In [1]:

```
from gensim.models import FastText
from gensim.models.word2vec import LineSentence
import multiprocessing

WIKI_SEG_TXT = "wiki_seg.txt"

max_cpu_counts = multiprocessing.cpu_count()
word_dim_size = 300

sentences = LineSentence(WIKI_SEG_TXT)

model = FastText(sentences, vector_size=word_dim_size, workers=max_cpu_counts)

output_model = f"fasttext.zh.{word_dim_size}.model"
model.save(output_model)

print(model.wv.vectors.shape)
model.wv.vectors

model = FastText.load('fasttext.zh.300.model')
vocab = model.wv.key_to_index

print(f"總共收錄了 {len(vocab)} 個詞彙")

print("印出 20 個收錄詞彙:")
print(list(vocab.keys())[:20])
```

```
(1281108, 300)
總共收錄了 1281108 個詞彙
印出 20 個收錄詞彙:
['年', '月', '日', '中', '10', '12', '11', '小行星', '中國', '時', '-', '日
本', '美國', '20', '香港', '臺灣', '15', '位於', '30', '站']
```

In [2]:

```
vec = model.wv['數學家']
print(vec.shape)
vec
```

```
(300,)
```

Out[2]:

In [3]:

```
word = "這肯定沒見過 "

# 若強行取值會報錯
try:
    vec = model.wv[word]
except KeyError as e:
    print(e)
```

In [4]:

```
model.wv.most_similar("飲料", topn=10)
```

Out[4]:

```
[('輝劍', 0.9708890914916992),
 ('名松', 0.9569700360298157),
 ('飲料類', 0.9305379986763),
 ('飲料機', 0.9210789799690247),
 ('飲料罐', 0.9001978635787964),
 ('軟飲料', 0.887269139289856),
 ('經米濱', 0.8780335783958435),
 ('茶飲料', 0.8691356182098389),
 ('飲品', 0.846047580242157),
 ('廣慈宮', 0.797712504863739)]
```

In [5]:
```
model.wv.most_similar("car")
```

Out[5]:
```
array([-2.40173921e-01, -1.11621618e-03, -1.26872182e+00,  6.97947383e-01,
       -1.07644582e+00, -6.57272637e-01, -2.25588131e+00, -7.36904502e-01,
        7.43719127e-01, -6.31970823e-01,  1.29935861e+00,  6.28995061e-01,
        1.59047210e+00,  6.71653748e-01, -7.50802875e-01,  1.89570022e+00,
        1.70877051e+00, -1.40614212e+00,  8.60415339e-01, -3.44749284e+00,
       -1.06719100e+00, -6.77665651e-01,  5.80408573e-01, -2.30039167e+00,
        1.19869712e+00,  9.24629271e-01,  4.42754894e-01,  1.17526305e+00,
        6.15279675e-01,  2.01202065e-01, -2.06836200e+00, -2.81031775e+00,
        8.38599268e-01,  6.31966512e-01, -1.56953883e+00, -1.37281775e+00,
       -7.60987341e-01,  6.63723284e-03,  9.61589813e-03,  5.70674241e-01,
        1.66490090e+00, -2.67470741e+00, -1.38995543e-01, -8.97718146e-02,
       -4.08933594e-01, -2.78019118e+00,  5.66998899e-01,  3.60473722e-01,
        4.64375496e-01,  3.86218280e-01, -3.82440257e+00, -5.71132042e-02,
       -4.35999593e-02,  1.75573319e-01,  8.23337913e-01,  1.77245364e-01,
       -1.38151395e+00,  1.44025218e+00,  2.73685193e+00, -6.67776048e-01,
       -8.69310871e-02, -1.98473620e+00, -1.77337909e+00, -2.42071271e+00,
        7.91258693e-01, -1.89115778e-01,  1.90285707e+00, -8.52158010e-01,
        6.32575572e-01, -1.63945884e-01,  9.86310959e-01, -1.97316635e+00,
        6.89252079e-01,  1.05064309e+00, -1.24747765e+00, -1.13373041e+00,
       -2.07794848e+00,  1.60250009e+00,  8.72203767e-01, -2.88194537e+00,
        4.51874733e-03, -6.11168981e-01,  1.51703703e+00, -2.97091174e+00,
        1.54577121e-01,  2.19664288e+00,  1.51999366e+00, -4.36329037e-01,
        1.01668239e+00, -3.19171458e-01, -3.18268865e-01,  8.07612240e-01,
       -4.88891554e+00,  2.64731336e+00, -1.38393426e+00,  1.84921968e+00,
       -9.35060501e-01,  5.24898887e-01, -1.44793165e+00, -3.41898012e+00,
        8.64959061e-01,  1.30871069e+00, -6.34120405e-01,  4.83684391e-01,
        2.37275052e+00,  4.42978591e-01, -1.28958905e+00,  2.66518998e+00,
        1.59283292e+00,  1.21687067e+00,  1.85326111e+00, -1.32756555e+00,
       -1.75584817e+00,  2.43160176e+00, -1.34762883e-01,  8.75572324e-01,
       -4.23799562e+00, -3.93460965e+00, -2.45714355e+00,  3.48230511e-01,
        6.10693634e-01,  2.01684546e+00,  5.46876013e-01, -5.55703975e-02,
        2.02441716e+00,  1.43272150e+00,  1.28788662e+00,  5.16789615e-01,
       -1.12526464e+00, -7.34690130e-01,  2.43126822e+00, -1.21860003e+00,
       -2.18015742e+00,  6.71792150e-01,  1.22030127e+00,  9.92207348e-01,
        2.90833330e+00, -1.21260059e+00,  7.89628386e-01,  1.06371987e+00,
       -1.82758462e+00,  1.92686990e-01,  2.14815760e+00, -2.64312953e-01,
        4.05896574e-01,  6.95940316e-01,  2.17425689e-01, -1.11351490e+00,
       -8.86758745e-01, -3.81736130e-01,  2.17350745e+00,  7.25531429e-02,
        4.34405535e-01,  3.13358808e+00,  1.46834159e+00,  2.05648232e+00,
        3.29993820e+00,  8.25851023e-01,  1.59554923e+00, -4.10528231e+00,
        9.53241467e-01, -3.05898802e-01, -1.07617736e+00, -1.50403726e+00,
        1.45217478e+00, -1.47536635e+00, -2.82843858e-01,  1.30832744e+00,
        1.39555168e+00, -1.41230956e-01,  1.61369324e+00, -7.41366029e-01,
       -1.65646732e+00, -1.52641326e-01, -1.68782938e+00,  1.66879940e+00,
       -7.90542305e-01,  3.22696614e+00, -1.72381973e+00,  9.37703717e-03,
        6.35905087e-01,  2.03491998e+00,  1.42329693e-01,  2.78502512e+00,
        8.72519076e-01, -5.82000315e-01,  2.32710576e+00, -1.78275990e+00,
        2.06135035e+00,  1.56157804e+00,  1.60534620e-01,  7.19626427e-01,
        7.50109851e-01,  2.29789987e-01, -5.42536616e-01, -1.82744801e+00,
       -2.77670121e+00, -4.16845530e-01,  1.96763837e+00, -7.03205943e-01,
       -1.55459511e+00, -1.69351652e-01, -1.05798459e+00,  2.66489983e+00,
       -1.10011645e-01, -3.40474755e-01, -2.57036233e+00,  1.03224409e+00,
        3.29389954e+00,  3.40971261e-01,  1.80962372e+00,  1.68237662e+00,
        8.88098963e-03, -1.17021151e-01, -5.03221214e-01,  4.47238386e-01,
        1.61854577e+00,  1.88970578e+00, -2.60758901e+00,  1.19623888e+00,
       -5.79448044e-01,  1.32295394e+00, -8.12097669e-01, -9.23748910e-01,
        2.70320082e+00,  1.30005896e+00,  1.14306772e+00,  5.78364670e-01,
       -5.36683621e-03,  1.24932110e+00, -5.76023102e-01,  3.55230880e+00,
       -3.41295511e-01, -1.17136073e+00,  4.16124254e-01, -1.87512577e+00,
       -5.55012561e-02,  2.52681088e+00,  1.17661476e+00, -1.10216725e+00,
        4.50842202e-01, -3.08442569e+00, -1.47741839e-01,  7.78728902e-01,
```

