In [1]:

```
# %load_ext memory_profiler
!pip install -q zhconv
```

In [2]:

```
!pip install gensim
```

```
Requirement already satisfied: gensim in c:\users\user\anaconda3\lib\site-
packages (4.3.0)
Requirement already satisfied: smart-open>=1.8.1 in c:\users\user\anaconda
3\lib\site-packages (from gensim) (5.2.1)
Requirement already satisfied: numpy>=1.18.5 in c:\users\user\anaconda3\li
b\site-packages (from gensim) (1.23.5)
Requirement already satisfied: FuzzyTM>=0.4.0 in c:\users\user\anaconda3\l
ib\site-packages (from gensim) (2.0.5)
Requirement already satisfied: scipy>=1.7.0 in c:\users\user\anaconda3\lib
\site-packages (from gensim) (1.10.0)
Requirement already satisfied: pyfume in c:\users\user\anaconda3\lib\site-
packages (from FuzzyTM>=0.4.0->gensim) (0.2.25)
Requirement already satisfied: pandas in c:\users\user\anaconda3\lib\site-
packages (from FuzzyTM>=0.4.0->gensim) (1.5.3)
Requirement already satisfied: pytz>=2020.1 in c:\users\user\anaconda3\lib
\site-packages (from pandas->FuzzyTM>=0.4.0->gensim) (2022.7)
Requirement already satisfied: python-dateutil>=2.8.1 in c:\users\user\ana
conda3\lib\site-packages (from pandas->FuzzyTM>=0.4.0->gensim) (2.8.2)
Requirement already satisfied: simpful in c:\users\user\anaconda3\lib\site
-packages (from pyfume->FuzzyTM>=0.4.0->gensim) (2.11.0)
Requirement already satisfied: fst-pso in c:\users\user\anaconda3\lib\site
-packages (from pyfume->FuzzyTM>=0.4.0->gensim) (1.8.1)
Requirement already satisfied: six>=1.5 in c:\users\user\anaconda3\lib\sit
e-packages (from python-dateutil>=2.8.1->pandas->FuzzyTM>=0.4.0->gensim)
(1.15.0)
Requirement already satisfied: miniful in c:\users\user\anaconda3\lib\site
-packages (from fst-pso->pyfume->FuzzyTM>=0.4.0->gensim) (0.0.6)
```

In [3]:

```
!pip install wget
```

```
Requirement already satisfied: wget in c:\users\user\anaconda3\lib\site-pa
ckages (3.2)
```

In [4]:

```
import urllib.request
url = "https://github.com/fxsjy/jieba/raw/master/extra_dict/dict.txt.big"
filename = "dict.txt.big"
urllib.request.urlretrieve(url, filename)
```

Out[4]:

```
('dict.txt.big', <http.client.HTTPMessage at 0x22907589cf0>)
```

In [5]:

```
!pip install jieba
```

```
Requirement already satisfied: jieba in c:\users\user\anaconda3\lib\site-p
ackages (0.42.1)
```

In [6]:

```
import os

# Packages
import gensim
import jieba
import zhconv
from gensim.corpora import WikiCorpus
from datetime import datetime as dt
from typing import List


if not os.path.isfile('dict.txt.big'):
    !wget https://github.com/fxsjy/jieba/raw/master/extra_dict/dict.txt.big
jieba.set_dictionary('dict.txt.big')

print("gensim", gensim.__version__)
print("jieba", jieba.__version__)
```

```
gensim 4.3.0
jieba 0.42.1
```

In [7]:

```
ZhWiki = r'C:\Users\User\Downloads\zhwiki-20230501-pages-articles-multistream.xml.bz2'
```

In [8]:

```
zhconv.convert("这原本是一段简体中文", "zh-tw")
```

Out[8]:

```
'這原本是一段簡體中文'
```

In [9]:

```
print(list(jieba.cut("中英夾雜的example，Word2Vec應該很interesting吧?")))
```

```
Building prefix dict from C:\Users\User\Downloads\dict.txt.big ...
Dumping model to file cache C:\Users\User\AppData\Local\Temp\jieba.u9f3f73
11224433a41d53eef275394fc7.cache
Loading model cost 1.190 seconds.
Prefix dict has been built successfully.

['中', '英', '夾雜', '的', 'example', '，', 'Word2Vec', '應該', '很', 'inte
resting', '吧', '?']
```

