

```
import gensim.downloader as api
wv = api.load('word2vec-google-news-300')
```

```
[=====] 100.0% 1662.8/1662.8MB downloaded
```

```
!python -m spacy download en_core_web_lg
```

```
2023-04-12 17:24:13.745947: W tensorflow/compiler/tf2tensorrt/utils/py_utils.cc:38] TF-TRT Warning: Could not find TensorRT
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting en-core-web-lg==3.5.0
  Downloading https://github.com/explosion/spacy-models/releases/download/en\_core\_web\_lg-3.5.0/en\_core\_web\_lg-3.5.0-py3-none-any.whl
    587.7/587.7 MB 2.4 MB/s eta 0:00:00
Requirement already satisfied: spacy<3.6.0,>=3.5.0 in /usr/local/lib/python3.9/dist-packages (from en-core-web-lg==3.5.0) (3.5.1)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in /usr/local/lib/python3.9/dist-packages (from spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (3.2.0)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.9/dist-packages (from spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (2.0.2)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.9/dist-packages (from spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (4.38.0)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /usr/local/lib/python3.9/dist-packages (from spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (0.9.1)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /usr/local/lib/python3.9/dist-packages (from spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (3.0.11)
Requirement already satisfied: setuptools in /usr/local/lib/python3.9/dist-packages (from spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (57.0.0)
Requirement already satisfied: packaging<20.0 in /usr/local/lib/python3.9/dist-packages (from spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (20.0)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3.9/dist-packages (from spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (1.0.0)
Requirement already satisfied: smart-open<7.0.0,>=5.2.1 in /usr/local/lib/python3.9/dist-packages (from spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (5.2.1)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.9/dist-packages (from spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (2.13.0)
Requirement already satisfied: numpy<1.15.0 in /usr/local/lib/python3.9/dist-packages (from spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (1.15.0)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.9/dist-packages (from spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (0.28.0)
Requirement already satisfied: Jinja2 in /usr/local/lib/python3.9/dist-packages (from spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (2.11.3)
Requirement already satisfied: typer<0.8.0,>=0.3.0 in /usr/local/lib/python3.9/dist-packages (from spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (0.3.0)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.9/dist-packages (from spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (3.0.2)
Requirement already satisfied: thinc<8.2.0,>=8.1.8 in /usr/local/lib/python3.9/dist-packages (from spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (8.1.8)
Requirement already satisfied: pathy<0.10.0 in /usr/local/lib/python3.9/dist-packages (from spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (0.10.0)
Requirement already satisfied: pydantic!=1.8,!=1.8.1,<1.11.0,>=1.7.4 in /usr/local/lib/python3.9/dist-packages (from spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (1.10.2)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /usr/local/lib/python3.9/dist-packages (from spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (2.0.6)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /usr/local/lib/python3.9/dist-packages (from spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (2.4.3)
Requirement already satisfied: typing-extensions<4.2.0 in /usr/local/lib/python3.9/dist-packages (from spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (4.1.1)
Requirement already satisfied: charset-normalizer<2.0.0 in /usr/local/lib/python3.9/dist-packages (from requests<3.0.0,>=2.13.0->spacy<3.6.0,>=3.5.0) (1.0.3)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.9/dist-packages (from requests<3.0.0,>=2.13.0->spacy<3.6.0,>=3.5.0) (2.10)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /usr/local/lib/python3.9/dist-packages (from requests<3.0.0,>=2.13.0->spacy<3.6.0,>=3.5.0) (1.26.13)
