

```

import gensim
from gensim.models import Word2Vec,KeyedVectors

!pip install wget

→ Requirement already satisfied: wget in /usr/local/lib/python3.7/dist-packages (3.2)

!wget -c "https://s3.amazonaws.com/dl4j-distribution/GoogleNews-vectors-negative300.bin.gz"
→ --2021-12-29 08:59:43-- https://s3.amazonaws.com/dl4j-distribution/GoogleNews-vectors-negative300.bin.gz
Resolving s3.amazonaws.com (s3.amazonaws.com)... 52.216.137.230
Connecting to s3.amazonaws.com (s3.amazonaws.com)|52.216.137.230|:443... connected.
HTTP request sent, awaiting response... 416 Requested Range Not Satisfiable

The file is already fully retrieved; nothing to do.

model = KeyedVectors.load_word2vec_format('GoogleNews-vectors-negative300.bin.gz',binary=True,limit=500000)

model['cricket']

→ array([-3.67187500e-01, -1.21582031e-01, 2.85156250e-01, 8.15429688e-02,
       3.19824219e-02, -3.19824219e-02, 1.34765625e-01, -2.73437500e-01,
       9.46044922e-03, -1.07421875e-01, 2.48046875e-01, -6.05468750e-01,
       5.02929688e-02, 2.98828125e-01, 9.57031250e-02, 1.39648438e-01,
      -5.41992188e-02, 2.91015625e-01, 2.85156250e-01, 1.51367188e-01,
     -2.89062500e-01, -3.46679688e-02, 1.81884766e-02, -3.92578125e-01,
     2.46093750e-01, 2.51953125e-01, -9.86328125e-02, 3.22265625e-01,
     4.49218750e-01, -1.36718750e-01, -2.34375000e-01, 4.12597656e-02,
    -2.15820312e-01, 1.69921875e-01, 2.56347656e-02, 1.50146484e-02,
    -3.75976562e-02, 6.95800781e-03, 4.00390625e-01, 2.09960938e-01,
   1.17675781e-01, -4.19921875e-02, 2.34375000e-01, 2.03125000e-01,
  -1.86523438e-01, -2.46093750e-01, 3.12500000e-01, -2.59765625e-01,
  -1.06933594e-01, 1.04003906e-01, -1.79687500e-01, 5.71289062e-02,
 -7.41577148e-03, -5.59082031e-02, 7.61718750e-02, -4.14062500e-01,
 -3.65234375e-01, -3.35937500e-01, -1.54296875e-01, -2.39257812e-01,
 -3.73046875e-01, 2.27355957e-03, -3.51562500e-01, 8.64257812e-02,
 1.26953125e-01, 2.21679688e-01, -9.86328125e-02, 1.08886719e-01,
 3.65234375e-01, -5.66406250e-02, 5.66406250e-02, -1.09375000e-01,
 -1.66992188e-01, -4.54101562e-02, -2.00195312e-01, -1.22558594e-01,
 1.31835938e-01, -1.31835938e-01, 1.03027344e-01, -3.41796875e-01,
 -1.57226562e-01, 2.04101562e-01, 4.39453125e-02, 2.44140625e-01,
 -3.19824219e-02, 3.20312500e-01, -4.41894531e-02, 1.08398438e-01,
 -4.98046875e-02, -9.52148438e-03, 2.46093750e-01, -5.59082031e-02,
 4.07714844e-02, -1.78222656e-02, -2.95410156e-02, 1.65039062e-01,
 5.03906250e-01, -2.81250000e-01, 9.81445312e-02, 1.80664062e-02,
 -1.83593750e-01, 2.53906250e-01, 2.25585938e-01, 1.63574219e-02,
 1.81640625e-01, 1.38671875e-01, 3.33984375e-01, 1.39648438e-01,
 1.45874023e-02, -2.89306641e-02, -8.39843750e-02, 1.50390625e-01,
 1.67968750e-01, 2.28515625e-01, 3.59375000e-01, 1.22558594e-01,
 -3.28125000e-01, -1.56250000e-01, 2.77343750e-01, 1.77001953e-02,
 -1.46484375e-01, -4.51660156e-03, -4.46777344e-02, 1.75781250e-01,
 -3.75000000e-01, 1.16699219e-01, -1.39648438e-01, 2.55859375e-01,
 -1.96289062e-01, -2.57568359e-02, -5.41992188e-02, -2.51464844e-02,
 -1.93359375e-01, -3.17382812e-02, -8.74023438e-02, -1.32812500e-01,
 -2.12402344e-02, 4.33593750e-01, -5.20019531e-02, 3.46679688e-02,
 8.00781250e-02, 3.41796875e-02, 1.99218750e-01, -2.39257812e-02,
 -2.37304688e-01, 1.93359375e-01, 7.32421875e-02, -2.87109375e-01,
 1.25000000e-01, 8.44726562e-02, 1.30859375e-01, -2.19726562e-01,
 -1.61132812e-01, -2.63671875e-01, -5.46875000e-01, -2.96875000e-01,
 3.44238281e-02, -2.87109375e-01, -1.93359375e-01, -1.61132812e-01,
 -3.84765625e-01, -2.14843750e-01, -6.22558594e-03, -1.27929688e-01,
 -1.00097656e-01, -6.21093750e-01, 3.78906250e-01, -4.58984375e-01,
