

HW4

22 апреля 2019 г.

```
In [57]: import gensim.downloader as api

In [77]: # загрузка модели Word2Vec из Gensim
         model = api.load('word2vec-ruscorpora-300')

In [90]: # печать наиболее похожих слов (для заданного слова)
         def print_most_similar(word):
             for w, dist in model.most_similar(word):
                 print('{:}: {:.3f}'.format(w, dist))

In [109]: # список рассматриваемых слов
          words = ['кот_NOUN', 'морковь_NOUN', 'ботинок_NOUN', 'красный_ADJ', 'большой_ADJ',
                  'идти_VERB', 'говорить_VERB', 'мало_NUM', 'хорошо_ADV']
          # вывод наиболее похожих слов для каждого слова из заданного списка
          for word in words:
              print('Наиболее похожие для "{:}":\n'.format(word))
              print_most_similar(word)
              print('\n')
```

Наиболее похожие для "кот_NOUN":

```
кошка_NOUN: 0.757
котенок_NOUN: 0.668
пес_NOUN: 0.563
мяукать_VERB: 0.562
тобик_NOUN: 0.559
фоксик_NOUN: 0.557
собака_NOUN: 0.557
мяучать_VERB: 0.554
харлашка_NOUN: 0.552
котяра_NOUN: 0.551
```

Наиболее похожие для "морковь_NOUN":

```
свекла_NOUN: 0.786
капуста_NOUN: 0.756
укроп::петрушка_NOUN: 0.739
```

репчатый::лук_NOUN: 0.736
огурец_NOUN: 0.736
сельдерей_NOUN: 0.735
помидор_NOUN: 0.730
овощ_NOUN: 0.723
картофель_NOUN: 0.722
лук::репчатый_NOUN: 0.722

Наиболее похожие для "ботинок_NOUN":

сапог_NOUN: 0.824
туфля_NOUN: 0.794
башмак_NOUN: 0.789
носок_NOUN: 0.770
брюки_NOUN: 0.726
кед_NOUN: 0.718
полуботинок_NOUN: 0.711
кроссовок_NOUN: 0.705
валенки_NOUN: 0.695
обувь_NOUN: 0.693

Наиболее похожие для "красный_ADJ":

оранжевый_ADJ: 0.660
желтый_ADJ: 0.646
малиновый_ADJ: 0.637
белый_ADJ: 0.629
зеленый_ADJ: 0.612
синий_ADJ: 0.609
красный_NOUN: 0.590
алый_ADJ: 0.576
фиолетовый_ADJ: 0.576
рубиновый_ADJ: 0.556

Наиболее похожие для "большой_ADJ":

огромный_ADJ: 0.763
громадный_ADJ: 0.691
небольшой_ADJ: 0.638
немалый_ADJ: 0.597
большой_ADJ: 0.588
колоссальный_ADJ: 0.579
значительный_ADJ: 0.567
маленький_ADJ: 0.532
меньший_ADJ: 0.530

малый_ADJ: 0.524

Наиболее похожие для "идти_VERB":

пойти_VERB: 0.729
брести_VERB: 0.570
плестись_VERB: 0.526
шагать_VERB: 0.524
тащиться_VERB: 0.522
бежать_VERB: 0.521
двигаться_VERB: 0.521
шествовать_VERB: 0.517
двинуться_VERB: 0.514
ехать_VERB: 0.513

Наиболее похожие для "говорить_VERB":

сказать_VERB: 0.746
заговаривать_VERB: 0.629
разговаривать_VERB: 0.611
думать_VERB: 0.595
толковать_VERB: 0.595
рассуждать_VERB: 0.590
рассказывать_VERB: 0.574
отвечать_VERB: 0.569
возражать_VERB: 0.567
уверять_VERB: 0.566

Наиболее похожие для "мало_NUM":

мало_ADV: 0.722
менее_ADV: 0.661
немногое_PRON: 0.587
среднее_NOUN: 0.583
превышать_VERB: 0.562
примерно_ADV: 0.545
полтора_NUM: 0.527
гораздо_ADV: 0.522
столько_NUM: 0.518
меньший_ADJ: 0.517

Наиболее похожие для "хорошо_ADV":

плохо_ADV: 0.766

отлично_ADV: 0.757
прекрасно_ADV: 0.753
хорошо_PART: 0.743
неплохо_ADV: 0.728
превосходно_ADV: 0.657
недурно_ADV: 0.622
худо_ADV: 0.615
великолепно_ADV: 0.611
недобдеть_VERB: 0.589

```
In [102]: print('Количество рассматриваемых слов:', len(words))
```

Количество рассматриваемых слов: 9

```
In [103]: # вспомогательные функции: голова и хвост списка, взятие третьего элемента кортежа  
# создание списка всех пар слов со значением их близости
```

```
def head(l):  
    return l[0]  
  
def tail(l):  
    return l[1:]  
  
def third(x):  
    return x[2]  
  
def all_pairs(words):  
    res = []  
    if words == []:  
        return res  
    else:  
        word = head(words)  
        others = tail(words)  
        for other in others:  
            res.append((word, other, model.similarity(word, other)))  
    return res + all_pairs(others)
```

```
In [107]: # список всевозможных пар заданных слов со значением их близости  
pairs = all_pairs(words)  
# топ 10 близких пар  
top_pairs = sorted(pairs, key=third, reverse=True)[:10]  
for w1, w2, sim in top_pairs:  
    print('{} и {}: {:.3f}'.format(w1, w2, sim))
```

делать_VERB и говорить_VERB: 0.420
говорить_VERB и хорошо_ADV: 0.373

мало_NUM и хорошо_ADV: 0.360
большой_ADJ и мало_NUM: 0.333
кот_NOUN и ботинок_NOUN: 0.326
делать_VERB и хорошо_ADV: 0.301
красный_ADJ и большой_ADJ: 0.280
большой_ADJ и хорошо_ADV: 0.241
кот_NOUN и говорить_VERB: 0.239
большой_ADJ и говорить_VERB: 0.231