



CYKL WEBINARIÓW

Pracowni Bibliografii Bieżącej Instytutu Badań Literackich PAN

Dane bibliograficzne w badaniach literackich



mgr Karolina Przysiecka
18.05.2021 godz. 13:00-14:30

Przetwarzanie języka naturalnego na potrzeby prac bibliograficznych. Przykład retrokonwersji PBL



mgr Patryk Hubar
15.06.2021 godz. 13:00-14:30

Bibliografia literacka wobec wyzwań literackiego Internetu



dr Paulina Czwordon-Lis
12.10.2021 godz. 13:00-14:30



CYKL WEBINARIÓW

Pracowni Bibliografii Bieżącej Instytutu Badań Literackich PAN

Dane bibliograficzne w badaniach literackich



mgr Karolina Przysiecka
18.05.2021 godz. 13:00-14:30

Dynamiczny rozwój rynku książki powoduje, że nikt nie jest w stanie przeczytać i przestudiować wszystkich książek. Z kolei analiza i interpretacja kanonu literackiego nie pozwala na scharakteryzowanie i zrozumienie systemu literackiego. Aby uzyskać jego pełny obraz, należy przeanalizować zarówno ważne dzieła badane przez literaturoznawców, jak i te, które nie są czytane, komentowane lub zostały zapomniane. To zadanie przekracza możliwości jednego naukowca lub nawet grupy naukowców. Z pomocą przychodzą nowe technologie dające możliwość prowadzenia analiz dużych zbiorów zróżnicowanych danych.

Seminarium będzie poświęcone bibliograficznej bazie danych jako źródle dużych zestawów danych badawczych. Jego celem jest zaprezentowanie możliwości wykorzystania danych bibliograficznych w badaniach literackich i odpowiedź na pytania: Jakie dane gromadzi bibliografia literacka? Jakich danych potrzebują badacze? Na jakie pytania badawcze można odpowiedzieć, wykorzystując dane bibliograficzne? Czy z analizy danych wyłonią się nowe perspektywy badawcze?



CYKL WEBINARIÓW

Pracowni Bibliografii Bieżącej Instytutu Badań Literackich PAN

Przetwarzanie języka naturalnego na potrzeby prac bibliograficznych. Przykład retrokonwersji PBL



mgr Patryk Hubar

15.06.2021 godz. 13:00-14:30

Dane zawarte w bibliografiach, które pierwotnie ukazywały się w formie drukowanej, stanowią niezwykle istotne źródło informacji dziedzinowej, niezbędne z perspektywy prowadzenia wyczerpujących badań naukowych. Niestety, ich obecna forma dystrybucji (skany w formacie PDF) nie pozwala na efektywne wyszukiwanie i wydobywanie informacji dziedzinowej, dlatego też niezbędne jest przeprowadzenie prac nad automatyczną konwersją zapisów bibliograficznych do postaci w pełni przeszukiwalnej bazy danych.

Zastosowanie półautomatycznych metod przetwarzania tekstu opartych o reguły (wyrażenia regularne) oraz prac manualnych, ze względu na stopień skomplikowania oraz objętość bibliografii, jest niewystarczające i nie pozwala na ekstrakcję informacji w wysokiej jakości.



CYKL WEBINARIÓW

Pracowni Bibliografii Bieżącej Instytutu Badań Literackich PAN

Rosnące nakłady pracy nad budowaniem kolejnych reguł oraz ich coraz większy stopień skomplikowania nie przynosiły spodziewanych efektów, a zmienny charakter zapisów bibliograficznych sprawiał, że poszczególne reguły kolidowały ze sobą. Wykorzystanie algorytmów uczenia maszynowego (Machine learning, ML) oraz przetwarzania języka naturalnego (Natural language processing, NLP) i dostosowanie ich do specyfiki konwersji danych bibliograficznych stanowiło więc następny krok w podjętym zadaniu.

W ramach seminarium zaprezentowane zostaną dotychczasowe efekty prac nad retrokonwersją „Polskiej Bibliografii Literackiej” z wykorzystaniem algorytmów uczenia maszynowego oraz przetwarzania języka naturalnego. Uczestnicy zapoznają się z podstawowymi rozwiązaniami NLP dostosowanymi do specyfiki prac nad przetwarzaniem danych bibliograficznych, w szczególności takimi algorytmami, jak klasyfikacja oraz rozpoznawanie jednostek nazewniczych (NER).



CYKL WEBINARIÓW

Pracowni Bibliografii Bieżącej Instytutu Badań Literackich PAN

Bibliografia literacka wobec wyzwań literackiego Internetu



dr Paulina Czwordon-Lis
12.10.2021 godz. 13:00-14:30

Nikogo nie trzeba obecnie przekonywać, że internet jest źródłem unikatowych i nierzadko wartościowych treści literackich i literaturoznawczych, a jego zasięg przewyższa oddziaływanie obiegu papierowego. Również z punktu widzenia bibliografii literackiej internet stanowi przestrzeń, której dokumentacja jawi się już jako konieczność, a jednocześnie wiążą się z nią wysokie wyzwania i dylematy.

W ramach niniejszego seminarium omówimy obszary literackiego internetu szczególnie bogate w treści interesujące z punktu widzenia „Polskiej Bibliografii Literackiej”. Następnie przedyskutujemy korzyści płynące z dokumentacji określonych materiałów internetowych na użytek historii literatury i życia literackiego po 1989 roku oraz teorii literatury. Na koniec przedmiotem dyskusji będą istotne problemy w tworzeniu polskiej bibliografii literackiego internetu: kwestie związane z nietrwałością i koniecznością archiwizowania stron (i możliwościami webarchive), problemy wartościowania treści amatorskich, dylematy związane z komercyjnością, zbieżnością ze źródłami papierowymi, jak również z kwestią praw autorskich. Punktem odniesienia będą przy tym doświadczenia twórców „Czeskiej Bibliografii Literackiej” w ich pracach nad dokumentacją czeskiego internetu literackiego.



CYKL WEBINARIÓW

Pracowni Bibliografii Bieżącej Instytutu Badań Literackich PAN

W celu zapisania się na spotkania prosimy
o wypełnienie krótkiego formularza:



<https://forms.gle/4yQ3uNHZYVne5iTV9>

Kilka godzin przed rozpoczęciem poszczególnych
spotkań otrzymają Państwo wiadomości e-mail
z linkiem i kodem do spotkania.

Serdecznie zapraszamy
Zespół PBB IBL PAN