In [1]:
```
WIKI_SEG_TXT = "wiki_seg.txt"
```

In [2]:
```
from gensim.models import FastText
from gensim.models.word2vec import LineSentence
import multiprocessing

max_cpu_counts = multiprocessing.cpu_count()
word_dim_size = 300  # 设置word vector维度
print(f"Use {max_cpu_counts} workers to train FastText (dim={word_dim_size})")

WIKI_SEG_TXT = "wiki_seg.txt"

# 读取训练语句
sentences = LineSentence(WIKI_SEG_TXT)

# 训练模型

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

# 保存模型
output_model = f"fasttext.zh.{word_dim_size}.model"
model.save(output_model)
```

```
Use 12 workers to train FastText (dim=300)
```

In [3]:
```
print(model.wv.vectors.shape)
model.wv.vectors
```

```
(1281108, 300)
```

Out[3]:
```
array([[-2.1143034 , -3.5834088 ,  0.9235744 , ..., -2.9513686 ,
         0.85165715,  0.7485893 ],
       [ 0.44912398,  1.157038  ,  0.17390107, ..., -1.5394628 ,
        -2.0758872 ,  2.3475122 ],
       [-4.7415714 ,  4.1082935 ,  3.0406713 , ..., -0.18160735,
         0.97919846, -3.1430063 ],
       ...,
       [-0.42775387, -0.4923242 ,  0.3848211 , ...,  0.34394506,
         0.08799452, -0.6257678 ],
       [ 0.3650192 , -0.02768007,  0.23861456, ...,  0.26748458,
         0.39309955,  1.2193569 ],
       [-0.14890154, -0.08819858, -0.0111459 , ...,  0.13148315,
        -0.0556561 , -0.13519895]], dtype=float32)
```

In [5]:
```
# 加载并训练FastText模型
model = FastText.load('fasttext.zh.300.model')  # 替换成您的模型路径
vocab = model.wv.key_to_index

# 打印总共收录的词汇数
print(f"總共收錄了 {len(vocab)} 個詞彙")

# 打印前20个收录的词汇
print("印出 20 個收錄詞彙:")
print(list(vocab.keys())[:20])
```

```
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 [6]:
```
vec = model.wv['數學家']
print(vec.shape)
vec
```