In [10]:

```
!pip install spacy
```

```
Requirement already satisfied: spacy in c:\users\user\anaconda3\lib\site-p
ackages (3.5.3)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in c:\users\user\an
aconda3\lib\site-packages (from spacy) (3.3.0)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in c:\users\use
r\anaconda3\lib\site-packages (from spacy) (3.0.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in c:\users\use
r\anaconda3\lib\site-packages (from spacy) (1.0.4)
Requirement already satisfied: numpy>=1.15.0 in c:\users\user\anaconda3\li
b\site-packages (from spacy) (1.23.5)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in c:\users\user\anaco
nda3\lib\site-packages (from spacy) (1.1.1)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in c:\users\user\an
aconda3\lib\site-packages (from spacy) (2.0.8)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in c:\users\user\an
aconda3\lib\site-packages (from spacy) (2.28.1)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in c:\users\user\anacon
da3\lib\site-packages (from spacy) (2.4.6)
Requirement already satisfied: pathy>=0.10.0 in c:\users\user\anaconda3\li
b\site-packages (from spacy) (0.10.1)
Requirement already satisfied: thinc<8.2.0,>=8.1.8 in c:\users\user\anacon
da3\lib\site-packages (from spacy) (8.1.10)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in c:\users\user\anac
onda3\lib\site-packages (from spacy) (3.0.8)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in c:\users\user
\anaconda3\lib\site-packages (from spacy) (1.0.9)
Requirement already satisfied: setuptools in c:\users\user\anaconda3\lib\s
ite-packages (from spacy) (65.6.3)
Requirement already satisfied: typer<0.8.0,>=0.3.0 in c:\users\user\anacon
da3\lib\site-packages (from spacy) (0.7.0)
Requirement already satisfied: pydantic!=1.8,!=1.8.1,<1.11.0,>=1.7.4 in
c:\users\user\anaconda3\lib\site-packages (from spacy) (1.10.7)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in c:\users\user\anacon
da3\lib\site-packages (from spacy) (4.64.1)
Requirement already satisfied: smart-open<7.0.0,>=5.2.1 in c:\users\user\a
naconda3\lib\site-packages (from spacy) (5.2.1)
Requirement already satisfied: packaging>=20.0 in c:\users\user\anaconda3
\lib\site-packages (from spacy) (22.0)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in c:\users\user\anacon
da3\lib\site-packages (from spacy) (2.0.7)
Requirement already satisfied: jinja2 in c:\users\user\anaconda3\lib\site-
packages (from spacy) (3.1.2)
Requirement already satisfied: typing-extensions>=4.2.0 in c:\users\user\a
naconda3\lib\site-packages (from pydantic!=1.8,!=1.8.1,<1.11.0,>=1.7.4->sp
acy) (4.4.0)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\user\anac
onda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy) (1.26.14)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\user\anacond
a3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy) (2022.12.7)
Requirement already satisfied: idna<4,>=2.5 in c:\users\user\anaconda3\lib
\site-packages (from requests<3.0.0,>=2.13.0->spacy) (3.4)
Requirement already satisfied: charset-normalizer<3,>=2 in c:\users\user\a
naconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy) (2.0.4)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in c:\users\user\a
naconda3\lib\site-packages (from thinc<8.2.0,>=8.1.8->spacy) (0.0.4)
Requirement already satisfied: blis<0.8.0,>=0.7.8 in c:\users\user\anacond
a3\lib\site-packages (from thinc<8.2.0,>=8.1.8->spacy) (0.7.9)
Requirement already satisfied: colorama in c:\users\user\anaconda3\lib\sit
e-packages (from tqdm<5.0.0,>=4.38.0->spacy) (0.4.6)
Requirement already satisfied: click<9.0.0,>=7.1.1 in c:\users\user\anacon
da3\lib\site-packages (from typer<0.8.0,>=0.3.0->spacy) (8.0.4)
```

```
Requirement already satisfied: MarkupSafe>=2.0 in c:\users\user\anaconda3
\lib\site-packages (from jinja2->spacy) (2.1.1)
```

