# Exercise 1 :)

Sol: Please find gate\_default.txt which contains POS tags. Tagging is done using opinion finder tool.

# Exercise2 :)

Sol: Please find list of nouns in **ex\_2\_noun.txt**. I have also kept count alongside the noun.

I have used **Exercise2.py** python script to parse the gate\_default.txt and get the noun (lemma).

Please keep gate\_default.txt in working directory before running the script.

The script will generate ex\_2\_noun.txt file containing list of nouns.



# Exercise3 :)

Please find top 10 frequent aspects in deceasing order of their support.

camera : 227

g3 : 79

picture : 64

canon : 56

lens : 44

flash : 41

quality : 40

camera,g3 : 38

time : 38

image : 33

Please find following script for exercise 3:

|  |  |
| --- | --- |
| **SCRIPT** | **FUNCTIONALITY** |
| Exercise3.py | It generates comma separated aspects from each tagged sentence which is used as input to fp growth algorithm. The file **fp\_growth\_input.txt** is generated which is given to fp growth algorithm. |
| Exercise3b.py | It also generates comma separated aspects form each sentence. But the script also includes consecutive nouns as single aspect. Example: “ picture quality”, “shutter speed”, “diopter adjustment dial” etc. |
| fp\_growth.py | It takes comma separated aspects as transaction and generates frequent aspects from all transactions i.e. sentences. We have to total for 598 sentences / transactions . |

\*Note: run **fp\_growth.py** script as: **python fp\_growth.py fp\_growth\_input.txt**

# Exercise 4 :)

Please find top 10 most frequently used opinion words:

not:63  
digital:56  
great:43  
very:29  
best:26  
easy:22  
so:22  
more:21  
external:17  
most:17  
raw:17  
excellent:16  
optical:16  
even:15  
just:15  
other:15

Please find **Exercise4.py**. It generates **opinion\_wrd.txt** containing list of all the opinion words which are found 5 words distance from frequent aspect which we got from fp growth algorithm.

# Exercise 5 :)

Please find the precision, recall and F1 score for the aspects identified in Exercise 3 and tagged data.

True Positive: 20

False Negative: 105

False Positive: 60

|  |
| --- |
| Precision: 0.25 |
| Recall: 0.16 |
| F1 score: 0.19512195122 |

Please find **Exercise5.py .** It reads and parses the frequent aspectfrom tagged data set. And also it reads the identified frequent aspect file. Then the script computes parameter like true positive, false positive and false negative for computing Precision, Recall and F1 score.

# Exercise 6 :)

Please find the positive and negative opinion words count:

Positive opinion words: 89

Negative opinion words: 28

**Positive opinion words:**

great ,best ,easy ,excellent ,good ,well ,better ,available ,enough ,perfect ,wonderful ,beautiful ,fairly ,amazing ,happy ,nice ,outstanding ,stunning ,worth ,accurate ,awesome ,comfortable ,compact ,consistently ,easier ,hot ,incredible ,positive ,proven ,responsive ,advanced ,clean ,cool ,dedicated ,fantastic ,faster ,fine ,finest ,greatest ,ideal ,intuitive ,lighter ,mind-blowing ,modern ,panoramic ,powerful ,reasonable ,right ,seamless ,sharp ,solid ,terrific ,useful ,well-behaved ,astonishingly ,astounding ,brilliant ,clear ,creative ,delicate ,fabulous ,finer ,flexible ,friendly ,glad ,helpful ,inexpensive ,keen ,leading ,neat ,nicely ,obtainable ,perfectly ,pleased ,precious ,pretty ,proper ,pure ,ready ,realistic ,sharper ,sophisticated ,spectacular ,steady ,stunned ,super ,topnotch ,user-friendly ,vivid

**Negative opinion words:**

['expensive', 'slow', 'fragile', 'harsh', 'negative', 'bad', 'blurry', 'disappointing', 'loose', 'poor', 'slowly', 'utterly', 'annoying', 'bulky', 'defective', 'die-hard', 'dim', 'insanely', 'limited', 'long-time', 'needless', 'pitiful', 'proprietary', 'slower', 'strictly', 'ugly', 'unnerving', 'unsatisfactory']

Please use **Exercises6.py** to obtain positive and negative word lists.

# Appendix:

To validate python scripts, please follow below steps:

1. Please keep the files CanonG3\_tagged.txt and gate\_default in working directory.
2. Run python scripts in below sequence

a) python Exercise2.py

b) python Exercise3.py

c) python fp\_growth.py fp\_growth\_input.txt

d) python Exercise4.py

e) python Exercise5.py

f) python Exercise6.py