**r1 = sc.textFile('Documents/a3\_IP')  
>>> r1.collect()**[u'doc6:of,crazy,fox,jumped,fox,ran', u'doc6:fox,ran,fast,over,fence,too,high', u'doc1:fox,jumped', u'doc1:fox,jumped,over,the,fence', u'doc1:of,crazy,fox,jumped,fox,ran', u'doc1:fox,ran,fast,over,fence,too,high', u'doc2:a,crazy,fox,jumped', u'doc2:a,crazy,fox,jumped,over,the,fence,again', u'doc3:fox,ran,fast,over,fence', u'doc1:fox,is,high,on,sugar,but,jumped,but,failed', u'doc2:a,crazy,fox,ran,ran,ran,fast', u'doc3:a,crazy,fox,jumped,jumped,jumped,jumped', u'doc4:a,crazy,fox,ran,jumped,ran,jumped', u'doc4:a,crazy,fox,jumped,jumped', u'doc4:a,crazy,fox,jumped,jumped', u'doc5:a,crazy,fox,jumped,over,fence,very,high', u'doc5:a,crazy,fox,jumped,over,the,fence,again', u'doc6:book,reading,about,fox,and,fence,jumped,but,failed', u'doc1:crazy,fox,jumped,but,failed']  
  
**>> r2 = r1.map(lambda x:(x.split(':')[0],x.split(':')[1].split(',')))  
>>> r2.collect()**  
[(u'doc6', [u'of', u'crazy', u'fox', u'jumped', u'fox', u'ran']), (u'doc6', [u'fox', u'ran', u'fast', u'over', u'fence', u'too', u'high']), (u'doc1', [u'fox', u'jumped']), (u'doc1', [u'fox', u'jumped', u'over', u'the', u'fence']), (u'doc1', [u'of', u'crazy', u'fox', u'jumped', u'fox', u'ran']), (u'doc1', [u'fox', u'ran', u'fast', u'over', u'fence', u'too', u'high']), (u'doc2', [u'a', u'crazy', u'fox', u'jumped']), (u'doc2', [u'a', u'crazy', u'fox', u'jumped', u'over', u'the', u'fence', u'again']), (u'doc3', [u'fox', u'ran', u'fast', u'over', u'fence']), (u'doc1', [u'fox', u'is', u'high', u'on', u'sugar', u'but', u'jumped', u'but', u'failed']), (u'doc2', [u'a', u'crazy', u'fox', u'ran', u'ran', u'ran', u'fast']), (u'doc3', [u'a', u'crazy', u'fox', u'jumped', u'jumped', u'jumped', u'jumped']), (u'doc4', [u'a', u'crazy', u'fox', u'ran', u'jumped', u'ran', u'jumped']), (u'doc4', [u'a', u'crazy', u'fox', u'jumped', u'jumped']), (u'doc4', [u'a', u'crazy', u'fox', u'jumped', u'jumped']), (u'doc5', [u'a', u'crazy', u'fox', u'jumped', u'over', u'fence', u'very', u'high']), (u'doc5', [u'a', u'crazy', u'fox', u'jumped', u'over', u'the', u'fence', u'again']), (u'doc6', [u'book', u'reading', u'about', u'fox', u'and', u'fence', u'jumped', u'but', u'failed']), (u'doc1', [u'crazy', u'fox', u'jumped', u'but', u'failed'])]  
  
  
  
/\*  
**def f(r2):  
    ngrams=dict()  
    n=3  
    for i in range(len(r2) - n + 1):  
        gram = tuple(r2[i:i+n])  
        if gram in ngrams:  
            ngrams[gram] += 1  
        else:  
            ngrams[gram] = 1  
    return ngrams  
  
r3 = r2.mapValues(f)**  
\*/  
/\*  
**r3 = r2.map(lambda x: (x[0], [((x[1][i],x[1][i+1],x[1][i+2])) for i in range (0, len(x[1])-2)]))  
r3.collect()**\*/  
  
