Problem Statement

Tasks

1. Find out the number of unique speakers in the sample conversation.
2. Create a new text file by the name of the speaker and store unique words spoken by the respective speaker in the text file. (Note: only one word in every line.)

In [82]:
```
Conversation = open("conv.txt")
#To open the txt file

line = Conversation.readlines()
#To store lines in list format
```

In [83]:
```
line
```

Out[83]:
```
['WILL: Iâ€™ve never seen wildlings do a thing like this. Iâ€™ve never seen a thing like this, not ever in my life.\n',
 '\n',
 'WAYMAR ROYCE: How close did you get?\n',
 '\n',
 'WILL: Close as any man would.\n',
 '\n',
 'GARED: We should head back to the wall.\n',
 '\n',
 'ROYCE: Do the dead frighten you?\n',
 '\n',
 'GARED: Our orders were to track the wildlings. We tracked them. They wonâ€™t trouble us no more.\n',
 '\n',
 'ROYCE: You donâ€™t think heâ€™ll ask us how they died? Get back on your horse.\n',
 '\n',
 'WILL: Whatever did it to them could do it to us. They even killed the children.\n',
 '\n',
 'ROYCE: Itâ€™s a good thing weâ€™re not children. You want to run away south, run away. Of course, they will behead you as a deserter â€¦ If I donâ€™t catch you first. Get back o
n your horse. I wonâ€™t say it again.\n',
 '\n',
```

In [84]:
```
speakers = set()
#A set to store the names of all unique speakers

for lines in line:
    if lines !='\n':
    # to find new lines
        lines = lines.split(":")
    # breaking the lines into two parts ie. speaker and their dialogue
        speaker = lines[0]
    # on 0th index, there is speaker
        dialogue = lines[1]
    # on 1st index, there is respective dialogue
        dialogue_lower= dialogue.lower()
    # To convert the dialogue into lower case to because python is case sensitive
        speakers.add(speaker)
    # Storing name of all unique speakers in speakers
print("Total number of speakers are:", len(speakers))
```

```
Total number of speakers are: 17
```

In [85]:
```
print(speakers)
```

```
{'THEON', 'CATELYN', 'JAIME', 'WILL', 'GARED', 'SEPTA MORDANE', 'WAYMAR ROYCE', 'NED', 'CASSEL', 'CERSEI', 'SANSA', 'ROBERT', 'ARYA', 'ROBB', 'JON', 'BRAN', 'ROYCE'}
```

In [86]:
```
data={}
# Dictionary for storing key value pairs

for speaker in speakers:
    data[speaker]=set()
# Creating a set for each speaker to store unique words
```

In [87]:
```
data
```

Out[87]:
```
{'THEON': set(),
 'CATELYN': set(),
 'JAIME': set(),
 'WILL': set(),
 'GARED': set(),
 'SEPTA MORDANE': set(),
 'WAYMAR ROYCE': set(),
 'NED': set(),
 'CASSEL': set(),
 'CERSEI': set(),
 'SANSA': set(),
 'ROBERT': set(),
 'ARYA': set(),
 'ROBB': set(),
 'JON': set(),
 'BRAN': set(),
 'ROYCE': set()}
```

In [88]:
```
for lines in line:
    if lines != '\n':
        lines = lines.split(":")
        speaker = lines[0]
        dialogue = lines[1]
        dialogue_lower= dialogue.lower()
        dialogue_lower_strip=[dialogue_lower.strip('.,!;()[]?…') for dialogue_lower in dialogue_lower]
        # for removing special characters from the dialogue

        string= ''.join(map(str, dialogue_lower_strip))
        # To convert list into string

        data[speaker].update(string.split())
        # To update each speaker set with unique words spoken by them by separating each word by split function
```

In [89]:
```
data['ROYCE']
```

Out[89]:
```
{'a',
 'again',
 'as',
 'ask',
 'away',
 'back',
 'behead',
 'camp',
 'catch',
 'children',
 'course',
 'dead',
 'deserter',
 'died',
 'do',
 'donâ€™t',
 'first',
 'frighten',
 'get',
 'good',
 'have',
 'heâ€™ll',
 'horse',
 'how',
 'i',
 'if',
 'is',
 'it',
 'itâ€™s',
 'men',
 'moved',
 'not',
 'of',
 'on',
 'run',
 'say',
 'seem',
 'south',
 'the',
 'they',
 'thing',
 'think',
 'to',
 'us',
 'want',
 'weâ€™re',
 'what',
 'will',
 'wonâ€™t',
 'you',
 'your',
 'â€¦'}
```

In [90]:
```
data.keys()
```

Out[90]:
```
dict_keys(['THEON', 'CATELYN', 'JAIME', 'WILL', 'GARED', 'SEPTA MORDANE', 'WAYMAR ROYCE', 'NED', 'CASSEL', 'CERSEI', 'SANSA', 'ROBERT', 'ARYA', 'ROBB', 'JON', 'BRAN', 'ROYCE'])
```

In [92]:
```
for speakers in data.keys():
    # To iterate loop for each speaker

    text_file = open(str(speaker)+ '.txt', mode= "w")
    # It will create a new text file for each speaker

    words = list(data[speaker])
    # To store all dialogue in list

    text_file.writelines("\n".join(words))
    # Writelines will insert words from list into text file
    # With "\n".join, all words will come one in a line
```

In [94]:
```
Conversation.close()
```