładu) lub wskazanie lokalnej ścieżki, w której umieszczone są pliki SFCR,
- 2) wybór plików SFCR, które będą ze sobą porównywane/analizowane (możliwość zaznaczenia wszystkich zakładów, kilku zakładów lub tylko jednego zakładu).

Dane zawarte w plikach wynikowych (CSV) powinny być zaprezentowane w przystępnej dla użytkownika formie umożliwiającej analizę na podstawie wybranych przez użytkownika danych. Dodatkowo w panelu wynikowym powinny pojawić się również dane statystyczne (np. liczba pobranych SFCR) i stopień wykonania (np. udział pobranych SFCR do SFCR ogółem).

JAKIE SĄ KORZYŚCI Z TEGO ROZWIĄZANIA?

Projekt ten umożliwi automatyzację wstępnej weryfikacji SFCR poszczególnych zakładów działających na polskim rynku ubezpieczeniowym. Posłuży też jako wsparcie dla analizy wykonywanej na podstawie danych ilościowych. Realizacja projektu ma na celu przyspieszenie wstępnej analizy, zwiększenie kompletności i jakości danych posiadanych przez Urząd Komisji Nadzoru Finansowego, automatyzację analizy dużych wolumenów dokumentów tekstowych oraz sprawdzenie dokumentów publikowanych dla interesariuszy zakładów (w tym klientów).

SPECYFIKACJA TECHNICZNA ZADANIA

Zadanie polega na wyszukaniu i analizie treści SFCR wraz z dokumentami powiązаныmi i składa się z następujących etapów:

- 1) automatyczne odszukanie podstrony ze sprawozdaniami SFCR¹ (pełna nazwa polska: „Sprawozdanie o wypłacalności i kondycji finansowej”, pełna nazwa angielska: „Solvency and financial condition report”, możliwe używanie nazwy „Sprawozdanie na temat wypłacalności i kondycji finansowej” lub formy skróconej „SFCR”) i opiniami biegłego (możliwe słowa kluczowe: Sprawozdanie z badania,

¹ Linki do listy aktywnych (działających) zakładów wraz z linkami do stron internetowych, które sporządziły SFCR za 2022 rok znajduje się w załączniku nr 1.

Sprawozdanie niezależnego biegłego rewidenta z badania, Sprawozdanie o wypłacalności i kondycji finansowej wraz z opinią biegłego rewidenta i raportem z badania, Raport z badania sprawozdania na temat, Sprawozdanie biegłego rewidenta wraz z opinią biegłego rewidenta, Opinia niezależnego biegłego rewidenta, Sprawozdanie audytora z badania sprawozdania, Raport wraz z opinią z badania, Sprawozdanie biegłego rewidenta z badania, Opinia i raport biegłego rewidenta dotyczący, Opinia biegłego rewidenta do Sprawozdania) na stronie zakładu (zakłady mają obowiązek przechowywać SFCR na stronie internetowej przez okres co najmniej 5 lat). Miejsce zapisu pliku SFCR informacji powinno być definiowane przez użytkownika a nazwa powinna zawierać następujące elementy: [rok]_SFCR_[kod zakładu]_[nazwę zakładu]².

