



04.11.2022

Korpus Dyskursu Literaturoznawczego

Korpus Dyskursu Literaturoznawczego (KDL) to jeden z kluczowych zasobów tekstowych opracowywanych przez Instytut Badań Literackich Polskiej Akademii Nauk w ramach projektu DARIAH-PL. Obejmuje on polski dyskurs literaturoznawczy XIX–XXI wieku (lata 1822–2022). Znajdą się w nim teksty poświęcone literaturze lub ukazujące się w czasopiśmie czy antologiach o profilu literaturoznawczym. Szeroki zakres czasowy (200 lat) pozwoli zarówno zgromadzić reprezentatywny materiał, jak i uchwycić procesy kształtowania się dyskursu literaturoznawczego w Polsce.

Korpus zostanie wykorzystany przede wszystkim do utworzenia nowego słownika terminów literackich i trenowania statystycznych modeli językowych. Efektem tych prac będą narzędzia służące do automatycznej ekstrakcji terminologii i bytów (m.in. postaci literackich, nazwisk autorów/ek, badaczy/ek i tłumaczy/ek literatury, tytułów utworów oraz nazw grup literackich i instytucji kultury), określania relacji między rozpoznanymi bytami oraz wskazywania tematyki tekstu. Dodatkowo użytkownicy/czki będą mogli przeszukiwać korpus m.in. za pomocą słów kluczowych, uzyskując dostęp do niewielkich fragmentów tekstów o objętości nieprzekraczającej akapitu. Powstanie również otwarta wersja KDL, zawierająca teksty (głównie starsze), które można udostępnić w całości.

Prace nad pierwszą wersją korpusu zostaną zakończone jesienią 2023 roku. W kolejnych latach KDL będzie rozszerzany o prace najnowsze, a równocześnie pogłębiany poprzez uzupełnienia w postaci wcześniej niedostępnych tekstów.

Osoby zainteresowane korpusem proszone są o kontakt z pracownikami Centrum Humanistyki Cyfrowej IBL PAN:

- Agnieszka Karlińska, agnieszka.karlinska@ibl.waw.pl
- Maciej Maryl, maciej.maryl@ibl.waw.pl.

Prezentacja na temat Korpusu Dyskursu Literaturoznawczego (w formacie .pdf) >>

Biuletyn Polonistyczny

Do pobrania:

- KDL_prezentacja.pdf

Linki:

- <https://biuletynpolonistyczny.pl/pl/projects/cyfrowa-infrastruktura-badawcza-dla-humanistyki-i-nauk-o-sztuce-dariah-pl,1796/details>

Słowa kluczowe: korpus, dyskurs literaturoznawczy, DARIAH-PL