Python Project for Data Science

Problem Statement ¶

Tasks

1. Find out the number of unique dialogue speakers in the sample conversation?
2. 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 [15]:

```python
script = open("conv.txt")
# to open the file## Problem Statement

line = script.readlines()
# using readlines to store lines in list format
```

In [16]:

```python
line
```

Out[16]:

```
['WILL: Iâ€™ve never seen wildlings do a thing like this. Iâ€™ve never see
n 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 won
â€™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 ch
```

In [17]:

```python
speakers = set()
# storing speakers in set to get only unique speakers

for lines in line:

    if lines != '\n':
    # to see the lines without new line character '\n'

        lines = lines.split(":")
    #   breaking lines in two part with the help of split

        speaker = lines[0]
    #   on 0th index it will be speakers

        dialogue = lines[1]
    #   on 1st index it will be dialogues it will be good to get data easily

        speakers.add(speaker)
    #   storing unique speakers on the speakers
```

In [18]:

```python
print("total no of speakers are: ", len(speakers))
```

```
total no of speakers are:   17
```

In [19]:

```python
print(speakers)
```

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

In [20]:

```python
data = {}
#created dictionary to store data in key value pair

for speaker in speakers:

    data[speaker] = set()
#   creating each set for every speaker to get unique words spoken by the speaker
```

In [21]:

```
data
```

Out[21]:

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

In [22]:

```python
for lines in line:

    if lines != '\n':

        lines = lines.split(":")

        speaker = lines[0]

        dialogue = lines[1]

        data[speaker].update(dialogue.split())
#       inserted dialogues for the every speaker and seperated each word by split function
```

In [23]:

```python
data['BRAN']
```

Out[23]:

```
{'But',
 'Down',
 'Heâ€™s',
 'I',
 'Is',
 'Jon',
 'NO!',
 'Our',
 'Please,',
 'So',
 'Their',
 'Walkers?',
 'What',
 'Where',
 'White',
 'a',
 'about',
 'coming',
 'dead.',
 'deserter.',
 'father!',
 'go?',
 'got',
 'he',
 'heâ€™s',
 'hundreds',
 'is',
 'it',
 'king!',
 'lying?',
 'motherâ€™s',
 'now!',
 'of',
 'old',
 'our',
 'people!',
 'right',
 'road!',
 'said',
 'saw',
 'the',
 'they',
 'true',
 'was',
 'way',
 'way?',
 'will',
 'you?'}
```

In [24]:

```python
data['JON']
```

Out[24]:

```
{'And',
 'Bran.',
 'Donâ€™t',
 'Father',
 'Fatherâ€™s',
 'Get',
 'Go',
 'House.',
 'Iâ€™m',
 'Lord',
 'Now',
 'One',
 'Stark',
 'Stark.',
 'Stark?',
 'The',
 'There',
 'They',
 'What',
 'You',
 'a',
 'are',
 'children.',
 'did',
 'direwolf',
 'do.',
 'each',
 'five',
 'five.',
 'for',
 'have',
 'hold',
 'if',
 'is',
 'it?',
 'know',
 'meant',
 'mother.',
 'much,',
 'not',
 'of',
 'on.',
 'pups.',
 'sigil',
 'the',
 'them.',
 'there',
 'think',
 'to',
 'too',
 'want',
 'watching.',
 'well.',
 'were',
 'will',
```

```
    'you',
    'your'}
```

In [10]:

```python
for speaker in data.keys():
#    used for loop to iterate to every speaker

    text_file = open(str(speaker) + '.txt', mode = "w")
#    it will create txt file for every speaker in write mode.

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

    words = [word.strip('.,!;()[]?') for word in words]
#    removed all special charaters from the words with the help of strip function

    text_file.writelines("\n".join(words))
#    with the help of writelines inserted all words from the list in text file.
#    with the help of join inserted "\n" so all words come one by one
```

```python
script.close()
```

In [ ]: