In [1]:
```python
data = open("C:/Users/PRABHAKAR VENKAT/Downloads/1661-0.txt", "r", encoding= 'utf8').read()
print (data[:56])
```

```
Project Gutenberg's The Adventures of Sherlock Holmes,
```

In [2]:
```python
text = open("C:/Users/PRABHAKAR VENKAT/Downloads/1661-0.txt", 'r', encoding='utf-8').read().lower()
print('corpus length:', len(text))
```

```
corpus length: 581888
```

In [3]:
```python
import string
```

In [4]:
```python
def clean_doc(doc):
    doc = doc.replace('--', ' ')
    tokens = doc.split()
    tokens = [' ' if w in string.punctuation else w for w in tokens]
    tokens = [word for word in tokens if word.isalpha()]
    tokens = [word.lower() for word in tokens]
    return tokens

tokens = clean_doc(text)

number_of_unique_tokens = len(set(tokens))
```

In [5]:
```python
print('Total Tokens: %d' % len(tokens))
print('Unique Tokens: %d' % number_of_unique_tokens)
print('These are the first 200 tokens: %s' % tokens[:200])
```

```
Total Tokens: 88473
Unique Tokens: 6672
These are the first 200 tokens: ['project', 'the', 'adventures', 'of', 'sherlock', 'by', 'arthur', 'conan', 'doyle', 'this', 'ebook', 'is', 'for', 'the', 'use', 'of', 'anyone', 'anywhere', 'at', 'no', 'cost', 'and', 'with', 'almost',
'no', 'restrictions', 'you', 'may', 'copy', 'give', 'it', 'away', 'or', 'it', 'under', 'the', 'terms', 'of', 'the', 'project', 'gutenberg', 'license', 'included', 'with', 'this', 'ebook', 'or', 'online', 'at', 'the', 'adventures', 'o
f', 'sherlock', 'holmes', 'arthur', 'conan', 'doyle', 'release', 'november', 'last', 'may', 'english', 'character', 'set', 'start', 'of', 'this', 'project', 'gutenberg', 'ebook', 'the', 'adventures', 'of', 'sherlock', 'holmes', 'produc
ed', 'by', 'an', 'anonymous', 'project', 'gutenberg', 'volunteer', 'and', 'jose', 'menendez', 'cover', 'the', 'adventures', 'of', 'sherlock', 'holmes', 'by', 'arthur', 'conan', 'doyle', 'contents', 'a', 'scandal', 'in', 'bohemia', 'th
e', 'league', 'a', 'case', 'of', 'identity', 'the', 'boscombe', 'valley', 'mystery', 'the', 'five', 'orange', 'pips', 'the', 'man', 'with', 'the', 'twisted', 'lip', 'the', 'adventure', 'of', 'the', 'blue', 'carbuncle', 'the', 'adventur
e', 'of', 'the', 'speckled', 'band', 'the', 'adventure', 'of', 'the', 'thumb', 'the', 'adventure', 'of', 'the', 'noble', 'bachelor', 'the', 'adventure', 'of', 'the', 'beryl', 'coronet', 'the', 'adventure', 'of', 'the', 'copper', 'beech
es', 'a', 'scandal', 'in', 'bohemia', 'to', 'sherlock', 'holmes', 'she', 'is', 'always', 'i', 'have', 'seldom', 'heard', 'him', 'mention', 'her', 'under', 'any', 'other', 'in', 'his', 'eyes', 'she', 'eclipses', 'and', 'predominates',
'the', 'whole', 'of', 'her', 'it', 'was', 'not', 'that', 'he', 'felt', 'any', 'emotion', 'akin', 'to', 'love', 'for', 'irene', 'all']
```

In [6]:
```python
sequence_length = 2
length = sequence_length + 1
sequences = list()
for i in range(length, len(tokens)):
    seq = tokens[i-length:i]
    line = ' '.join(seq)
    sequences.append(line)
```

In [7]:
```python
print ('Total Sequences: %d' % len(sequences))
print ('This is the first sequence: {0}'.format(sequences[0]))
```

```
Total Sequences: 88470
This is the first sequence: project the adventures
```