In [11]:

```
import spacy

# 下載語言模組
spacy.cli.download("zh_core_web_sm")  # 下載 spacy 中文模組
spacy.cli.download("en_core_web_sm")  # 下載 spacy 英文模組

nlp_zh = spacy.load("zh_core_web_sm") # 載入 spacy 中文模組
nlp_en = spacy.load("en_core_web_sm") # 載入 spacy 英文模組

# 印出前20個停用詞
print('--\n')
print(f"中文停用詞 Total={len(nlp_zh.Defaults.stop_words)}: {list(nlp_zh.Defaults.stop_wo
print("--")
print(f"英文停用詞 Total={len(nlp_en.Defaults.stop_words)}: {list(nlp_en.Defaults.stop_wo
```

```
✓ Download and installation successful
You can now load the package via spacy.load('zh_core_web_sm')
✓ Download and installation successful
You can now load the package via spacy.load('en_core_web_sm')
--

中文停用詞 Total=1891: ['最大', '一来', '日渐', '挨个', '从古到今', '尽管如
此', '无论', '哎呀', '不惟', '方才', '着', '大概', '等到', '屡屡', '挨家挨
户', '挨次', '/', '（', '连连', '忽地'] ...
--
英文停用詞 Total=326: ['perhaps', 'take', 'never', 'meanwhile', 'due', 'aft
er', 'hers', 'those', 'could', 'keep', 'around', 'herein', 'whither', 'fro
m', 'behind', 'into', 'alone', 'now', 'everything', 'other'] ...
```

In [12]:

```
STOPWORDS =  nlp_zh.Defaults.stop_words | \
             nlp_en.Defaults.stop_words | \
             set(["\n", "\r\n", "\t", " ", ""])
print(len(STOPWORDS))

# 將簡體停用詞轉成繁體，擴充停用詞表
for word in STOPWORDS.copy():
    STOPWORDS.add(zhconv.convert(word, "zh-tw"))

print(len(STOPWORDS))
```

```
2222
3005
```

In [13]:

```
def preprocess_and_tokenize(
    text: str, token_min_len: int=1, token_max_len: int=15, lower: bool=True) -> List[st
    if lower:
        text  = text.lower()
    text = zhconv.convert(text, "zh-tw")
    return [
        token for token in jieba.cut(text, cut_all=False)
        if token_min_len <= len(token) <= token_max_len and \
            token not in STOPWORDS
    ]
```

In [14]:

```
print(preprocess_and_tokenize("歐幾里得，西元前三世紀的古希臘數學家，現在被認為是幾何之父，此
print(preprocess_and_tokenize("我来到北京清华大学"))
print(preprocess_and_tokenize("中英夾雜的example，Word2Vec應該很interesting吧?"))
```

```
['歐幾', '裡得', '西元前', '世紀', '古希臘', '數學家', '幾何', '父', '此畫',
'拉斐爾']
['來到', '北京', '清華大學']
['中', '英', '夾雜', 'example', 'word2vec', 'interesting']
```

In [15]:

```
print(f"Parsing {ZhWiki}...")
wiki_corpus = WikiCorpus(ZhWiki, token_min_len=1)
```

```
Parsing C:\Users\User\Downloads\zhwiki-20230501-pages-articles-multistrea
m.xml.bz2...

C:\Users\User\anaconda3\lib\site-packages\gensim\utils.py:1333: UserWarnin
g: detected Windows; aliasing chunkize to chunkize_serial
  warnings.warn("detected %s; aliasing chunkize to chunkize_serial" % enti
ty)
```

In [16]:

```
g = wiki_corpus.get_texts()
print(next(g)[:10])
print(next(g)[:10])
print(next(g)[:10])



# print(jieba.lcut("".join(next(g))[:50]))
# print(jieba.lcut("".join(next(g))[:50]))
```

```
['歐幾里得', '西元前三世紀的古希臘數學家', '現在被認為是幾何之父', '此畫為拉斐爾
的作品', '雅典學院', '数学', '是研究數量', '屬於形式科學的一種', '數學利用抽象
化和邏輯推理', '從計數']
['蘇格拉底之死', '由雅克', '路易', '大卫所繪', '年', '哲學', '是研究普遍的',
'基本问题的学科', '包括存在', '知识']
['文學', '在狭义上', '是一种语言艺术', '亦即使用语言文字为手段', '形象化地反映
客观社会生活', '表达主观作者思想感情的一种艺术', '文学不仅强调传达思想观念', '更
强调传达方式的独特性', '且讲究辞章的美感', '文学']
```