[(u'doc6', {(u'crazy', u'fox', u'jumped'): 1, (u'jumped', u'fox', u'ran'): 1, (u'fox', u'jumped', u'fox'): 1, (u'of', u'crazy', u'fox'): 1}), (u'doc6', {(u'over', u'fence', u'too'): 1, (u'fast', u'over', u'fence'): 1, (u'ran', u'fast', u'over'): 1, (u'fence', u'too', u'high'): 1, (u'fox', u'ran', u'fast'): 1}), (u'doc1', {}), (u'doc1', {(u'jumped', u'over', u'the'): 1, (u'fox', u'jumped', u'over'): 1, (u'over', u'the', u'fence'): 1}), (u'doc1', {(u'crazy', u'fox', u'jumped'): 1, (u'jumped', u'fox', u'ran'): 1, (u'fox', u'jumped', u'fox'): 1, (u'of', u'crazy', u'fox'): 1}), (u'doc1', {(u'over', u'fence', u'too'): 1, (u'fast', u'over', u'fence'): 1, (u'ran', u'fast', u'over'): 1, (u'fence', u'too', u'high'): 1, (u'fox', u'ran', u'fast'): 1}), (u'doc2', {(u'crazy', u'fox', u'jumped'): 1, (u'a', u'crazy', u'fox'): 1}), (u'doc2', {(u'a', u'crazy', u'fox'): 1, (u'jumped', u'over', u'the'): 1, (u'fox', u'jumped', u'over'): 1, (u'the', u'fence', u'again'): 1, (u'crazy', u'fox', u'jumped'): 1, (u'over', u'the', u'fence'): 1}), (u'doc3', {(u'fast', u'over', u'fence'): 1, (u'ran', u'fast', u'over'): 1, (u'fox', u'ran', u'fast'): 1}), (u'doc1', {(u'high', u'on', u'sugar'): 1, (u'fox', u'is', u'high'): 1, (u'jumped', u'but', u'failed'): 1, (u'but', u'jumped', u'but'): 1, (u'sugar', u'but', u'jumped'): 1, (u'on', u'sugar', u'but'): 1, (u'is', u'high', u'on'): 1}), (u'doc2', {(u'crazy', u'fox', u'ran'): 1, (u'ran', u'ran', u'ran'): 1, (u'a', u'crazy', u'fox'): 1, (u'fox', u'ran', u'ran'): 1, (u'ran', u'ran', u'fast'): 1}), (u'doc3', {(u'fox', u'jumped', u'jumped'): 1, (u'jumped', u'jumped', u'jumped'): 2, (u'crazy', u'fox', u'jumped'): 1, (u'a', u'crazy', u'fox'): 1}), (u'doc4', {(u'fox', u'ran', u'jumped'): 1, (u'crazy', u'fox', u'ran'): 1, (u'ran', u'jumped', u'ran'): 1, (u'a', u'crazy', u'fox'): 1, (u'jumped', u'ran', u'jumped'): 1}), (u'doc4', {(u'crazy', u'fox', u'jumped'): 1, (u'a', u'crazy', u'fox'): 1, (u'fox', u'jumped', u'jumped'): 1}), (u'doc4', {(u'crazy', u'fox', u'jumped'): 1, (u'a', u'crazy', u'fox'): 1, (u'fox', u'jumped', u'jumped'): 1}), (u'doc5', {(u'a', u'crazy', u'fox'): 1, (u'fence', u'very', u'high'): 1, (u'over', u'fence', u'very'): 1, (u'jumped', u'over', u'fence'): 1, (u'fox', u'jumped', u'over'): 1, (u'crazy', u'fox', u'jumped'): 1}), (u'doc5', {(u'a', u'crazy', u'fox'): 1, (u'jumped', u'over', u'the'): 1, (u'fox', u'jumped', u'over'): 1, (u'the', u'fence', u'again'): 1, (u'crazy', u'fox', u'jumped'): 1, (u'over', u'the', u'fence'): 1}), (u'doc6', {(u'fence', u'jumped', u'but'): 1, (u'about', u'fox', u'and'): 1, (u'reading', u'about', u'fox'): 1, (u'fox', u'and', u'fence'): 1, (u'jumped', u'but', u'failed'): 1, (u'and', u'fence', u'jumped'): 1, (u'book', u'reading', u'about'): 1}), (u'doc1', {(u'fox', u'jumped', u'but'): 1, (u'crazy', u'fox', u'jumped'): 1, (u'jumped', u'but', u'failed'): 1})]  
  