In [8]:
```python
import numpy as np
from keras.preprocessing.text import Tokenizer
from keras.utils import to_categorical
from keras.models import Sequential
from keras.layers import Dense
from keras.layers import LSTM
from keras.layers import Embedding
```

In [9]:
```python
tokenizer = Tokenizer()
tokenizer.fit_on_texts(sequences)
sequences = tokenizer.texts_to_sequences(sequences)
vocab_size = number_of_unique_tokens + 1

sequences0 = np.array(sequences)
X, y = sequences0[:,:-1], sequences0[:,-1]
y = to_categorical(y, num_classes=vocab_size)
```

In [10]:
```python
dimensions_to_represent_word = 100

model = Sequential()
model.add(Embedding(vocab_size, sequence_length, input_length=sequence_length))
model.add(LSTM(100, return_sequences=True))
model.add(LSTM(100))
model.add(Dense(100, activation='relu'))
model.add(Dense(vocab_size, activation='softmax'))
print(model.summary())

model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
```

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding (Embedding)       (None, 2, 2)              13346

 lstm (LSTM)                 (None, 2, 100)            41200

 lstm_1 (LSTM)               (None, 100)               80400

 dense (Dense)               (None, 100)               10100

 dense_1 (Dense)             (None, 6673)              673973

=================================================================
Total params: 819,019
Trainable params: 819,019
Non-trainable params: 0
_________________________________________________________________
None
```

In [11]:
```python
model.fit(X, y, batch_size=201, epochs=100)
```

