Biuletyn
Polonistyczny

09.05.2021
Data wydarzenia: 15.06.2021 g.13:00 - 15.06.2021 g.14:30

Patryk Hubar: Przetwarzanie języka naturalnego na potrzeby prac bibliograficznych. Przykład retrokonwersji PBL / z cyklu webinariów Pracowni Bibliografii Bieżącej Instytutu Badań Literackich PAN

Typ wydarzenia: Spotkanie
Organizatorzy: Instytut Badań Literackich PAN

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.

Rosnące nakłady pracy nad budowaniem kolejnych reguł orazich 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 stanowiłowięc następny krokwpodjętym 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 orazrozpoznawanie jednostek nazewniczych (NER).

Webinaria będą nagrywane.

W celu zapisania się na webinaria, prosimy o wypełnienie krótkiego formularza: https://forms.gle/4yQ3uNHZYVne5iTV9.

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

Opłata: bez opłat
Do pobrania:

- Ulotka_cyklu_webinariów_PBB_2021_TC8H1Cc.pdf

Słowa kluczowe: Pracownia Bibliografii Bieżącej, retrokonwersja, przetwarzanie języka naturalnego, Hubar Patryk