**  
  
Eksploracja danych tekstowych - laboratoria 6: Tokenizacja tekstu**

*Do realizacji zadań wykorzystaj środowisko jupyter notebook. Po zakończeniu zadań wyślij odpowiedzi na adres michal.zarnecki@cdv.pl.*

1. Wybierz i pobierz zbiór danych zawierających dane tekstowe ze strony: https://www.kaggle.com/datasets?tags=13204-NLP

2. Skorzystaj z notatnika Jupyter "Eksploracja danych tekstowych - Jupyter 5-6 Normalizacja i tokenizacja tekstu.ipynb" i skopiuj kod dotyczący tokenizacji tekstu, następnie zastosuj go na swoim zbiorze danych.  
Wykorzystaj metody tokenizacji:

* bag of words
* TF-IDF
* word embeddings