```
Epoch 1/100
441/441 [==============================] - 26s 39ms/step - loss: 6.4478 - accuracy: 0.0598
Epoch 2/100
441/441 [==============================] - 16s 36ms/step - loss: 6.0789 - accuracy: 0.0645
Epoch 3/100
441/441 [==============================] - 15s 35ms/step - loss: 6.0035 - accuracy: 0.0645
Epoch 4/100
441/441 [==============================] - 15s 35ms/step - loss: 5.9503 - accuracy: 0.0712
Epoch 5/100
441/441 [==============================] - 16s 35ms/step - loss: 5.8279 - accuracy: 0.0852
Epoch 6/100
441/441 [==============================] - 16s 37ms/step - loss: 5.6777 - accuracy: 0.0923
Epoch 7/100
441/441 [==============================] - 16s 37ms/step - loss: 5.5631 - accuracy: 0.0983
Epoch 8/100
441/441 [==============================] - 16s 35ms/step - loss: 5.4659 - accuracy: 0.1030
Epoch 9/100
441/441 [==============================] - 15s 34ms/step - loss: 5.3943 - accuracy: 0.1073
Epoch 10/100
441/441 [==============================] - 15s 35ms/step - loss: 5.3347 - accuracy: 0.1111
Epoch 11/100
441/441 [==============================] - 15s 34ms/step - loss: 5.2827 - accuracy: 0.1137
Epoch 12/100
441/441 [==============================] - 17s 38ms/step - loss: 5.2358 - accuracy: 0.1163
Epoch 13/100
441/441 [==============================] - 15s 34ms/step - loss: 5.1946 - accuracy: 0.1185
Epoch 14/100
441/441 [==============================] - 15s 34ms/step - loss: 5.1530 - accuracy: 0.1211
Epoch 15/100
441/441 [==============================] - 17s 39ms/step - loss: 5.1108 - accuracy: 0.1261
Epoch 16/100
441/441 [==============================] - 16s 36ms/step - loss: 5.0687 - accuracy: 0.1286
Epoch 17/100
441/441 [==============================] - 15s 35ms/step - loss: 5.0285 - accuracy: 0.1310
Epoch 18/100
441/441 [==============================] - 15s 35ms/step - loss: 4.9883 - accuracy: 0.1326
Epoch 19/100
441/441 [==============================] - 15s 35ms/step - loss: 4.9462 - accuracy: 0.1348
Epoch 20/100
441/441 [==============================] - 16s 36ms/step - loss: 4.9049 - accuracy: 0.1370
Epoch 21/100
441/441 [==============================] - 15s 35ms/step - loss: 4.8619 - accuracy: 0.1388
Epoch 22/100
441/441 [==============================] - 15s 35ms/step - loss: 4.8191 - accuracy: 0.1408
Epoch 23/100
441/441 [==============================] - 15s 35ms/step - loss: 4.7772 - accuracy: 0.1422
Epoch 24/100
441/441 [==============================] - 15s 35ms/step - loss: 4.7381 - accuracy: 0.1442
Epoch 25/100
441/441 [==============================] - 15s 35ms/step - loss: 4.6996 - accuracy: 0.1445
Epoch 26/100
441/441 [==============================] - 15s 35ms/step - loss: 4.6615 - accuracy: 0.1461
Epoch 27/100
441/441 [==============================] - 15s 35ms/step - loss: 4.6229 - accuracy: 0.1477
Epoch 28/100
441/441 [==============================] - 15s 35ms/step - loss: 4.5873 - accuracy: 0.1492
Epoch 29/100
441/441 [==============================] - 15s 35ms/step - loss: 4.5521 - accuracy: 0.1497
Epoch 30/100
441/441 [==============================] - 15s 35ms/step - loss: 4.5206 - accuracy: 0.1514
Epoch 31/100
441/441 [==============================] - 16s 35ms/step - loss: 4.4877 - accuracy: 0.1512
Epoch 32/100
441/441 [==============================] - 15s 35ms/step - loss: 4.4568 - accuracy: 0.1536
Epoch 33/100
441/441 [==============================] - 15s 35ms/step - loss: 4.4291 - accuracy: 0.1541
Epoch 34/100
441/441 [==============================] - 15s 35ms/step - loss: 4.4020 - accuracy: 0.1550
Epoch 35/100
441/441 [==============================] - 15s 35ms/step - loss: 4.3769 - accuracy: 0.1560
Epoch 36/100
441/441 [==============================] - 15s 35ms/step - loss: 4.3518 - accuracy: 0.1572
Epoch 37/100
441/441 [==============================] - 15s 35ms/step - loss: 4.3323 - accuracy: 0.1577
Epoch 38/100
441/441 [==============================] - 15s 35ms/step - loss: 4.3091 - accuracy: 0.1592
Epoch 39/100
441/441 [==============================] - 16s 35ms/step - loss: 4.2878 - accuracy: 0.1604
Epoch 40/100
441/441 [==============================] - 15s 35ms/step - loss: 4.2705 - accuracy: 0.1607
Epoch 41/100
441/441 [==============================] - 15s 35ms/step - loss: 4.2510 - accuracy: 0.1617
Epoch 42/100
441/441 [==============================] - 15s 35ms/step - loss: 4.2348 - accuracy: 0.1622
Epoch 43/100
441/441 [==============================] - 16s 37ms/step - loss: 4.2183 - accuracy: 0.1630
Epoch 44/100
441/441 [==============================] - 15s 35ms/step - loss: 4.2023 - accuracy: 0.1642
Epoch 45/100
441/441 [==============================] - 18s 41ms/step - loss: 4.1891 - accuracy: 0.1649
Epoch 46/100
441/441 [==============================] - 17s 38ms/step - loss: 4.1743 - accuracy: 0.1654
Epoch 47/100
441/441 [==============================] - 16s 36ms/step - loss: 4.1600 - accuracy: 0.1660
Epoch 48/100
441/441 [==============================] - 16s 35ms/step - loss: 4.1482 - accuracy: 0.1670
Epoch 49/100
441/441 [==============================] - 17s 38ms/step - loss: 4.1362 - accuracy: 0.1673
Epoch 50/100
441/441 [==============================] - 18s 42ms/step - loss: 4.1233 - accuracy: 0.1678
Epoch 51/100
441/441 [==============================] - 18s 40ms/step - loss: 4.1113 - accuracy: 0.1681
