Pomysły do pracy

1. Topic modeling

Porównać co będzie gdy wytrenuje się word2vec na korpusie przemowien a co gdy wezmie się gotowe z facebooka.

Czy lepsze klastry będą się pojawiać?

1. Procesowanie

- wytrenować model tylko na tych słowach

- segmentacja

- Opis grup

- wizualizacja grup

- określanie tematu na podstawie grup

1. Laguage as a sequence

Generator przemówień – jak dla piosenek. Można zrobić tak,że dla danego tematu mieć różne zbiory prawdopodobieństw!

1. Stopwords

Warto pomyśleć o rozszerzeniu listy stopwords o „sejm” „marszałek”.

Stopwords = [‘Oklaski’, ‘oklaski’,’Panie Marszałku!’, ‘Panie Ministrze!’, ‘Wysoka Izbo’, ‘Pani Marszałek!’, ‘Szanowna Pani Marszałek!‘]