```
(300,)
```

Out[6]:
```
array([-0.36248013,  0.98012364,  0.55402476,  0.4230872 , -0.08536739,
        0.37592587, -1.7252742 ,  0.35973006,  1.0355555 ,  0.20347224,
        0.89693195,  2.3107479 ,  1.1560698 ,  0.43550915, -1.8510116 ,
        1.8772343 ,  2.1570227 , -0.8266401 , -0.4693025 , -2.9896805 ,
       -0.76506203,  0.55268776, -0.6439924 , -2.1026824 ,  0.82134354,
       -0.03820011,  1.9603138 ,  3.5589116 ,  1.4944109 , -1.1495035 ,
       -2.2098894 , -2.3953311 , -0.26409993,  0.7407733 ,  0.37456003,
       -3.1395462 ,  0.02966588,  1.4000149 , -0.7277578 ,  0.29956517,
        0.23028906, -3.5200157 , -1.7080253 ,  1.3862095 , -0.80813515,
       -2.0764017 ,  1.3242884 ,  0.78814757,  0.70669633,  0.8410793 ,
       -2.373359  ,  0.17978494, -1.7763336 , -0.5197104 ,  0.02317583,
       -0.61163217, -0.24958536,  1.333366  ,  0.6208196 ,  0.62406814,
        0.20208071, -0.8206698 , -1.3902856 , -1.7098118 ,  0.6752405 ,
        2.733623  ,  1.2758677 , -0.6498876 ,  0.2879006 ,  0.38086182,
        0.99583673, -0.8345073 , -0.37709767,  0.33063385, -0.92941964,
        0.00550079,  0.5862416 ,  0.87920475,  0.52489877, -3.6531105 ,
        0.4016076 , -0.6659497 ,  0.82306755, -3.8803885 , -1.2592509 ,
        1.3713334 ,  1.7969382 , -0.40232936, -0.06662682, -1.0427252 ,
        1.8947021 ,  0.33118635, -3.2141137 ,  0.13048308, -2.2011406 ,
        0.77231276,  1.3026292 ,  1.8185749 , -0.91952515, -1.3727586 ,
        1.6142715 , -0.35515103, -1.1541601 ,  0.8725955 ,  2.8665287 ,
       -1.7573503 ,  0.9622064 ,  2.2132113 ,  2.2489228 ,  1.295852  ,
        0.41261426,  1.413129  , -0.97843295,  0.48973542,  1.3581734 ,
        0.71574175, -3.1343539 , -2.4122615 , -2.9045713 ,  0.8909527 ,
       -1.7113088 ,  1.910596  ,  1.9032061 ,  1.8908036 ,  1.0927012 ,
        1.2950226 ,  1.3321568 ,  0.9011788 , -0.04983101, -1.31874   ,
        1.4708176 ,  0.09124085, -2.5455387 ,  1.1389534 ,  2.9139862 ,
        1.9673725 ,  2.7447848 , -1.3354753 ,  3.1814198 ,  0.14665474,
       -1.8544286 ,  0.8223278 ,  3.2614264 , -0.8112384 , -0.09817249,
        1.4645436 ,  0.75998557, -0.8995562 ,  0.11551278,  1.5027435 ,
        1.8521479 ,  1.0239952 ,  1.5470933 ,  1.8559628 ,  1.3798646 ,
        3.0782177 ,  4.1210227 ,  1.4251592 ,  0.9985844 , -2.2082667 ,
       -2.6183095 , -1.1970755 , -3.152174  , -2.6965501 ,  0.17867514,
       -0.9372266 , -0.32081595, -1.0925673 ,  2.88764   ,  0.38610265,
        0.05830714, -1.0056459 ,  1.7154334 , -1.44236   , -0.5403752 ,
        0.42215803,  1.2282223 ,  3.1132028 , -1.5636071 ,  0.40505138,
        1.432704  ,  2.3415618 ,  2.3565447 ,  2.7195437 ,  0.8814994 ,
       -0.4873589 ,  1.1314331 ,  0.57285845,  0.74154717,  0.41416332,
        0.7483571 ,  0.31172642,  0.6211714 ,  0.21189976,  1.6088046 ,
        0.5581938 , -1.8051902 , -1.5454528 ,  2.6511977 , -0.30181536,
       -1.6230351 ,  0.8690149 , -0.6757468 ,  2.7721062 ,  1.1747085 ,
```

```
 1.4936074 , -1.3174665 , -0.05859082,  3.9109533 ,  0.04994953,
 0.3170531 ,  1.0511549 , -0.42471218, -0.27838972,  1.7440172 ,
 0.719544  ,  0.64353424, -0.34134546, -1.178439  , -0.9027263 ,
-2.2345853 ,  1.2022699 ,  1.0734307 , -0.41167647,  2.8903198 ,
 1.2145857 ,  0.11285645, -2.2847824 ,  0.1919113 ,  0.16430385,
 0.9349448 ,  1.0600609 ,  0.52170885, -0.38075987,  2.165146  ,
-2.3794723 , -0.91174155,  1.0703781 ,  2.0973678 ,  0.55109   ,
 0.52043056, -0.31678596, -0.1718689 ,  2.059049  , -0.04883113,
 2.1690748 ,  3.3277364 ,  1.179526  , -1.1479416 , -0.26288223,
-1.8426039 ,  2.4506278 , -0.9216002 , -0.31827846, -2.2192037 ,
-0.6037803 ,  1.1627191 , -0.30118388,  1.6683006 ,  1.8911893 ,
 2.2625875 ,  0.31399542, -0.2117326 ,  0.24281022,  3.30119   ,
 2.025536  ,  0.39673144,  0.04100621,  0.77655894, -0.6997056 ,
-0.37861657,  0.8386439 , -1.5636232 ,  0.66094923, -0.30948257,
 2.08865   ,  1.3413908 ,  0.11441649,  0.10324325,  0.53062344,
-0.269244  , -0.46463028,  1.6779274 , -3.0914824 , -1.921438  ,
 1.3992082 , -0.23602292,  0.7996966 , -1.2923689 ,  0.68264467,
-2.070288  , -0.6330929 ,  0.01733787,  0.82230055, -2.3566875 ,
 1.3539721 , -1.0342708 ,  0.3492705 ,  1.1793842 ,  0.5706733 ],
dtype=float32)
```

In [7]:

