In [1]:     1  !pip install gensim

```
Requirement already satisfied: gensim in c:\users\yashm\anaconda3\lib\site-packages (3.8.3)
Requirement already satisfied: smart-open>=1.8.1 in c:\users\yashm\anaconda3\lib\site-packages (from gensim) (2.0.0)
Requirement already satisfied: six>=1.5.0 in c:\users\yashm\anaconda3\lib\site-packages (from gensim) (1.15.0)
Requirement already satisfied: numpy>=1.11.3 in c:\users\yashm\anaconda3\lib\site-packages (from gensim) (1.19.5)
Requirement already satisfied: scipy>=0.18.1 in c:\users\yashm\anaconda3\lib\site-packages (from gensim) (1.4.1)
Requirement already satisfied: Cython==0.29.14 in c:\users\yashm\anaconda3\lib\site-packages (from gensim) (0.29.14)
Requirement already satisfied: requests in c:\users\yashm\anaconda3\lib\site-packages (from smart-open>=1.8.1->gensim)
(2.25.1)
Requirement already satisfied: boto3 in c:\users\yashm\anaconda3\lib\site-packages (from smart-open>=1.8.1->gensim) (1.
14.0)
Requirement already satisfied: boto in c:\users\yashm\anaconda3\lib\site-packages (from smart-open>=1.8.1->gensim) (2.4
9.0)
Requirement already satisfied: botocore<1.18.0,>=1.17.0 in c:\users\yashm\anaconda3\lib\site-packages (from boto3->smar
t-open>=1.8.1->gensim) (1.17.0)
Requirement already satisfied: jmespath<1.0.0,>=0.7.1 in c:\users\yashm\anaconda3\lib\site-packages (from boto3->smart-
open>=1.8.1->gensim) (0.10.0)
Requirement already satisfied: s3transfer<0.4.0,>=0.3.0 in c:\users\yashm\anaconda3\lib\site-packages (from boto3->smar
t-open>=1.8.1->gensim) (0.3.3)
Requirement already satisfied: urllib3<1.26,>=1.20 in c:\users\yashm\anaconda3\lib\site-packages (from botocore<1.18.0,
>=1.17.0->boto3->smart-open>=1.8.1->gensim) (1.25.8)
Requirement already satisfied: docutils<0.16,>=0.10 in c:\users\yashm\anaconda3\lib\site-packages (from botocore<1.18.
0,>=1.17.0->boto3->smart-open>=1.8.1->gensim) (0.15.2)
Requirement already satisfied: python-dateutil<3.0.0,>=2.1 in c:\users\yashm\anaconda3\lib\site-packages (from botocore
<1.18.0,>=1.17.0->boto3->smart-open>=1.8.1->gensim) (2.8.1)
Requirement already satisfied: chardet<5,>=3.0.2 in c:\users\yashm\anaconda3\lib\site-packages (from requests->smart-op
en>=1.8.1->gensim) (3.0.4)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\yashm\anaconda3\lib\site-packages (from requests->smart-o
pen>=1.8.1->gensim) (2019.11.28)
Requirement already satisfied: idna<3,>=2.5 in c:\users\yashm\anaconda3\lib\site-packages (from requests->smart-open>=
1.8.1->gensim) (2.8)

WARNING: You are using pip version 20.3.3; however, version 22.2.2 is available.
You should consider upgrading via the 'C:\Users\yashm\anaconda3\python.exe -m pip install --upgrade pip' command.
```

In [3]:

```python
## Requirements: nltk,gensim
from nltk.tokenize import sent_tokenize, word_tokenize
import nltk
nltk.download('punkt')

import gensim
from gensim.models import Word2Vec


## procedure to read text files in python:
sample = open("Allice.txt", "r")
s = sample.read()

f = s.replace("\n", " ")

#Creating a bag of words
data = []

for i in sent_tokenize(f):
    temp = []

    for j in word_tokenize(i):
        temp.append(j.lower())

    data.append(temp)
```

```
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\yashm\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
```

In [4]:
```
1  data
```

```
       'and',
       'stupid',
       ')',
       ',',
       'whether',
       'the',
       'pleasure',
       'of',
       'making',
       'a',
       'daisy-chain',
       'would',
       'be',
       'worth',
       'the',
       'trouble',
       'of',
       'getting',
       'up',
       'and',
```

In [5]:
```
1  # from beginning to end
2  # from start to end
```

In [6]:
```
1  s
```

Out[6]: "Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do: once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, `and what is the use of a book,' thought Alice `without pictures or conversation?'\nSo she was considering in her own mind (as well as she could, for the hot day made her feel very sleepy and stupid), whether the pleasure of making a daisy-chain would be worth the trouble of getting up and picking the daisies, when suddenly a White Rabbit with pink eyes ran close by her.\n\nThere was nothing so very remarkable in that; nor did Alice think it so very much out of the way to hear the Rabbit say to itself, `Oh dear! Oh dear! I shall be late!' (when she thought it over afterwards, it occurred to her that she ought to have wondered at this, but at the time it all seemed quite natural); but when the Rabbit actually took a watch out of its waistcoat-pocket, and looked at it, and then hurried on, Alice started to her feet, for it flashed across her mind that she had never before seen a rabbit with either a waistcoat-pocket, or a watch to take out of it, and burning with curiosity, she ran across the field after it, and fortunately was just in time to see it pop down a large rabbit-hole under the hedge."

In [7]:
```
1  model1 = gensim.models.Word2Vec(data, min_count = 1,
2                                   size = 32, window = 5)
```

In [11]:
```
1  model1.similarity('pictures','conversation')
```

C:\Users\yashm\anaconda3\lib\site-packages\ipykernel_launcher.py:1: DeprecationWarning: Call to deprecated `similarity`
(Method will be removed in 4.0.0, use self.wv.similarity() instead).
  """Entry point for launching an IPython kernel.

Out[11]: 0.12516128

In [12]:
```
1  model1.similar_by_word('sister',topn = 10)
```

C:\Users\yashm\anaconda3\lib\site-packages\ipykernel_launcher.py:1: DeprecationWarning: Call to deprecated `similar_by_word` (Method will be removed in 4.0.0, use self.wv.similar_by_word() instead).
  """Entry point for launching an IPython kernel.

Out[12]: [('use', 0.39400291442871094),
          ('her', 0.38451123237609863),
          (';', 0.37227243185043335),
          ('so', 0.3567379117012024),
          ('shall', 0.35244184732437134),
          ('afterwards', 0.31063932180404663),
          ('day', 0.30770426988601685),
          (')', 0.2983238101005554),
          ('conversation', 0.27072620391845703),
          ('is', 0.2541230320930481)]

In [15]:
```
1 model1['alice']
```

C:\Users\yashm\anaconda3\lib\site-packages\ipykernel_launcher.py:1: DeprecationWarning: Call to deprecated `__getitem__
` (Method will be removed in 4.0.0, use self.wv.__getitem__() instead).
  """Entry point for launching an IPython kernel.

Out[15]: array([-0.01013712, -0.01046759,  0.00857046,  0.00508187, -0.00330526,
        0.01029478, -0.01233976,  0.00070719,  0.01480515, -0.00172649,
        0.00740297, -0.00809265,  0.01083408, -0.00253887,  0.00499096,
       -0.01478628, -0.00967111, -0.0061741 , -0.01077706, -0.00760007,
        0.01031225,  0.01317534,  0.01314594,  0.01114891, -0.00448742,
       -0.00840383,  0.00666611, -0.00639092, -0.01258383, -0.00228606,
       -0.01345703,  0.00101697], dtype=float32)

In [ ]:
```
1
```