  
**>>> r4 = r3.map(lambda x: (x[0],len(x[1])))  
>>> r4.collect()**[(u'doc6', 4), (u'doc6', 5), (u'doc1', 0), (u'doc1', 3), (u'doc1', 4), (u'doc1', 5), (u'doc2', 2), (u'doc2', 6), (u'doc3', 3), (u'doc1', 7), (u'doc2', 5), (u'doc3', 4), (u'doc4', 5), (u'doc4', 3), (u'doc4', 3), (u'doc5', 6), (u'doc5', 6), (u'doc6', 7), (u'doc1', 3)]  
  
  
**r5 = r4.reduceByKey(lambda x,y:(x+y))  
>>> r5.collect()**[(u'doc2', 13), (u'doc3', 7), (u'doc1', 22), (u'doc6', 16), (u'doc4', 11), (u'doc5', 12)]

**>>> r6 = r5.map(lambda x:(1,x))  
>>> r6.collect()**[(1, (u'doc2', 13)), (1, (u'doc3', 7)), (1, (u'doc1', 22)), (1, (u'doc6', 16)), (1, (u'doc4', 11)), (1, (u'doc5', 12))]  
  
  
**>>> r7 = r6.join(r6)  
>>> r7.collect()**[(1, ((u'doc2', 13), (u'doc2', 13))), (1, ((u'doc2', 13), (u'doc3', 7))), (1, ((u'doc2', 13), (u'doc1', 22))), (1, ((u'doc2', 13), (u'doc6', 16))), (1, ((u'doc2', 13), (u'doc4', 11))), (1, ((u'doc2', 13), (u'doc5', 12))), (1, ((u'doc3', 7), (u'doc2', 13))), (1, ((u'doc3', 7), (u'doc3', 7))), (1, ((u'doc3', 7), (u'doc1', 22))), (1, ((u'doc3', 7), (u'doc6', 16))), (1, ((u'doc3', 7), (u'doc4', 11))), (1, ((u'doc3', 7), (u'doc5', 12))), (1, ((u'doc1', 22), (u'doc2', 13))), (1, ((u'doc1', 22), (u'doc3', 7))), (1, ((u'doc1', 22), (u'doc1', 22))), (1, ((u'doc1', 22), (u'doc6', 16))), (1, ((u'doc1', 22), (u'doc4', 11))), (1, ((u'doc1', 22), (u'doc5', 12))), (1, ((u'doc6', 16), (u'doc2', 13))), (1, ((u'doc6', 16), (u'doc3', 7))), (1, ((u'doc6', 16), (u'doc1', 22))), (1, ((u'doc6', 16), (u'doc6', 16))), (1, ((u'doc6', 16), (u'doc4', 11))), (1, ((u'doc6', 16), (u'doc5', 12))), (1, ((u'doc4', 11), (u'doc2', 13))), (1, ((u'doc4', 11), (u'doc3', 7))), (1, ((u'doc4', 11), (u'doc1', 22))), (1, ((u'doc4', 11), (u'doc6', 16))), (1, ((u'doc4', 11), (u'doc4', 11))), (1, ((u'doc4', 11), (u'doc5', 12))), (1, ((u'doc5', 12), (u'doc2', 13))), (1, ((u'doc5', 12), (u'doc3', 7))), (1, ((u'doc5', 12), (u'doc1', 22))), (1, ((u'doc5', 12), (u'doc6', 16))), (1, ((u'doc5', 12), (u'doc4', 11))), (1, ((u'doc5', 12), (u'doc5', 12)))]  
  
  
**>>> r8 = r7.filter(lambda x: x[1][0][0]<x[1][1][0])  
>>> r8.collect()**[(1, ((u'doc2', 13), (u'doc3', 7))), (1, ((u'doc2', 13), (u'doc6', 16))), (1, ((u'doc2', 13), (u'doc4', 11))), (1, ((u'doc2', 13), (u'doc5', 12))), (1, ((u'doc3', 7), (u'doc6', 16))), (1, ((u'doc3', 7), (u'doc4', 11))), (1, ((u'doc3', 7), (u'doc5', 12))), (1, ((u'doc1', 22), (u'doc2', 13))), (1, ((u'doc1', 22), (u'doc3', 7))), (1, ((u'doc1', 22), (u'doc6', 16))), (1, ((u'doc1', 22), (u'doc4', 11))), (1, ((u'doc1', 22), (u'doc5', 12))), (1, ((u'doc4', 11), (u'doc6', 16))), (1, ((u'doc4', 11), (u'doc5', 12))), (1, ((u'doc5', 12), (u'doc6', 16)))]  
>>>   
  
