Coccurance Matrix function

In [1]:
```
#Import library
import pandas as pd
import numpy as np
```

In [2]:
```
#Function
def cooccurance_matrix(Corpus,feature_list,window_size): #Input variables Corp
us - Whole string array,feature_list-Feature list array,window_size-Window siz
e
    #Convert to DataFrame
    df_Corpus=pd.DataFrame(Corpus)
    #Creating two list of features
    all_feature=feature_list
    all_feature_2=all_feature

    #Create an list of features combination
    index_column_set = [[index, column] for index in all_feature
              for column in all_feature_2 if index != column]

    #Created an dataframe
    array_of0=np.zeros((len(feature_list),len(feature_list)))
    df_all = pd.DataFrame(array_of0, columns =all_feature, index=all_feature )

    if_else_lamda=lambda x: x if (x>0) else 0 #Lamda function made the index t
o zero if leass then zero

    for l in index_column_set: #For every index and column set
        f_count=0
        for sentance in df_Corpus[0].values: #For each sentances
            k=sentance.split() #Split the sentances
            res_list = list(filter(lambda x: k[x] == l[0], range(len(k)))) #If
index found in sentance then create the array of positions
            if len(res_list) >0:#if the array not null
                c=0
                for i in res_list:#For every potion
                    ind_x=if_else_lamda(i - window_size) #call Lamda function
                    c=c+k[ind_x:i+window_size].count(l[1]) #Count the column o
ccurance
                f_count=f_count+c #Sum with previous count
        df_all[l[0]][l[1]]=f_count#Finally assign the value in the dataframe i
ndex column
    return df_all
```

In [3]:
```
#Test Purpose
whole_doc=['ABC DEF IJK PQR','PQR KLM OPQ','LMN PQR XYZ ABC DEF PQR ABC']
feature_list=['ABC','PQR','DEF']
k=cooccurance_matrix(whole_doc,feature_list,6)
```

In [4]:
```
#Print Coccurance Matrix
k
```

Out[4]:

	ABC	PQR	DEF
ABC	0.0	5.0	3.0
PQR	5.0	0.0	3.0
DEF	3.0	3.0	0.0