Requirement already satisfied: certifi<2017.4.17 in /usr/local/lib/python3.9/dist-packages (from requests<3.0.0,>=2.13.0->spacy<3.6.0,>=3.5.0) (2017.4.17)
Requirement already satisfied: blis<0.8.0,>=0.7.8 in /usr/local/lib/python3.9/dist-packages (from thinc<8.2.0,>=8.1.8->spacy<3.6.0,>=3.5.0) (0.7.8)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /usr/local/lib/python3.9/dist-packages (from thinc<8.2.0,>=8.1.8->spacy<3.6.0,>=3.5.0) (0.0.1)
Requirement already satisfied: click<9.0.0,>=7.1.1 in /usr/local/lib/python3.9/dist-packages (from typer<0.8.0,>=0.3.0->spacy<3.6.0,>=3.5.0) (7.1.1)
Requirement already satisfied: MarkupSafe<2.0 in /usr/local/lib/python3.9/dist-packages (from Jinja2->spacy<3.6.0,>=3.5.0) (2.0.1)
Installing collected packages: en-core-web-lg
Successfully installed en-core-web-lg-3.5.0
✓ Download and installation successful
You can now load the package via spacy.load('en_core_web_lg')
```

```
import spacy
```

```
nlp = spacy.load("en_core_web_lg")
doc = nlp("dog cat banana kalliiii")
```

```
for token in doc:
    print(token.text, "Vector:", token.has_vector, "OOV:", token.is_oov)
```

```
dog Vector: True OOV: False
cat Vector: True OOV: False
banana Vector: True OOV: False
kalliiii Vector: False OOV: True
```

```
#gensim
wv.similarity(w1="great", w2="good")
```

```
0.729151
```

```
#spacy
def print_similarity(base_word, words_to_compare):
    base_token = nlp(base_word)
    doc = nlp(words_to_compare)
    for token in doc:
        print(f"{token.text} <-> {base_token.text}: ", token.similarity(base_token))
    print_similarity("great", "good")
```

```
good <-> great: 0.7551303681455304
```

```
#gensim
wv.most_similar("good")
```

```

[('great', 0.7291510105133057),
 ('bad', 0.7190051078796387),
 ('terrific', 0.6889115571975708),
 ('decent', 0.6837348341941833),
 ('nice', 0.6836092472076416),
 ('excellent', 0.644292950630188),
 ('fantastic', 0.6407778263092041),
 ('better', 0.6120728850364685),
 ('solid', 0.5806034803390503),
 ('lousy', 0.576420247554779)]

wv.most_similar("dog")

[('dogs', 0.8680489659309387),
 ('puppy', 0.8106428384780884),
 ('pit_bull', 0.780396044254303),
 ('pooch', 0.7627376914024353),
 ('cat', 0.7609457969665527),
 ('golden_retriever', 0.7500901818275452),
 ('German_shepherd', 0.7465174198150635),
 ('Rottweiler', 0.7437615394592285),
 ('beagle', 0.7418621778488159),
 ('pup', 0.740691065788269)]

wv.most_similar(positive=['king', 'woman'], negative=['man'], topn=5)

[('queen', 0.7118193507194519),
 ('monarch', 0.6189674139022827),
 ('princess', 0.5902431011199951),
 ('crown_prince', 0.5499460697174072),
 ('prince', 0.5377321839332581)]

wv.most_similar(positive=['france', 'berlin'], negative=['paris'], topn=5)

[('germany', 0.5094343423843384),
 ('european', 0.48650455474853516),
 ('german', 0.4714890420436859),
 ('austria', 0.46964022517204285),
 ('swedish', 0.4645182490348816)]

wv.doesnt_match(["facebook", "cat", "google", "microsoft"])

'cat'

wv.doesnt_match(["dog", "cat", "google", "mouse"])

'google'

king = nlp.vocab["king"].vector
man = nlp.vocab["man"].vector
woman = nlp.vocab["woman"].vector
queen = nlp.vocab["queen"].vector

result = king - man + woman
from sklearn.metrics.pairwise import cosine_similarity

cosine_similarity([result], [queen])

array([[0.61780137]], dtype=float32)

glv = api.load("glove-twitter-25")

[=====] 100.0% 104.8/104.8MB downloaded

glv.most_similar("good")

[('too', 0.9648017287254333),
 ('day', 0.9533665180206299),
 ('well', 0.9503170847892761),
 ('nice', 0.9438973665237427),
 ('better', 0.9425962567329407),
 ('fun', 0.9418926239013672),
 ('much', 0.9413353800773621),
 ('this', 0.9387555122375488),
 ('hope', 0.9383506774902344),
 ('great', 0.9378516674041748)]

```

```
glv.doesnt_match("breakfast cereal dinner lunch".split())
```

```
'cereal'
```

```
glv.doesnt_match("facebook cat google microsoft".split())
```

```
'cat'
```

```
glv.doesnt_match("banana grapes orange human".split())
```

```
'human'
```

