In [3]:
```
pip install gensim
#Gensim is an open-source Python library designed for vector space modeling
#It is particularly focused on unsupervised machine learning tasks such as c
#and topic modeling. Gensim is developed by Radim Řehůřek and is widely usec
#information retrieval applications.
```

```
Defaulting to user installation because normal site-packages is not writea
ble
Requirement already satisfied: gensim in c:\programdata\anaconda3\lib\site
-packages (4.1.2)
Requirement already satisfied: smart-open>=1.8.1 in c:\programdata\anacond
a3\lib\site-packages (from gensim) (5.2.1)
Requirement already satisfied: scipy>=0.18.1 in c:\programdata\anaconda3\l
ib\site-packages (from gensim) (1.9.1)
Requirement already satisfied: numpy>=1.17.0 in c:\users\dell\appdata\roam
ing\python\python39\site-packages (from gensim) (1.24.3)
Note: you may need to restart the kernel to use updated packages.
```

In [33]:
```
pip install python-Levenshtein
#If you're working on a project that involves comparing or matching strings,
#or if you're using a library or tool that depends on python-Levenshtein,
```

```
Defaulting to user installation because normal site-packages is not writea
bleNote: you may need to restart the kernel to use updated packages.

Requirement already satisfied: python-Levenshtein in c:\users\dell\appdata
\roaming\python\python39\site-packages (0.23.0)
Requirement already satisfied: Levenshtein==0.23.0 in c:\users\dell\appdat
a\roaming\python\python39\site-packages (from python-Levenshtein) (0.23.0)
Requirement already satisfied: rapidfuzz<4.0.0,>=3.1.0 in c:\users\dell\ap
pdata\roaming\python\python39\site-packages (from Levenshtein==0.23.0->pyt
hon-Levenshtein) (3.5.2)
```

dataset Link:
http://snap.stanford.edu/data/amazon/productGraph/categoryFiles/reviews_Cell_Phones_and
(http://snap.stanford.edu/data/amazon/productGraph/categoryFiles/reviews_Cell_Phones_and

In [4]:
```
import pandas as pd
import numpy as np
import gensim
```

In [5]:
```
df=pd.read_json("Cell_Phones_and_Accessories_5.json",lines=True)#shift tab
```

In [6]:
```
df.head()
```

Out[6]:

	reviewerID	asin	reviewerName	helpful	reviewText	overall	summary	uni
0	A30TL5EWN6DFXT	120401325X	christina	[0, 0]	They look good and stick good! I just don't li...	4	Looks Good	
1	ASY55RVNIL0UD	120401325X	emily l.	[0, 0]	These stickers work like the review says they ...	5	Really great product.	
2	A2TMXE2AFO7ONB	120401325X	Erica	[0, 0]	These are awesome and make my phone look so st...	5	LOVE LOVE LOVE	
3	AWJ0WZQYMYFQ4	120401325X	JM	[4, 4]	Item arrived in great time and was in perfect ...	4	Cute!	
4	ATX7CZYFXI1KW	120401325X	patrice m rogoza	[2, 3]	awesome! stays on, and looks great. can be use...	5	leopard home button sticker for iphone 4s	

In [7]:
```
df.shape
```

Out[7]:
```
(194439, 9)
```

In [8]:
```
df.reviewText[0]
```

Out[8]:
```
"They look good and stick good! I just don't like the rounded shape becaus
e I was always bumping it and Siri kept popping up and it was irritating.
I just won't buy a product like this again"
```

In [9]:
```
df.reviewText[1]
```

Out[9]:
```
'These stickers work like the review says they do. They stick on great and
they stay on the phone. They are super stylish and I can share them with m
y sister. :)'
```

In [10]:
```
# first step- Preprocessing.
```

In [11]:
```
gensim.utils.simple_preprocess("They look good and stick good! I just don't
```

Out[11]:
```
['they',
 'look',
 'good',
 'and',
 'stick',
 'good',
 'just',
 'don',
 'like',
 'the',
 'rounded',
 'shape',
 'because',
 'was',
 'always',
 'bumping',
 'it',
 'and',
 'siri',
 'kept',
 'popping',
 'up',
 'and',
 'it',
 'was',
 'irritating',
 'just',
 'won',
 'buy',
 'product',
 'like',
 'this',
 'again']
```

In [12]:
```
#this we done on single review But if want to apply on entire dataset
```

In [14]:
```
reviews=df.reviewText.apply(gensim.utils.simple_preprocess)
reviews
```

Out[14]:
```
0         [they, look, good, and, stick, good, just, don...
1         [these, stickers, work, like, the, review, say...
2         [these, are, awesome, and, make, my, phone, lo...
3         [item, arrived, in, great, time, and, was, in,...
4         [awesome, stays, on, and, looks, great, can, b...
                                ...
194434    [works, great, just, like, my, original, one, ...
194435    [great, product, great, packaging, high, quali...
194436    [this, is, great, cable, just, as, good, as, t...
194437    [really, like, it, becasue, it, works, well, w...
194438    [product, as, described, have, wasted, lot, of...
Name: reviewText, Length: 194439, dtype: object
```

In [16]:
```
model=gensim.models.Word2Vec(
window=10,
min_count=2,
workers=4)
```

In [18]:
```
model.build_vocab(reviews,progress_per=1000) #initiallizing model
```