  
**r9 = r8.map(lambda x:((x[1][0][0],x[1][1][0]),(x[1][0][1],x[1][1][1])))  
>>> r9.collect()**[((u'doc2', u'doc3'), (13, 7)), ((u'doc2', u'doc6'), (13, 16)), ((u'doc2', u'doc4'), (13, 11)), ((u'doc2', u'doc5'), (13, 12)), ((u'doc3', u'doc6'), (7, 16)), ((u'doc3', u'doc4'), (7, 11)), ((u'doc3', u'doc5'), (7, 12)), ((u'doc1', u'doc2'), (22, 13)), ((u'doc1', u'doc3'), (22, 7)), ((u'doc1', u'doc6'), (22, 16)), ((u'doc1', u'doc4'), (22, 11)), ((u'doc1', u'doc5'), (22, 12)), ((u'doc4', u'doc6'), (11, 16)), ((u'doc4', u'doc5'), (11, 12)), ((u'doc5', u'doc6'), (12, 16))]  
>>>   
  
**>> r10 = r9.map(lambda x:((x[0]),(x[1][0]+x[1][1])))  
>>> r10.collect()**[((u'doc2', u'doc3'), 20), ((u'doc2', u'doc6'), 29), ((u'doc2', u'doc4'), 24), ((u'doc2', u'doc5'), 25), ((u'doc3', u'doc6'), 23), ((u'doc3', u'doc4'), 18), ((u'doc3', u'doc5'), 19), ((u'doc1', u'doc2'), 35), ((u'doc1', u'doc3'), 29), ((u'doc1', u'doc6'), 38), ((u'doc1', u'doc4'), 33), ((u'doc1', u'doc5'), 34), ((u'doc4', u'doc6'), 27), ((u'doc4', u'doc5'), 23), ((u'doc5', u'doc6'), 28)]  
>>>   
  
