Find out the number of unique dialogue speakers in the sample conversation?

In [1]:
```
import pandas as pd
```

In [2]:
```
df = pd.read_csv(r'C:\Users\kantj\OneDrive\Desktop\Internshala\Python\Project\Python for DS- Project\Project Document\conv.txt',
```

In [3]:
```
df.rename(columns = {0:'Name',1:'Conv'}, inplace = True)
```

In [4]:
```
df
```

Out[4]:

	Name	Conv
0	WILL	I've never seen wildlings do a thing like thi...
1	WAYMAR ROYCE	How close did you get?
2	WILL	Close as any man would.
3	GARED	We should head back to the wall.
4	ROYCE	Do the dead frighten you?
...	...	...
106	NED	Tell me about Jon Arryn.
107	ROBERT	One minute he was fine, and then ... Burned rig...
108	NED	We both did.
109	ROBERT	He never had to teach you much, but me ... You ...
110	NED	Aye.

111 rows × 2 columns

In [5]:
```
len(df.Name.unique().tolist())
```

Out[5]:
```
17
```

Create a new text file by the name of the dialogue speaker and store the unique words spoken by that character in the respective text file. Make

sure there is only one word every line.

In [6]:
```
import nltk
```

In [7]:
```
def preprocess_text(text):
    lower_text = text.lower()
    tokens = nltk.tokenize.word_tokenize(lower_text)
    return tokens
```

In [8]:
```
for i in df.Name.unique():
    i = df[df['Name'] == i]['Conv'].to_string()
```

In [9]:
```
tokenized_final = []

for i in df.Name.unique():

    f= open("%s.txt" % i,"a+")

    token = preprocess_text(df[df['Name'] == i]['Conv'].to_string())

    tokenized_final.append(token)

    flattened_tokenized_final = [i for j in tokenized_final for i in j]

    for words in flattened_tokenized_final:
        f.write("%s\n" %words)
    f.close()
```

In [ ]:
```

```