```python
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 [8]:
```
model.wv.most_similar("飲料", topn=10)
```

Out[8]:
```
[('輝劍', 0.9710693359375),
 ('名松', 0.9501043558120728),
 ('飲料類', 0.9315357208251953),
 ('飲料機', 0.9279479384422302),
 ('飲料罐', 0.8953368663787842),
 ('軟飲料', 0.8831291198730469),
 ('茶飲料', 0.8725141882896423),
 ('經米濱', 0.8699420690536499),
 ('飲品', 0.8453800082206726),
 ('飲料瓶', 0.7923927307128906)]
```

In [9]:
```
model.wv.most_similar("car")
```

Out[9]:
```
[('hcar', 0.8572422862052917),
 ('carcar', 0.8509683012962341),
 ('ccar', 0.849014401435852),
 ('jetcar', 0.8113610148429871),
 ('tramcar', 0.8039993643760681),
 ('zipcar', 0.8031772971153259),
 ('motorcar', 0.8004679083824158),
 ('boxcar', 0.8001610636711121),
 ('indycar', 0.7994945645332336),
 ('cars', 0.7986459136009216)]
```

In [10]:
```
model.wv.most_similar("facebook")
```

Out[10]:
```
[('youtubefacebook', 0.927127480506897),
 ('thefacebook', 0.8969534635543823),
 ('facebookpage', 0.8882699012756348),
 ('facebox', 0.8686223030090332),
 ('instagram', 0.7984750270843506),
 ('twitteryoutube', 0.7682653069496155),
 ('googleyoutube', 0.7594155669212341),
 ('twitter', 0.7524656057357788),
 ('youtube', 0.7465772032737732),
 ('lnstagram', 0.7246598601341248)]
```

In [11]:
```
model.wv.most_similar("詐欺")
```

Out[11]:
```
[('賈邱', 0.8884372115135193),
 ('赤坑鎮', 0.831690788269043),
 ('中境', 0.8139835000038147),
 ('越中境', 0.7967145442962646),
 ('詐欺罪', 0.7719191908836365),
 ('他魚', 0.7685927152633667),
 ('欺詐', 0.7375842332839966),
 ('抱出', 0.7236839532852173),
 ('欺詐案', 0.6801178455352783),
 ('義德堂', 0.6536234021186829)]
```

In [12]:
```
model.wv.most_similar("合約")
```

Out[12]:
```
[('德康', 0.9192003607749939),
 ('合同', 0.8019339442253113),
 ('綠蠅', 0.749859631061554),
 ('合同期', 0.7271698713302612),
 ('合同額', 0.717912495136261),
 ('合同商', 0.7071802616119385),
 ('簽約', 0.7068753242492676),
 ('續約', 0.7045730948448181),
 ('籤合同', 0.6919059753417969),
 ('合同制', 0.6848821640014648)]
```

In [13]:
```
model.wv.similarity("連結", "鍵接")
```

Out[13]:
```
0.4270057
```

In [14]:
```
model.wv.similarity("連結", "陰天")
```

Out[14]:
```
-0.015750855
```

In [17]:
```
print(f"Loading {output_model}...")
new_model = FastText.load(output_model)
```

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

In [18]:
```
model.wv.similarity("連結", "陰天") == new_model.wv.similarity("連結", "陰天")
```

Out[18]:
```
True
```

In [ ]:
```

```