 1.44531250e-01, -9.13085938e-02, -3.08593750e-01, 2.23632812e-01,
 7.86132812e-02, -2.16796875e-01, 8.78906250e-02, -1.66992188e-01,
 1.14746094e-02, -2.53906250e-01, -6.25000000e-02, 6.04248047e-03,
 1.56250000e-01, 4.37500000e-01, -2.23632812e-01, -2.32421875e-01,
 2.75390625e-01, 2.39257812e-01, 4.49218750e-02, -7.51953125e-02,
 5.74218750e-01, -2.61230469e-02, -1.21582031e-01, 2.44140625e-01,
 -3.37890625e-01, 8.59375000e-02, -7.71484375e-02, 4.85839844e-02,
 1.43554688e-01, 4.25781250e-01, -4.29687500e-02, -1.08398438e-01,
 1.19628906e-01, -1.91406250e-01, -2.12890625e-01, -2.87109375e-01,
 -1.14746094e-01, -2.04101562e-01, -2.06298828e-02, -2.53906250e-01,
 8.25195312e-02, -3.97949219e-02, -1.57226562e-01, 1.34765625e-01,
 2.08007812e-01, -1.78710938e-01, -2.00195312e-02, -8.34960938e-02,
 -1.20605469e-01, 4.29687500e-02, -1.94335938e-01, -1.32812500e-01,
 -2.17285156e-02, -2.35351562e-01, -3.63281250e-01, 1.51367188e-01,
 9.32617188e-02, 1.63085938e-01, 1.02050781e-01, -4.27734375e-01,
 2.83203125e-01, 2.74658203e-04, -3.20312500e-01, 1.68457031e-02,
model.most_similar('man')

```

```

→ [('woman', 0.7664012908935547),
 ('boy', 0.6824870109558105),
 ('teenager', 0.6586930155754089),
 ('teenage_girl', 0.6147903800010681),
 ('girl', 0.5921714305877686),
 ('robber', 0.5585119128227234),
 ('teen_agen', 0.5549196600914001),
 ('men', 0.5489763021469116),
 ('guy', 0.5420035123825073),
 ('person', 0.5342026352882385)]
```

```
model.most_similar('cricket')
```

```

→ [('cricketing', 0.8372225165367126),
 ('cricketers', 0.8165745735168457),
 ('Test_cricket', 0.8094818592071533),
 ('Twenty##_cricket', 0.8068488240242004),
 ('Twenty##', 0.7624266147613525),
 ('Cricket', 0.7541396617889404),
 ('cricketer', 0.7372579574584961),
 ('West_Indies_cricket', 0.698798656463623),
 ('Cricket_Board', 0.687838613986969),
 ('ODI', 0.6851663589477539)]
```

```
model.most_similar('facebook')
```

```

→ [('Facebook', 0.7563532590866089),
 ('twitter', 0.6988551616668701),
 ('Twitter', 0.6642444729804993),
 ('Facebook.com', 0.6529868245124817),
 ('facebook.com', 0.6135972142219543),
 ('Twitter.com', 0.6102107763290405),
 ('Myspace', 0.5829181671142578),
 ('Youtube', 0.5796945095062256),
 ('social_networking', 0.5592854022979736),
 ('social_networking_sites', 0.5527533292770386)]
```

```
model.similarity('man', 'woman')
```

```
→ 0.76640123
```

```
model.similarity('man', 'PHP')
```

```
→ -0.032995153
```

```
model.doesnt_match(['PHP', 'java', 'monkey'])
```

```

→ /usr/local/lib/python3.7/dist-packages/gensim/models/keyedvectors.py:895: FutureWarning: arrays to stack must be passed as a "sequence"
      vectors =.vstack(self.word_vec(word, use_norm=True) for word in used_words).astype(REAL)
      'monkey'
```

```
◀ ▶
```

```
vec = model['king'] - model['man'] + model['woman']
model.most_similar([vec])
```

```

→ [('king', 0.8449392318725586),
 ('queen', 0.7300517559051514),
 ('monarch', 0.6454660892486572),
 ('princess', 0.6156251430511475),
 ('crown_prince', 0.5818676948547363),
 ('prince', 0.5777117609977722),
 ('kings', 0.5613663792610168),
 ('sultan', 0.5376776456832886),
 ('queens', 0.5289887189865112),
 ('ruler', 0.5247419476509094)]
```

```
vec = model['INR'] - model['India'] + model['England']
model.most_similar([vec])
```

```

→ [('INR', 0.6442340612411499),
 ('GBP', 0.5040826201438904),
 ('England', 0.44649264216423035),
 ('£', 0.4334099888015747),
 ('Â_£', 0.430719792842865),
 ('£_#.##m', 0.42561304569244385),
 ('Pounds_Sterling', 0.42512616515159607),
 ('GBP##', 0.42464494705200195),
 (''stg', 0.42324796319007874),
 ('£_#.##m', 0.4201711118221283)]
```

Start coding or [generate](#) with AI.