

Propozycje projektów

Poniżej znajdziecie Państwo propozycje 3 projektów z obszaru NLP:

- Document semantic search
- Aspect based sentiment analysis
- Text generation

Projekty mogą być zrealizowane w wybranej technologii, wynikowy model powinien być dostępny w formie aplikacji dostępnej z poziomu przeglądarki internetowej.

W naszej opinii, zadania mogą być rozwiązywane w małych grupach (2-3 osoby).

Po wstępnej prezentacji, jesteśmy gotowi spotkać się online na konsultacje ze studentami w grudniu.

Kontakt:

mateusz.kroczyński@pl.ey.com

Document semantic search

Zadanie:

Zaimplementuj wyszukiwanie znaczeniowe (semantic search) na zbiorze dokumentów w **języku polskim**. Twój silnik powinien wyszukiwać fragmenty tekstu z dokumentów.

Zbiór danych:

Korpus polskojęzycznej [Wikipedii](#).

Wyniki:

Prezentacja propozycji dotyczących sposobów przygotowania i pre procesowania danych, które mogłyby poprawić wyniki.

Prosta aplikacja webowa implementująca interfejs wyszukiwania, prezentacja użytych rozwiązań oraz otrzymanych wyników i kod użytego modelu.

Aspect based sentiment analysis

Zadanie:

Wykrywanie sentymentu wobec wspomnianych hashtagów we wspomnianych w tweetach w języku polskim. Przypisywanie pozytywnego/neutralnego/negatywnego sentymentu do wypowiedzi dotyczących hashtagów w treści tweeta.

Zbiór danych:

Zbiór tweetów w **języku polskim** pobrany z publicznie dostępnego [API Twittera](#). Model może być trenowany przy użyciu innych korpusów języka polskiego.

Wyniki:

Prezentacja propozycji dotyczących sposobów przygotowania i pre procesowania danych, które mogłyby poprawić wyniki.

Aplikacja webowa zwracająca ogólny sentyment, graficzną prezentację sentymentu w czasie i przykładowe tweety dla danego hashtagu/sentymentu i kod modelu. Prezentacja użytych rozwiązań oraz otrzymanych wyników.

Jakościowa prezentacja wyników dla wybranych 20 hash tagów z następnego slajdu.

Aspect based sentiment analysis

Hashtagi do prezentacji:

#kwarantanna	#vege
#IgaŚwiątek	#hot16challenge
#fitness	#krolowezycia
#kryzys	#ikea
#łódź	#halloween
#kawa	#radom
#karmieniepiersia	#pomidorowa
#COVID19	#nvidia
#poniedziałek	#biedronka

Text generation

Zadanie:

Zbuduj aplikację uzupełniającą zdania w **języku polskim** rozpoczęte przez użytkownika – generator tekstu. Na potrzeby trenowania modelu, można sięgać po inne korpusy.

Zbiór danych:

Korpus dyskursu parlamentarnego. Model może być trenowany przy użyciu innych korpusów języka polskiego.

Wyniki:

Propozycje dotyczące sposobów przygotowania i pre procesowania danych, które mogłyby poprawić wyniki.

Aplikacja generująca tekst i prezentacja dotycząca zastosowanych rozwiązań i wyników.