In [19]:
```
model.epochs
```

Out[19]:
```
5
```

In [20]:
```
model.corpus_count
```

Out[20]:
```
194439
```

In [21]:
```
model.train(reviews, total_examples=model.corpus_count, epochs=model.epochs
```

Out[21]:
```
(61501730, 83868975)
```

In [23]:
```
#save the model so to use again for other purpose.
model.save("./Word2Vec.model")
```

In [24]:
```
#prdict words similar to ==> "bad"
model.wv.most_similar('bad')
```

Out[24]:
```
[('terrible', 0.6661187410354614),
 ('shabby', 0.6488698124885559),
 ('horrible', 0.5989841818809509),
 ('good', 0.5797889232635498),
 ('lame', 0.5598593950271606),
 ('crappy', 0.5590352416038513),
 ('funny', 0.5565664768218994),
 ('awful', 0.5403169393539429),
 ('okay', 0.5346116423606873),
 ('disappointing', 0.530344545841217)]
```

In [25]:
```
#to print similarity score of 2 word
model.wv.similarity(w1="cheap",w2="inexpensive")
```

Out[25]:
```
0.5211316
```

In [26]:
```
model.wv.similarity(w1="great",w2="great")
```

Out[26]:
```
0.7905774
```

In [27]:
```
model.wv.similarity(w1="great",w2="product")
```

Out[27]:
```
-0.045096543
```

In [28]:
```
model.wv.similarity(w1="great",w2="awesome")
```

Out[28]:
```
0.755809
```

In [1]:
```
pip install Pyppeteer
```

```
Defaulting to user installation because normal site-packages is not writea
ble
Collecting Pyppeteer
  Downloading pyppeteer-1.0.2-py3-none-any.whl (83 kB)
     ---------------------------------------- 83.4/83.4 kB 360.3 kB/s eta 0:
00:00
Collecting pyee<9.0.0,>=8.1.0
  Downloading pyee-8.2.2-py2.py3-none-any.whl (12 kB)
Collecting websockets<11.0,>=10.0
  Downloading websockets-10.4-cp39-cp39-win_amd64.whl (101 kB)
     ---------------------------------------- 101.4/101.4 kB 1.9 MB/s eta 0:
00:00
Requirement already satisfied: tqdm<5.0.0,>=4.42.1 in c:\programdata\anaco
nda3\lib\site-packages (from Pyppeteer) (4.64.1)
Requirement already satisfied: importlib-metadata>=1.4 in c:\programdata\a
naconda3\lib\site-packages (from Pyppeteer) (4.11.3)
Requirement already satisfied: urllib3<2.0.0,>=1.25.8 in c:\programdata\an
aconda3\lib\site-packages (from Pyppeteer) (1.26.11)
Requirement already satisfied: certifi>=2021 in c:\programdata\anaconda3\l
ib\site-packages (from Pyppeteer) (2022.9.14)
Requirement already satisfied: appdirs<2.0.0,>=1.4.3 in c:\programdata\ana
conda3\lib\site-packages (from Pyppeteer) (1.4.4)
Requirement already satisfied: zipp>=0.5 in c:\programdata\anaconda3\lib\s
ite-packages (from importlib-metadata>=1.4->Pyppeteer) (3.8.0)
Requirement already satisfied: colorama in c:\programdata\anaconda3\lib\si
te-packages (from tqdm<5.0.0,>=4.42.1->Pyppeteer) (0.4.5)
Installing collected packages: pyee, websockets, Pyppeteer
Successfully installed Pyppeteer-1.0.2 pyee-8.2.2 websockets-10.4
Note: you may need to restart the kernel to use updated packages.

  WARNING: The script pyppeteer-install.exe is installed in 'C:\Users\Dell
\AppData\Roaming\Python\Python39\Scripts' which is not on PATH.
  Consider adding this directory to PATH or, if you prefer to suppress thi
s warning, use --no-warn-script-location.
```

In [2]:
```
pip install pandoc
```

```
Defaulting to user installation because normal site-packages is not writea
bleNote: you may need to restart the kernel to use updated packages.

Collecting pandoc
  Downloading pandoc-2.3.tar.gz (33 kB)
  Preparing metadata (setup.py): started
  Preparing metadata (setup.py): finished with status 'done'
Collecting plumbum
  Downloading plumbum-1.8.2-py3-none-any.whl (127 kB)
     ---------------------------------- 127.0/127.0 kB 679.1 kB/s eta 0:
00:00
Collecting ply
  Downloading ply-3.11-py2.py3-none-any.whl (49 kB)
     ---------------------------------------- 49.6/49.6 kB 1.3 MB/s eta 0:
00:00
Requirement already satisfied: pywin32 in c:\programdata\anaconda3\lib\sit
e-packages (from plumbum->pandoc) (302)
Building wheels for collected packages: pandoc
  Building wheel for pandoc (setup.py): started
  Building wheel for pandoc (setup.py): finished with status 'done'
  Created wheel for pandoc: filename=pandoc-2.3-py3-none-any.whl size=3326
3 sha256=ab0bd0b162a436790fcf6478b41ec91db2c18904ded365ff09635ff071dd21b6
  Stored in directory: c:\users\dell\appdata\local\pip\cache\wheels\69\e6
\a1\1daa96d919c9e09a71473649b717b8da286f3f8d7719d1cfc5
Successfully built pandoc
Installing collected packages: ply, plumbum, pandoc
Successfully installed pandoc-2.3 plumbum-1.8.2 ply-3.11
```

In [ ]:
```
```