

Zbiory tekstów uważane są za cenne źródło informacji, które wykorzystać można podczas modelowania ekonomicznego. Rosnący dostęp do dużych korpusów językowych złożonych z np. abstraktów i artykułów naukowych, wiadomości, treści z mediów społecznościowych lub komunikatów różnych instytucji oraz rozwój metod służących do pozyskiwania informacji z tekstów, wpływają na zainteresowanie tego typu danymi. Jednak wiedza o działaniu metod eksploracji tekstów, w szczególności w połączeniu z tradycyjnymi procedurami ekonometrycznymi jest ciągle ograniczona. Z tego powodu, celem projektu TEXTMOD jest rozwinięcie metod i wiedzy związanych z włączaniem informacji pochodzących z korpusów językowych do modeli ekonometrycznych. Szczególna uwaga skupiona będzie na wielowymiarowych modelach szeregów czasowych. Wskaźniki tekstowe konstruowane będą z użyciem narzędzi służących do identyfikacji najważniejszych tematów poruszanych w dużych zbiorach dokumentów. Przykładem szeregu czasowego uzyskanego na podstawie tekstów, który może mieć zastosowanie w badaniach ekonomicznych i wносить dodatkowe informacje w stosunku do zmiennych realnych, jest trend związany z wybranym tematem opisujący jak istotność tego tematu (np. dotyczącego inflacji) zmieniała się w czasie. Podczas, gdy w ostatnich latach zaproponowano wiele metod modelowania tematów i ich trendów, wiedza o ich właściwościach statystycznych, względnej efektywności i interakcjach z tradycyjnymi podejściami do modelowania jest niewielka. W związku z tym, głównym celem projektu jest zbadanie wrażliwości tych procedur na zmiany wartości parametrów, ich odporności ze względu na modyfikacje próby i niepewności związanej z uzyskanymi wynikami. Zaproponowane zostaną nowe metody służące do porównywania wyników modelowania tematów przeprowadzonych na podstawie różnych prób i z użyciem alternatywnych algorytmów. W kolejnym, ważnym kroku, rozważone zostaną różne procedury wyznaczania trendów związanych z tematami. Następnie, zbadane zostaną konsekwencje uwzględniania trendów w modelach szeregów czasowych np. w popularnym modelu wektorowej autoregresji. Szczególny nacisk położony będzie na odpowiednią interpretację wyników, ocenę korzyści związanych z wykorzystaniem w modelach zmiennych tekstowych i prawidłowy pomiar niepewności dotyczącej uzyskanych ocen parametrów, która przedstawiona zostanie za pomocą pasm ufności. Zaproponowane metody zostaną zastosowane do zbadania zależności pomiędzy zmiennymi ekonomicznymi i trendami związanymi z tematami zidentyfikowanymi dla korpusów naukowych z Polski i Niemiec.