In [4]:

```
WIKI_SEG_TXT = "wiki_seg.txt"
```

In [5]:

```
%%time

from gensim.models import word2vec
import multiprocessing

max_cpu_counts = multiprocessing.cpu_count()
word_dim_size = 300  #  設定 word vector 維度
print(f"Use {max_cpu_counts} workers to train Word2Vec (dim={word_dim_size})")


# 讀取訓練語句
sentences = word2vec.LineSentence(WIKI_SEG_TXT)

# 訓練模型
model = word2vec.Word2Vec(sentences, vector_size=word_dim_size, workers=max_cpu_counts)

# 儲存模型
output_model = f"word2vec.zh.{word_dim_size}.model"
model.save(output_model)
```

```
Use 12 workers to train Word2Vec (dim=300)
CPU times: total: 1h 9min 37s
Wall time: 29min 4s
```

In [6]:

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

```
(1281108, 300)
```

Out[6]:

```
array([[ 2.8076830e+00,  1.2115554e+00, -3.0168431e+00, ...,
         2.1298341e-01,  1.0413009e+00, -1.3318332e+00],
       [ 2.3509305e+00, -7.0053291e-01, -2.4598410e+00, ...,
         8.5789061e-01,  2.8712637e+00, -3.0379291e+00],
       [ 9.7158843e-01, -3.6645389e-01, -1.2535313e+00, ...,
        -2.2477122e-02,  6.4034611e-03, -1.3746341e-01],
       ...,
       [ 6.7367656e-03,  4.7338571e-02,  9.2347644e-02, ...,
        -2.3707920e-03, -5.1538050e-02, -8.7550983e-02],
       [-1.8512698e-02, -1.0023722e-02, -1.1426814e-02, ...,
        -5.8442885e-05,  2.7651146e-02, -9.6827056e-03],
       [-5.0373469e-02,  2.7495909e-02,  3.9573699e-02, ...,
         1.7450697e-03, -5.5091362e-02, -2.2538140e-02]], dtype=float32)
```

In [7]:

```
print(f"總共收錄了 {len(model.wv.key_to_index)} 個詞彙")

print("印出 20 個收錄詞彙:")
print(list(model.wv.key_to_index.keys())[:10])
```

```
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']
```

In [8]:

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

```
(300,)
```

Out[8]:

In [9]:

```
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)
```

```
"Key '這肯定沒見過 ' not present"
```

In [10]:

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

Out[10]:

```
[('飲品', 0.8093394041061401),
 ('果汁', 0.7081009745597839),
 ('軟飲料', 0.6987029910087585),
 ('酒精類', 0.6691874265670776),
 ('含酒精', 0.6534177660942078),
 ('瓶裝', 0.6331033706665039),
 ('酒類', 0.6328833699226379),
 ('提神', 0.6264342069625854),
 ('酒水', 0.6196794509887695),
 ('罐裝', 0.6145046353340149)]
```