Epoch 52/100
441/441 [==============================] - 19s 43ms/step - loss: 4.1024 - accuracy: 0.1685
Epoch 53/100
441/441 [==============================] - 17s 39ms/step - loss: 4.0903 - accuracy: 0.1698
Epoch 54/100
441/441 [==============================] - 17s 38ms/step - loss: 4.0797 - accuracy: 0.1707
Epoch 55/100
441/441 [==============================] - 16s 35ms/step - loss: 4.0696 - accuracy: 0.1717
Epoch 56/100
441/441 [==============================] - 16s 35ms/step - loss: 4.0593 - accuracy: 0.1717
Epoch 57/100
441/441 [==============================] - 15s 35ms/step - loss: 4.0510 - accuracy: 0.1722
Epoch 58/100
441/441 [==============================] - 15s 35ms/step - loss: 4.0402 - accuracy: 0.1724
Epoch 59/100
441/441 [==============================] - 15s 35ms/step - loss: 4.0316 - accuracy: 0.1731
Epoch 60/100
441/441 [==============================] - 15s 35ms/step - loss: 4.0247 - accuracy: 0.1729
Epoch 61/100
441/441 [==============================] - 15s 35ms/step - loss: 4.0178 - accuracy: 0.1741
Epoch 62/100
441/441 [==============================] - 15s 35ms/step - loss: 4.0084 - accuracy: 0.1737
Epoch 63/100
441/441 [==============================] - 15s 35ms/step - loss: 4.0006 - accuracy: 0.1754
Epoch 64/100
441/441 [==============================] - 15s 35ms/step - loss: 3.9924 - accuracy: 0.1752
Epoch 65/100
441/441 [==============================] - 15s 35ms/step - loss: 3.9849 - accuracy: 0.1755
Epoch 66/100
441/441 [==============================] - 15s 35ms/step - loss: 3.9784 - accuracy: 0.1765
Epoch 67/100
441/441 [==============================] - 15s 35ms/step - loss: 3.9696 - accuracy: 0.1768
Epoch 68/100
441/441 [==============================] - 15s 35ms/step - loss: 3.9641 - accuracy: 0.1775
Epoch 69/100
441/441 [==============================] - 16s 36ms/step - loss: 3.9560 - accuracy: 0.1776
Epoch 70/100
441/441 [==============================] - 16s 36ms/step - loss: 3.9495 - accuracy: 0.1790
Epoch 71/100
441/441 [==============================] - 16s 36ms/step - loss: 3.9419 - accuracy: 0.1784
Epoch 72/100
441/441 [==============================] - 16s 36ms/step - loss: 3.9359 - accuracy: 0.1804
Epoch 73/100
441/441 [==============================] - 18s 42ms/step - loss: 3.9307 - accuracy: 0.1799
Epoch 74/100
441/441 [==============================] - 15s 35ms/step - loss: 3.9243 - accuracy: 0.1789
Epoch 75/100
441/441 [==============================] - 16s 35ms/step - loss: 3.9177 - accuracy: 0.1807
Epoch 76/100
441/441 [==============================] - 16s 35ms/step - loss: 3.9115 - accuracy: 0.1812
Epoch 77/100
441/441 [==============================] - 15s 35ms/step - loss: 3.9067 - accuracy: 0.1806
Epoch 78/100
441/441 [==============================] - 15s 35ms/step - loss: 3.9008 - accuracy: 0.1819
Epoch 79/100
441/441 [==============================] - 15s 34ms/step - loss: 3.8951 - accuracy: 0.1823
Epoch 80/100
441/441 [==============================] - 15s 34ms/step - loss: 3.8883 - accuracy: 0.1825
Epoch 81/100
441/441 [==============================] - 15s 34ms/step - loss: 3.8827 - accuracy: 0.1832
Epoch 82/100
441/441 [==============================] - 15s 34ms/step - loss: 3.8790 - accuracy: 0.1834
Epoch 83/100
441/441 [==============================] - 15s 34ms/step - loss: 3.8727 - accuracy: 0.1843
Epoch 84/100
441/441 [==============================] - 15s 34ms/step - loss: 3.8673 - accuracy: 0.1845
Epoch 85/100
441/441 [==============================] - 15s 34ms/step - loss: 3.8633 - accuracy: 0.1843
Epoch 86/100
441/441 [==============================] - 15s 34ms/step - loss: 3.8580 - accuracy: 0.1845
Epoch 87/100
441/441 [==============================] - 15s 34ms/step - loss: 3.8524 - accuracy: 0.1848
Epoch 88/100
441/441 [==============================] - 15s 34ms/step - loss: 3.8487 - accuracy: 0.1854
Epoch 89/100
441/441 [==============================] - 15s 34ms/step - loss: 3.8425 - accuracy: 0.1853
Epoch 90/100
441/441 [==============================] - 15s 34ms/step - loss: 3.8364 - accuracy: 0.1870
Epoch 91/100
441/441 [==============================] - 15s 34ms/step - loss: 3.8339 - accuracy: 0.1864
Epoch 92/100
441/441 [==============================] - 15s 34ms/step - loss: 3.8295 - accuracy: 0.1867
Epoch 93/100
441/441 [==============================] - 15s 34ms/step - loss: 3.8253 - accuracy: 0.1877
Epoch 94/100
441/441 [==============================] - 15s 34ms/step - loss: 3.8212 - accuracy: 0.1881
Epoch 95/100
441/441 [==============================] - 15s 35ms/step - loss: 3.8174 - accuracy: 0.1872
Epoch 96/100
441/441 [==============================] - 15s 35ms/step - loss: 3.8110 - accuracy: 0.1887
Epoch 97/100
441/441 [==============================] - 15s 35ms/step - loss: 3.8075 - accuracy: 0.1878
Epoch 98/100
441/441 [==============================] - 15s 34ms/step - loss: 3.8051 - accuracy: 0.1891
Epoch 99/100
441/441 [==============================] - 15s 35ms/step - loss: 3.7983 - accuracy: 0.1893
Epoch 100/100
441/441 [==============================] - 15s 35ms/step - loss: 3.7944 - accuracy: 0.1899
```

Out[11]:
```
<keras.callbacks.History at 0x2e35e033510>
```

In [12]:
```python
print (X.shape)
prediction = model.predict(X[0].reshape(1,sequence_length))
print (prediction.shape)
print (prediction)
```

