



18.05.2021

#PolskaHumanistykaCyfrowa || Automatyczna analiza fleksyjna tekstów polskich z lat 1830-1918 z uwzględnieniem zmian w odmianie i pisowni

Uniwersytet Warszawski

Celem projektu jest stworzenie analizatora fleksyjnego dla polszczyzny drugiej połowy doby nowopolskiej, co jest wstępnym krokiem dla komputerowej analizy tekstów dawnych. Analizator ten będzie rozpoznawać formy z zachowaniem osobliwości odmiany i pisowni lat 1830-1918. Jego struktura zostanie tak zaprojektowana, by mógł on być podstawą dalszych rozszerzeń o stany wcześniejsze (przede wszystkim o pierwszą połowę doby nowopolskiej). Jednocześnie w ramach projektu, obok analizatora morfologicznego, powstanie nieduży (1 mln segmentów), oznakowany korpus tekstów z lat 1830-1918. Zasób ten zostanie udostępniony wraz z przeszukiwarką według kryteriów fleksyjnych i metatekstowych (periodyzacja próbek, ich przyporządkowanie do podzbiorów stylów funkcjonalnych, składniki opisu bibliograficznego).
