

Домашнее задание 2 (к 23/02/2016). Модель Word2vec

Составьте достаточно большую коллекцию текстов на любом языке или используйте тексты отсюда: <https://www.kaggle.com/c/word2vec-nlp-tutorial/data>.

1. **[1 балла]** Обучите модель word2vec. Оцените время обучения модели, используя модуль time.
2. **[2 балла]** Приведите 5-10 примеров использования `.most_common` для определения близких слов. Корректно ли они найдены? Являются ли синонимами исходного слова?
3. **[2 балла]** Приведите 5-10 примеров использования `.most_common` для определения ассоциаций (А к Б, как В к?). Корректно ли найдены ассоциации?
4. **[2 балла]** Приведите 5-10 примеров использования `.doesnt_match` для определения лишнего слова. Корректно ли найдены лишние слова?
5. **[3 балла]** Попробуйте найти такие пары и тройки слов, для которых не выполняются свойства коммутативности и транзитивности относительно операции определения близких слов.

Обозначим отношение “быть близким” символом \circ .

Коммутативность $x \circ y \implies y \circ x$

Транзитивность $x \circ y, y \circ z \implies x \circ z$