In [6]:
```
model.wv.most_similar("facebook")
```

Out[6]:
```
[('youtubefacebook', 0.9263174533843994),
 ('thefacebook', 0.9004828929901123),
 ('facebookpage', 0.8896870613098145),
 ('facebox', 0.8654385805130005),
 ('instagram', 0.8053536415100098),
 ('twitteryoutube', 0.7763373851776123),
 ('twitter', 0.7609679102897644),
 ('googleyoutube', 0.7585869431495667),
 ('youtube', 0.7452121973037722),
 ('lnstagram', 0.7385021448135376)]
```

In [7]:
```
model.wv.most_similar("詐欺")
```

Out[7]:
```
[('賈邱', 0.8894781470298767),
 ('赤坑鎮', 0.8356450796127319),
 ('中境', 0.8105662465095552),
 ('越中境', 0.7931427955627441),
 ('詐欺罪', 0.7840371727943420),
 ('他魚', 0.7666366696357727),
 ('欺詐', 0.7374577522277832),
 ('抱出', 0.7082515954971313),
 ('欺詐案', 0.6659218072891235),
 ('義德堂', 0.6421015858650208)]
```

In [8]:
```
model.wv.most_similar("合約")
```

Out[8]:
```
[('德康', 0.7152364349365234),
 ('合同', 0.6814538836479187),
 ('綠蠅', 0.6765901446342468),
 ('合同額', 0.6385767459869385),
 ('合同期', 0.6291205883026123),
 ('合同商', 0.6239711046218872),
 ('籤合同', 0.6218411326408386),
 ('合同制', 0.6184295415878296),
 ('簽約', 0.6969977617263794),
 ('合同條款', 0.6926885843467407)]
```

In [9]:
```
model.wv.similarity("連結", "鍵接")
```

Out[9]:
```
0.42695913
```

In [10]:
```
model.wv.similarity("連結", "陰天")
```

Out[10]:
```
-0.03664172
```

In [11]:
```
print(f"Loading {output_model}...")
new_model = FastText.load(output_model)
```

```
Loading fasttext.zh.300.model...
```

In [12]:
```
model.wv.similarity("連結", "陰天") == new_model.wv.similarity("連結", "陰天")
```

Out[12]:
```
True
```

In [ ]:
```

```