- 2) pobranie raportów SFCR. Po pobraniu plików i zapisie ich we wskazanej lokalizacji narzędzie powinno dla każdego zakładu osobno:
 - a. Odczytać zawartość tekstową SFCR, podzielić ją zgodnie z załącznikiem nr 5 i zapisać do pliku CSV. ID oraz nazwy sekcji wskazane w załączniku nr 2 są niezmiennie i obowiązkowe, tzn. należy zawsze dopasować zawartość sekcji do struktury z tego załącznika. Zakład może dodać dowolną ilość sekcji niższego poziomu w ramach sekcji obowiązkowych. Dla każdego zakładu sekcje niższego poziomu niż wskazane w załączniku nr 2, np. A.1.1., powinny mieć nadawane kolejne nowe numery ID i przypisywane nazwy sekcji. Dzięki temu możliwa będzie analiza kompletności zgodnie z załącznikiem nr 3. Narzędzie powinno też identyfikować, że w tekście znajdują się obrazki/grafiki/schematy i umieszczać numer takiego obrazka oraz jego tytuł w danych sekcjach, w miejscu w którym występuje on w tekście. Tabele zawarte w treści SFCR, które nie są stałymi obligatoryjnymi tabelami określonymi w załączniku nr 4, powinny zostać odpowiednio oznaczone w treści w miejscu gdzie występują (numer tabeli, licząc od początku dokumentu), a ich zawartość zapisana w pliku CSV.
 - b. Odnaleźć obligatoryjne tabele wskazane w załączniku nr 4 – tabele te zwykle znajdują się na końcu sprawozdania, lecz mogą występować też wcześniej, wklejone do sekcji. Narzędzie powinno umieć odróżnić je od innych tabel, które mogą być w tekście. Ponadto tabele te mogą zawierać błędy w swojej strukturze, np. w ilości wierszy, kolumn, numerach i nazwach wierszy i kolumn (tutaj przydadzą się techniki uczenia maszynowego do rozpoznawania właściwych tabel i umieszczonych w nich danych).
 - c. Pobrać tabele obligatoryjne stanowiące załącznik do SFCR oraz utworzenie plików CSV z danymi dla każdego zakładu zgodnie z formatkami określonymi w załączniku nr 5. Oznacza to również, że w przypadku błędnych struktur tabel narzędzie powinno próbować rozpoznawać w jaki sposób przyporządkować dane do prawidłowych miejsc w tabelach. Gdy wystąpią braki tabel lub błędy w strukturze tabel, informacja ta powinna odkładać się w osobnym pliku CSV, z oznaczeniem konkretnej tabeli. Trzeba zwrócić uwagę, którego roku dotyczą tabele umieszczone w SFCR (może zdarzyć się, że w ramach tego samego raportu są tabele np. za rok bieżący i za rok poprzedni) i wybrać rok odpowiadający temu, za który jest sporządzony SFCR.
- 3) weryfikacja czy sprawozdanie SFCR jest kompletne tzn. zawiera wszystkie elementy z:
 - a) Rozporządzenia delegowanego Komisji (UE) 2015/35 z dnia 10 października 2014 r. uzupełniającego dyrektywę Parlamentu Europejskiego i Rady 2009/138/WE w sprawie podejmowania i prowadzenia działalności ubezpieczeniowej i reasekuracyjnej (Wypłacalność II),

² Dane znajdują się w załączniku nr 1.

b) Rozporządzenia wykonawczego Komisji (UE) 2015/2452 z dnia 2 grudnia 2015 r. ustanawiającego wykonawcze standardy techniczne w odniesieniu do procedur, formatów i wzorów formularzy dotyczących sprawozdania na temat wypłacalności i kondycji finansowej zgodnie z dyrektywą Parlamentu Europejskiego i Rady 2009/138/WE (Tekst mający znaczenie dla EOG) (2015/2452),

w tym, zweryfikuje:

- datę zamieszczenia SFCR na stronie www zakładu,

Szczegółowe wymagania zostały określone w załączniku nr 2 oraz w załączniku nr 3.

Wynik kompletności powinien być zaprezentowany w pliku CSV, w podziale na poszczególne sekcje oraz całościowo (struktura pliku została zamieszczona w załączniku nr 5).

4) pobranie opinii biegłego rewidenta. Może się zdarzyć, że opinie biegłego rewidenta są w formie skanu. Narzędzie powinno zweryfikować:

- datę opinii audytora oraz czy opinia biegłego rewidenta została wydana z zastrzeżeniem (słowa kluczowe: opinia z zastrzeżeniem, opinia z zastrzeżeniami, podstawa opinii z zastrzeżeniem, opinia negatywna, podstawa opinii negatywnej, niepewność; odmowa wyrażenia opinii; odmowa) czy bez zastrzeżenia.

5) porówna zmiany w opisie ogółem oraz w poszczególnych sekcjach w stosunku do wybranego wcześniejszego roku. Sekcje to struktura sprawozdania SFCR (załącznik nr 2). Narzędzie powinno wskazać jaką część opisu w ramach sekcji jest powtórzeniem tekstu w porównaniu do poprzedniej wybranej przez użytkownika wersji (wskaźnik podobieństwa tekstu).

Wynik porównania powinien zostać zwizualizowany, tak aby dało się odczytać podobieństwa/różnice między sekcjami dla wybranych SFCR.

Zadania dodatkowe po ukończeniu zadania podstawowego:

Zadanie dodatkowe:

6) przygotowanie pobranej treści SFCR dla krajowych zakładów (z tekstu oraz obrazu) do analizy tekstu oraz przygotowanie analizy w zakresie:

- skomplikowania tekstu (długość zdań, wykorzystane słownictwo) – narzędzie powinno weryfikować m.in.:
 - długość zdań, które zawarte są w SFCR (w założeniu im dłuższe zdanie tym bardziej skomplikowana/mniej przejrzysta treść),
 - jak często różne znaki interpunkcyjne występują w tekście w zależności od jego trudności/długości zdania,
 - trudność słów wykorzystywanych w sprawozdaniu (na podstawie częstości jego wykorzystywania).