In [11]:

```
model.wv.most_similar("car")
```

Out[11]:

```
[('truck', 0.6725255250930786),
 ('cab', 0.6530406475067139),
 ('seat', 0.6503428220748901),
 ('tikita', 0.642068624496460),
 ('motor', 0.634936511516571),
 ('wagon', 0.6274935603141785),
 ('vehicle', 0.6248626708984375),
 ('driving', 0.623626708984375),
 ('cabriolet', 0.6228964328765869),
 ('luxury', 0.6170464754104614)]
```

```
array([-2.3187249 ,  0.01615903, -1.9287746 , -1.3532702 ,  2.0199182 ,
       -0.6695111 , -1.1122613 ,  0.8779186 , -1.8652799 , -1.3066971 ,
       -0.47660009, -1.4137077 ,  0.57277536, -0.08464076, -1.9201062 ,
        0.70093197, -1.9009708 , -1.0594636 , -2.0958002 , -0.7651449 ,
       -2.5452335 , -0.9265208 , -0.8671445 ,  0.06142833, -0.82512   ,
       -0.26285926,  1.2750754 ,  1.7273247 ,  0.0280052 , -0.04842537,
       -0.8294002 , -0.9745456 ,  1.3239641 ,  0.89738506, -0.8455278 ,
       -2.8672667 , -1.0797536 ,  3.526261  , -1.1574554 , -0.4385191 ,
       -2.2924526 , -4.1270447 , -1.3164153 ,  0.82559234, -0.57494384,
       -2.0428135 ,  2.0156    , -0.21082444,  1.9229635 , -0.36971295,
       -1.5027969 ,  1.4621071 , -1.2802663 ,  0.21305539,  0.4693426 ,
       -0.52149993, -2.2269216 , -1.7401706 ,  0.51544434,  2.2579443 ,
       -1.5485755 ,  1.8927777 , -0.52080435, -2.2128117 , -1.3925284 ,
       -2.726877  ,  0.45157835,  0.4734782 ,  2.0502794 ,  1.6137005 ,
       -1.4361435 , -1.3721389 , -0.5339236 ,  0.89922655,  0.19019386,
        0.81141675, -3.3819478 ,  0.7973358 , -0.2520694 , -0.4523835 ,
       -1.1291028 ,  0.14261152,  1.5904088 , -3.4663181 ,  0.7920404 ,
       -1.6483672 ,  0.941491  , -0.48369545,  1.6809987 ,  0.92586243,
        1.6481433 , -0.31991285, -3.3504539 ,  2.2591102 , -0.02753139,
        1.2861175 , -2.7851511 , -0.59754294, -0.96135455, -2.6263528 ,
        0.25832286,  1.6388592 ,  0.40657586, -0.38751096,  0.40434107,
        0.21460548, -2.683086  ,  2.6807296 , -1.8877826 ,  1.7389461 ,
        0.23634869, -0.77310044, -1.8935976 , -0.42175317, -2.5223544 ,
       -0.73790294, -0.05748505, -2.0688727 ,  1.0402049 ,  0.05082105,
       -0.810206  ,  1.5652857 ,  0.287196  , -0.7655122 , -2.583647  ,
       -0.6039236 ,  1.1259085 ,  1.0755616 , -1.9423015 , -2.0135927 ,
        0.38187397, -0.03653628, -3.3630455 ,  1.0415881 ,  0.6791245 ,
        1.1275458 ,  0.90551025, -1.9495416 ,  2.2610254 ,  0.30972818,
        0.32236293, -0.14912032,  0.3086139 , -2.8236384 , -0.06726016,
        1.0591159 ,  1.3239223 ,  0.63874996,  0.5161228 , -0.38972253,
        0.5241747 , -0.6194351 ,  3.0056975 , -1.6143725 ,  0.54825586,
        4.078273  ,  2.3946247 ,  1.9459795 ,  2.3803315 , -0.4702034 ,
       -0.1714735 , -1.3943998 , -2.09173   ,  1.2360635 , -0.07071591,
       -0.9863926 ,  1.2416483 , -1.9572176 ,  1.0829248 , -1.7989526 ,
        0.2781602 , -3.0023117 , -0.5433178 , -2.0394232 , -1.7159337 ,
        0.75989205,  1.1329938 ,  0.9445522 , -0.23089415, -2.2350576 ,
       -1.2938199 , -1.2625726 , -0.84647465, -1.9061563 ,  2.0666242 ,
       -0.578414  ,  0.24752066, -0.5919141 , -0.24072686, -0.66556716,
       -0.49562958,  1.0262786 ,  0.31940255, -0.21806663, -0.5607456 ,
       -1.0929866 , -1.8445252 ,  0.52612793, -1.1904604 , -2.542954  ,
       -1.6011027 , -2.6444583 , -0.2614324 ,  0.4526377 , -0.6890156 ,
        0.3497628 ,  2.5980887 , -0.29343086,  3.001629  , -1.0457222 ,
        0.00963606,  1.2653441 , -0.644599  , -0.6582231 ,  1.3897427 ,
        0.5474573 ,  2.0522857 ,  0.22220452, -1.166292  ,  0.7192794 ,
       -1.3088368 , -0.04805864, -1.0686829 , -1.9971707 ,  2.1239347 ,
        1.3761659 ,  1.4310707 , -0.24958465, -0.48361585,  0.34051472,
        2.9673486 ,  2.6818237 , -1.224066  ,  0.6426097 ,  1.6297886 ,
       -0.92241246, -1.0050594 , -1.5973707 ,  0.73252434, -0.7662014 ,
       -0.8832713 ,  0.31644768, -0.8506618 , -2.2681773 ,  2.0381017 ,
        1.5878602 , -0.95660865, -0.3437465 ,  0.8569719 , -0.7194139 ,
        0.89409435,  2.8180773 ,  1.0417285 ,  0.751422  , -0.03957854,
       -0.7609537 , -0.35797483, -2.989081  , -1.594717  , -0.1451011 ,
        2.1480892 , -0.0475874 , -1.913769  ,  0.23995681,  0.5169853 ,
       -0.8153051 ,  1.0542337 ,  1.0361558 , -0.3984292 ,  2.046264  ,
        1.7930588 , -0.87417823, -0.5523076 ,  1.0972489 ,  0.46799898,
        1.5372856 , -0.96159875, -0.5039287 , -2.7191503 ,  1.2790406 ,
        0.5903456 , -0.089865  ,  0.00835543,  0.69932765, -0.14111193,
        4.5176687 ,  0.93654454, -1.5810127 , -0.4929825 ,  0.6508359 ,
       -3.0301054 , -0.4947826 ,  0.63825494,  0.6202983 , -0.44677562,
```