```
(88470, 2)
1/1 [==============================] - 4s 4s/step
(1, 6673)
[[4.4535409e-19 1.6637336e-02 2.6662663e-02 ... 2.7235183e-33
  2.2977804e-29 1.0968266e-32]]
```

In [13]:
```python
test = ['thank you',
'welcome to',
'when there',
'more than',
'it cannot',
'is that',
'although this',
'do you',
'I was',
'the only',
'a great',
'thats very']
```

In [14]:
```python
for t in test:
    example = tokenizer.texts_to_sequences([t])
    prediction = model.predict(np.array(example))
    predicted_word = np.argmax(prediction)
    reverse_word_map = dict(map(reversed, tokenizer.word_index.items()))
    print ("{0} -> {1}".format(t, reverse_word_map[predicted_word]))
```

```
1/1 [==============================] - 0s 15ms/step
thank you -> will
1/1 [==============================] - 0s 15ms/step
welcome to -> the
1/1 [==============================] - 0s 15ms/step
when there -> is
1/1 [==============================] - 0s 15ms/step
more than -> the
1/1 [==============================] - 0s 14ms/step
it cannot -> be
1/1 [==============================] - 0s 16ms/step
is that -> i
1/1 [==============================] - 0s 19ms/step
although this -> time
1/1 [==============================] - 0s 15ms/step
do you -> have
1/1 [==============================] - 0s 15ms/step
I was -> a
1/1 [==============================] - 0s 15ms/step
the only -> drawback
1/1 [==============================] - 0s 15ms/step
a great -> prison
1/1 [==============================] - 0s 469ms/step
thats very -> visitor
```

In [ ]: