Maitrey Prajapati - 10445262

Program:

sc = SparkContext.getOrCreate()

f = sc.textFile("data.txt")

f = f.map(lambda x : x.replace(".","").replace(" ","")).collect()

bigrams = sc.parallelize((element[i]+element[i+1],1) for element in f for i in range(0,len(element),2))

bigrams = bigrams.reduceByKey(lambda x,y: x+y)

sortedBigrams = bigrams.sortBy(lambda x : x[1]).collect()

##Printing lowest occurring bigrams

previous = sortedBigrams[0][1]

index = 0

print("Lowest Occuring")

while(sortedBigrams[index][1]==previous):

print(sortedBigrams[index])

index+=1

##Printing highest occurring bigrams

previous = sortedBigrams[-1][1]

index = -1

print("Highest Occuring")

while(sortedBigrams[index][1]==previous):

print(sortedBigrams[index])

index-=1

Output:

Lowest Occurring

('66', 1)

('93', 1)

('19', 1)

('38', 1)

('51', 1)

Highest Occurring

('68', 10)

('21', 10)