

Warsztaty z transfer learningu w NLP (EuroCC)

Report of Contributions

Contribution ID : 1

Type : **not specified**

Wprowadzenie

Monday, 27 September 2021 10:00 (20)

Presenter(s) : PRZEWLOCKI, Pawel (NCBJ)

Contribution ID : 2

Type : **not specified**

Ćwiczenia z tokenizacji

Monday, 27 September 2021 10:20 (30)

Notatnik: using-tokenizer.ipynb

Modele: Roberta (polska i angielska), BERT, itp.

- podstawy działania Google Collab, ładowanie bibliotek, ściąganie modeli
- ładowanie modeli tokenizacji
- rozumienie outputu tokenizatorów
- porównanie działania różnych modeli

Presenter(s) : PRZEWLOCKI, Pawel (NCBJ)

Contribution ID : 3

Type : **not specified**

Ładowanie i używanie modeli

Monday, 27 September 2021 10:50 (30)

Notatnik: using-models.ipynb

Modele: BERT, polska Roberta

- inicjalizacja modeli
- używanie modeli do predykcji wybranych słów
- używanie modeli do weryfikacji, czy zdania są ze sobą powiązane

Presenter(s) : PRZEWLOCKI, Pawel (NCBJ)

Contribution ID : 4

Type : **not specified**

Klasyfikacja zdań za pomocą douczonego wcześniej modelu

Monday, 27 September 2021 11:40 (90)

Notatnik: polish-Roberta-classification-simp.ipynb

Model: polska Roberta

Korpus: <https://clarin-pl.eu/dspace/handle/11321/700>

- przetwarzanie zbioru danych do użycia przy douczaniu modelu
- douczanie modelu, aby rozpoznawał nastawienie emocjonalne zdań z recenzji internetowych
- obliczanie miar skuteczności selekcji pozytywnych recenzji

Możliwe zadania dodatkowe: obserwacja zależności wyników od parametrów uczenia, wielkości zbioru uczącego, itd; próba nauczenia modelu na całych akapitach tekstu

Presenter(s) : PRZEWLOCKI, Pawel (NCBJ)