

Rynki finansowe przetwarzają informacje dostępne poszczególnym uczestnikom rynku poszukując ceny równoważącej popyt z podażą. Naukowe metody badania zmian cen aktywów finansowych pozwalają m. in. na ocenę znaczenia ekonomicznego różnych informacji, ich porównanie i zastosowanie w prognozowaniu. Jednak w miarę postępu technologii cyfrowych zwiększa się zakres i zróżnicowanie dostępnych informacji. Klasyczne metody analizy finansowej wykorzystują niewielką część dostępnych informacji, ponieważ brakuje metod ich integracji. Szczególnym wyzwaniem jest analiza danych niestrukturyzowanych, takich jak teksty zawarte w sprawozdaniach publikowanych obowiązkowo przez spółki giełdowe i instytucje finansowe. Problem jest o tyle istotny, że w ostatnich latach pojawiają się wątpliwości, co do użyteczności informacji dostępnej inwestorom, podczas gdy spółki są zobowiązane do publikowania coraz większej liczby stron tekstu objaśniającego zdarzenia biznesowe, finansowe i społeczne.

Celem naszego projektu jest wyjaśnienie jak informacje tekstowe ewoluują, co wpływa na ich zmiany, oraz jakie jest ich znaczenie dla rynków finansowych. Teksty finansowe nie są tak dobrze zbadane jak liczby, ponieważ ludzie zajmujący się finansami zwracają większą uwagę na liczby. Tekstów jest jednak coraz więcej, więc ich znaczenie rośnie. Nasz projekt proponuje innowacyjne rozwiązania: nowe metody badawcze i nowy sposób zbierania danych. Chcemy zobaczyć jakie wzorce występują w informacjach tekstowych oraz sprawdzić jak motywacje przedsiębiorstw wpływają na ich zmiany. Będziemy chcieli uzyskać konkretne dowody na znaczenie tekstów finansowych dla funkcjonowania rynków.

Wykonanie projektu wymaga połączenia teorii finansów z językoznawstwem oraz wykorzystania komputerowych metod automatycznego przetwarzania tekstu. Z finansów wykorzystujemy teorię przepływu informacji przez rynku, która pokazuje nam jak przetworzyć teksty na sygnały. Stworzymy duży zbiór raportów, który zorganizujemy zgodnie z zasadami konstrukcji korpusów. Z językoznawstwa wykorzystamy teorię pozwalającą nam systematycznie badać teksty. Sięgamy do teorii opisującej sformułowania wartościujące, czyli takie, w których przedsiębiorstwo dzieli się pewnymi opiniami o wynikach finansowych, zdarzeniach czy działaniach. Dzięki tej teorii będziemy mogli zidentyfikować kluczowe elementy tekstu. Z metod komputerowych wykorzystujemy narzędzia, które umożliwiają tworzenie algorytmów przetwarzających tekst. Współpracujemy z centrum CLARIN przy Politechnice Wrocławskiej, które oferuje pełną infrastrukturę do badania języków. Wykorzystujemy też WordNet, czyli dużą bazę danych zawierającą tysiące słów połączonych w grupy znaczeniowe. WordNet to jedna z technik umożliwiających komputerom zrozumienie języka ludzkiego. Ważne dla nas jest to, że WordNety dla różnych języków są połączone, dzięki czemu możemy prowadzić nasze badania na danych z różnych krajów.