In [12]:

```
model.wv.most_similar("facebook")
```

Out[12]:

```
[('臉書', 0.8149164319038391),
 ('專頁', 0.7451366782188416),
 ('instagram', 0.7248266935348511),
 ('面書', 0.708460271358490),
 ('貼文', 0.7016114592552185),
 ('twitter', 0.6950445175170898),
 ('推特', 0.6841502785682678),
 ('粉絲團', 0.6587467193603516),
 ('粉專', 0.6559129357337952),
 ('tumblr', 0.6437175869941711)]
```

In [13]:

```
model.wv.most_similar("詐欺")
```

Out[13]:

```
[('欺詐', 0.7152416110038757),
 ('詐騙', 0.6116501688957214),
 ('竊盜', 0.5657426118850708),
 ('慣犯', 0.5543817281723022),
 ('詐欺罪', 0.538599431514740),
 ('金光黨', 0.5374152660369873),
 ('信用調查', 0.5151287913322449),
 ('師篇', 0.5096631646156311),
 ('騙徒', 0.5079110860824585),
 ('前科', 0.5054307579994202)]
```

In [14]:

```
model.wv.most_similar("合約")
```

Out[14]:

```
[('合同', 0.7782608270645142),
 ('簽約', 0.7128676772117615),
 ('續約', 0.6848387122154236),
 ('續簽', 0.6119592785835266),
 ('租約', 0.6113901734352112),
 ('選擇權', 0.5982978343963623),
 ('簽下', 0.5979309082031250),
 ('買斷', 0.5780441164970398),
 ('解約', 0.5734624862670898),
 ('短約', 0.5693488717079163)]
```

In [15]:

```
model.wv.similarity("連結", "鏈接")
```

Out[15]:

```
0.7164953
```

```
        -0.3824682 ,  0.06467538, -0.6349842 ,  0.04732496, -1.2558346 ],
      dtype=float32)
```

In [16]:

```
model.wv.similarity("連結", "陰天")
```

Out[16]:

```
0.022799488
```

In [17]:

```
print(f"Loading {output_model}...")
new_model = word2vec.Word2Vec.load(output_model)
```

```
Loading word2vec.zh.300.model...
```

In [18]:

```
model.wv.similarity("連結", "陰天") == new_model.wv.similarity("連結", "陰天")
```

Out[18]:

```
True
```

In [ ]:

In [ ]: