Assignment 4

Steps to Install standfordCoreNLP:

1. Download Stanford CoreNLP (link)

2. Install Java 8 (for mac): Open terminal and use below commands:
   a. brew update
   b. brew install jenv
   c. brew cask install java
3. (if brew is not installed then first intsall brew on mac)

4. Running Stanford CoreNLP Server: change directory to the path of the unzipped Stanford CoreNLP and execute the below command:
   a. java -mx4g -cp "*" edu.stanford.nlp.pipeline.StanfordCoreNLPServer -annotators "tokenize,ssplit,pos,lemma,parse,sentiment" -port 9000 -timeout 30000

5. Accessing Stanford CoreNLP Server using Python
   a. Download pycorenlp: pip install pycorenlp

Designing of Opinion Extraction Module with CoreNLP

For task 1, first we need to install standfordCoreNLP (instructions given above)

Annotator used: tokenize, sentiment, ner, pos
Reason why I chose above annotators for designing my opinion extraction module:
tokenize: To detect words in a sentence
pos: To get the tagging of a word which I have further utilized to extract only relevant 'enhanced dependencies'.
Sentiment: To identify the sentiment of a query and retrieve only matching sentiment query.

Objective: To create a dictionary 'extracted_opinions' of format {opinion: [list_of_docs]} where opinion = 'attribute', 'quality'

Process:
To create a dictionary (extracted_opinions)"

1. First a dictionary mapping all the words in a review with its corresponding 'pos' is created.
2. for each sentence in a review, extracted 'enhanced dependencies' of following types:
   a) 'amod': Retrieved all amod type dependencies without any conditions. In this case opinion is stored as 'governorGloss', 'dependentGloss'
   b) 'nsubj': Retrieved only 'governorGloss' = 'JJ' ('adjective') and 'dependentGloss' = 'NN' ('noun'). In this case opinion is stored as 'dependentGloss', 'governorGloss'
   c) 'discourse': Retrieved only 'governorGloss' = 'NN' ('noun') and 'dependentGloss' = 'UH' ('interjection'). In this case opinion is stored as 'dependentGloss', 'governorGloss'

   To extract above dependencies, I have referred research paper on 'Aspect-Based Opinion Mining Using Dependency Relations" by Amani K Samha.('Amani K Samha, "Aspect-Based Opinion Mining Using Dependency Relations", International Journal of Computer Science Trends and Technology (IJCST) – Volume 4 Issue 1, Jan - Feb 2016')
3. After extracting relevant opinions from a review, I stored it in the extracted_opinions along with the review_id. So, for each review we have multiple dictionary entry.

Opinion Similarity Score Calculation:

1. Split query_opinion to get 'attribute' and 'quality' of the query_opinion
2. Split each opinion in extracted_opinion into 2 parts – 'attribute' and 'quality'
3. Check if the 'attribute' and 'quality' of an opinion exist in our word embedding – 'assign4_word2vec1.bin'
4. Now check the sentiment of the query quality and opinion quality.
5. If opinion's 'attribute' and 'quality' exist in our word embedding and sentiments of the query and the opinion is same, then calculate similarity between query and opinion. To calculate similarity between a query and an opinion calculate:
   a. Attribute similarity between the query and the opinion
   b. Quality similarity between the query and the opinion
6. Decide a threshold of similarity to retrieve only relevant opinions.
7. Tuning cosine similarity threshold:
   a. cosine similarity threshold for attribute is kept at 0.33 to retrieve opinion attributes like 'salad', 'fries', 'taco' for query attribute 'food'
   Below is an output snippet to support my logic:
   ```
   query opinion [food, delicious] has similar opinions:
   query attribute: food, opinion attribute: salad,similarity
   score: 0.36177757
   query attribute: food, opinion attribute: turkey,similarity
   score: 0.3526584
   query attribute: food, opinion attribute: taco,similarity
   score: 0.34466738
   ```

```
query attribute: food, opinion attribute: fries,similarity
score: 0.337115
```
b. cosine similarity threshold for attribute is kept at 0.2 to retrieve opinion
   attributes like 'tender ', 'excellent', 'good' for query attribute 'delicious'
   Below is an output snippet to support my logic:
```
query opinion [food, delicious] has similar opinions:
query quality: delicious, opinion quality: tender,similarity s
core: 0.20179237
query quality: delicious, opinion quality: excellent,similarit
y score: 0.29940972
query quality: delicious, opinion quality: good,similarity sco
re: 0.28821617
```

Successful Cases:

- I am filtering by common sentiments to remove cases where there are strong
  similarities between negative and positive words. For eg for query "service
  good" similarity between 'bad' and 'good' > 0.7. Filtering only on the basis
  of scores is not efficient here that's why I implement filtering by common
  sentiments. So a positive query will extract only positive opinions.

```
query opinion [service, good] has similar opinions:
[good, excellent: 0.6442926
[good, great: 0.72915095
[good, warm: 0.3033841
[good, solid: 0.5806034
[good, good: 1.0

        [service, excellent] appears in review  [ 1 ,   2 ]

        [service, great] appears in review      [ 5 ,   1 4 ]

        [service, warm] appears in review       [ 8 ]

        [service, solid] appears in review      [ 1 3 ]

        [service, good] appears in review       [ 2 0 ]
```
Note: [service, bad] is not retrieved

- I decided to have separate cosine similarity threshold value for attribute and
  quality to design a better retrieval system. A lower cosine similarity threshold
  is good for quality but not for attributes (given a common sentiment) as
  threshold at 0.2 will retrieve some irrelevant attributes like 'food' for

'service' and a higher threshold in quality will reject some relevant opinions as well. For instance 'good' and 'delicious' have similarity score 0.28.

```
query opinion [food, delicious] has similar opinions:
[delicious, wonderful: 0.50828356
[delicious, delicious: 1.0
[delicious, tender: 0.20179237
[delicious, flavorful: 0.74283195
[delicious, delicious: 1.0
[delicious, nice: 0.36657465
[delicious, great: 0.34220225
[delicious, excellent: 0.29940972
[delicious, hearty: 0.5545052
[delicious, roasted: 0.46554118
[delicious, fresh: 0.33691075
[delicious, interesting: 0.31214893
[delicious, good: 0.28821617
[delicious, satisfying: 0.47659993
[delicious, great: 0.34220225
```

Cases where system fails:

- My algorithm fails to retrieve opinion attribute 'waiter', 'server' for query attribute 'service' as they have a very low cosine similarity and lowering down threshold to retrieve these cases will retrieve some irrelevant opinions as well.
- My opinion retrieval design system is limited to 2 words opinion only and hence attributes like 'bread stick', 'California salad' are stored in the system as 'stick' and 'salad' respectively.

Output:

(base) rishusingh@Rishus-MacBook-Pro Assignment 4 % python Assignment4Main.py

[co-worker, old] appears in review [1]

[meal, wonderful] appears in review   [1]

[salad, huge] appears in review   [1]

[salad, delicious] appears in review   [1]

[stick, little] appears in review   [1]

[stick, delicious] appears in review   [1]

[service, excellent] appears in review   [1 , 2]

[place, great] appears in review   [1]

[experience, upscale] appears in review   [1]

[spots, other] appears in review   [1]

[school, old] appears in review   [2]

[woodwork, ornate] appears in review   [2]

[tablecloths, white] appears in review   [2]

[potatoes, red] appears in review   [2]

[skinned, mashed] appears in review   [2]

[meat, tender] appears in review   [2]

[meat, flavorful] appears in review [2]

[slaw, delicious] appears in review   [2]

[atmosphere, nice] appears in review   [3]

[list, local] appears in review   [3]

[list, great] appears in review   [3]

[potions, huge] appears in review [3]

[menu, huge] appears in review   [3]

[thing, good] appears in review   [3]

[note, positive] appears in review [4]

[meal, delicious] appears in review   [4]

[sandwich, grilled] appears in review   [4]

[toppings, own] appears in review [4]

[ambiance, nice] appears in review   [4]

[selection, great] appears in review   [4]

[restaurant, small] appears in review   [4 ,   5]

[table, own] appears in review   [4]

[ride, useless] appears in review   [5]

[meals, best] appears in review   [5]

[fan, huge] appears in review   [5]

[service, great] appears in review   [5 ,   1 4]

[sandwiches, french] appears in review   [5]

[restaurant, nice] appears in review   [5]

[atmosphere, great] appears in review   [5]

[food, great] appears in review   [5]

[rotation, regular] appears in review   [6]

[food, excellent] appears in review   [6]

[deserts, recent] appears in review   [6]

[deserts, mini] appears in review   [6]

[deal, great] appears in review   [6]

[selection, good] appears in review   [6]

[quality, bad] appears in review   [7]

[service, rude] appears in review   [7]

[day, bad] appears in review   [7]

[food, hearty] appears in review   [8]

[prices, decent] appears in review [8]

[service, warm] appears in review [8]

[fish, battered] appears in review   [8]

[good, bad] appears in review   [8]

[service, bad] appears in review   [9]

[fingers, homemade] appears in review [1 0]

[value, great] appears in review   [1 0]

[menu, large] appears in review   [1 1]

[quality, excellent] appears in review   [1 1]

[menu, entire] appears in review   [1 1]

[specialities, other] appears in review   [1 1]

[5, top] appears in review   [1 2]

[favorites, local] appears in review   [1 2]

[feeling, warm] appears in review [1 2]

[deal, cheapest] appears in review   [1 2]

[order, particular] appears in review   [1 2]

[course, main] appears in review   [1 2]

[onion, french] appears in review   [1 2]

[summer, last] appears in review   [1 3]

[time, wonderful] appears in review   [1 3]

[service, solid] appears in review   [1 3]

[group, larger] appears in review   [1 3]

[turkey, roasted] appears in review   [1 4]

[food, fresh] appears in review   [1 4]

[food, interesting] appears in review   [1 4]

[taco, good] appears in review   [1 5]

[waiter, slow] appears in review   [1 5]

[time, long] appears in review   [1 5]

[thing, second] appears in review [1 5]

[food, cold] appears in review   [1 5]

[group, large] appears in review   [1 6]

[food, satisfying] appears in review   [1 6]

[food, typical] appears in review   [1 6]

[food, american] appears in review   [1 6]

[way, bad] appears in review   [1 6]

[options, vegan] appears in review   [1 6]

[salad, large] appears in review   [1 7]

[salad, greek] appears in review   [1 7]

[nachos, good] appears in review [1 7]

[fries, great] appears in review   [1 7]

[waiter, friendly] appears in review   [1 7]

[waiter, attentive] appears in review   [1 7]

[wines, decent] appears in review [1 8]

[restaurants, more] appears in review   [1 9]

[restaurants, busy] appears in review   [1 9]

[hopes, high] appears in review   [1 9]

[service, good] appears in review [2 0]

[ambience, pleasant] appears in review [2 0]

[pittsburghers, older] appears in review [2 0]

[lot, expensive] appears in review [2 0]

----

query opinion [service, good] has similar opinions:

    [service, excellent] appears in review   [1 ,   2]

    [service, great] appears in review [5 ,   1 4]

    [service, warm] appears in review [8]

    [service, solid] appears in review   [1 3]

[service, good] appears in review [2 0]

query opinion [service, bad] has similar opinions:

[service, rude] appears in review   [7]

[service, bad] appears in review   [9]

query opinion [atmosphere, good] has similar opinions:

[atmosphere, nice] appears in review   [3]

[ambiance, nice] appears in review   [4]

[atmosphere, great] appears in review   [5]

[feeling, warm] appears in review [1 2]

[ambience, pleasant] appears in review [2 0]

query opinion [food, delicious] has similar opinions:

[meal, wonderful] appears in review   [1]

[salad, delicious] appears in review   [1]

[meat, tender] appears in review   [2]

[meat, flavorful] appears in review [2]

[meal, delicious] appears in review   [4]

[restaurant, nice] appears in review   [5]

[food, great] appears in review   [5]

[food, excellent] appears in review   [6]

[food, hearty] appears in review   [8]

[turkey, roasted] appears in review   [1 4]

[food, fresh] appears in review   [1 4]

[food, interesting] appears in review   [1 4]

[taco, good] appears in review   [1 5]

[food, satisfying] appears in review   [1 6]

[fries, great] appears in review   [1 7]