**def f(r3): return r3  
...   
>>> r11 = r3.flatMapValues(f)  
r11.collect()**

[(u'doc6', (u'crazy', u'fox', u'jumped')), (u'doc6', (u'jumped', u'fox', u'ran')), (u'doc6', (u'fox', u'jumped', u'fox')), (u'doc6', (u'of', u'crazy', u'fox')), (u'doc6', (u'over', u'fence', u'too')), (u'doc6', (u'fast', u'over', u'fence')), (u'doc6', (u'ran', u'fast', u'over')), (u'doc6', (u'fence', u'too', u'high')), (u'doc6', (u'fox', u'ran', u'fast')), (u'doc1', (u'jumped', u'over', u'the')), (u'doc1', (u'fox', u'jumped', u'over')), (u'doc1', (u'over', u'the', u'fence')), (u'doc1', (u'crazy', u'fox', u'jumped')), (u'doc1', (u'jumped', u'fox', u'ran')), (u'doc1', (u'fox', u'jumped', u'fox')), (u'doc1', (u'of', u'crazy', u'fox')), (u'doc1', (u'over', u'fence', u'too')), (u'doc1', (u'fast', u'over', u'fence')), (u'doc1', (u'ran', u'fast', u'over')), (u'doc1', (u'fence', u'too', u'high')), (u'doc1', (u'fox', u'ran', u'fast')), (u'doc2', (u'crazy', u'fox', u'jumped')), (u'doc2', (u'a', u'crazy', u'fox')), (u'doc2', (u'a', u'crazy', u'fox')), (u'doc2', (u'jumped', u'over', u'the')), (u'doc2', (u'the', u'fence', u'again')), (u'doc2', (u'fox', u'jumped', u'over')), (u'doc2', (u'crazy', u'fox', u'jumped')), (u'doc2', (u'over', u'the', u'fence')), (u'doc3', (u'fast', u'over', u'fence')), (u'doc3', (u'ran', u'fast', u'over')), (u'doc3', (u'fox', u'ran', u'fast')), (u'doc1', (u'high', u'on', u'sugar')), (u'doc1', (u'fox', u'is', u'high')), (u'doc1', (u'jumped', u'but', u'failed')), (u'doc1', (u'is', u'high', u'on')), (u'doc1', (u'sugar', u'but', u'jumped')), (u'doc1', (u'on', u'sugar', u'but')), (u'doc1', (u'but', u'jumped', u'but')), (u'doc2', (u'crazy', u'fox', u'ran')), (u'doc2', (u'ran', u'ran', u'ran')), (u'doc2', (u'a', u'crazy', u'fox')), (u'doc2', (u'ran', u'ran', u'fast')), (u'doc2', (u'fox', u'ran', u'ran')), (u'doc3', (u'jumped', u'jumped', u'jumped')), (u'doc3', (u'crazy', u'fox', u'jumped')), (u'doc3', (u'a', u'crazy', u'fox')), (u'doc3', (u'fox', u'jumped', u'jumped')), (u'doc4', (u'fox', u'ran', u'jumped')), (u'doc4', (u'crazy', u'fox', u'ran')), (u'doc4', (u'ran', u'jumped', u'ran')), (u'doc4', (u'a', u'crazy', u'fox')), (u'doc4', (u'jumped', u'ran', u'jumped')), (u'doc4', (u'crazy', u'fox', u'jumped')), (u'doc4', (u'a', u'crazy', u'fox')), (u'doc4', (u'fox', u'jumped', u'jumped')), (u'doc4', (u'crazy', u'fox', u'jumped')), (u'doc4', (u'a', u'crazy', u'fox')), (u'doc4', (u'fox', u'jumped', u'jumped')), (u'doc5', (u'a', u'crazy', u'fox')), (u'doc5', (u'fence', u'very', u'high')), (u'doc5', (u'over', u'fence', u'very')), (u'doc5', (u'jumped', u'over', u'fence')), (u'doc5', (u'fox', u'jumped', u'over')), (u'doc5', (u'crazy', u'fox', u'jumped')), (u'doc5', (u'a', u'crazy', u'fox')), (u'doc5', (u'jumped', u'over', u'the')), (u'doc5', (u'the', u'fence', u'again')), (u'doc5', (u'fox', u'jumped', u'over')), (u'doc5', (u'crazy', u'fox', u'jumped')), (u'doc5', (u'over', u'the', u'fence')), (u'doc6', (u'fence', u'jumped', u'but')), (u'doc6', (u'about', u'fox', u'and')), (u'doc6', (u'reading', u'about', u'fox')), (u'doc6', (u'fox', u'and', u'fence')), (u'doc6', (u'jumped', u'but', u'failed')), (u'doc6', (u'and', u'fence', u'jumped')), (u'doc6', (u'book', u'reading', u'about')), (u'doc1', (u'fox', u'jumped', u'but')), (u'doc1', (u'crazy', u'fox', u'jumped')), (u'doc1', (u'jumped', u'but', u'failed'))]  
>>> 

**r12 = r11.map(lambda x:(x,1))  
>>> r12.collect()**[((u'doc6', (u'crazy', u'fox', u'jumped')), 1), ((u'doc6', (u'jumped', u'fox', u'ran')), 1), ((u'doc6', (u'fox', u'jumped', u'fox')), 1), ((u'doc6', (u'of', u'crazy', u'fox')), 1), ((u'doc6', (u'over', u'fence', u'too')), 1), ((u'doc6', (u'fast', u'over', u'fence')), 1), ((u'doc6', (u'ran', u'fast', u'over')), 1), ((u'doc6', (u'fence', u'too', u'high')), 1), ((u'doc6', (u'fox', u'ran', u'fast')), 1), ((u'doc1', (u'jumped', u'over', u'the')), 1), ((u'doc1', (u'fox', u'jumped', u'over')), 1), ((u'doc1', (u'over', u'the', u'fence')), 1), ((u'doc1', (u'crazy', u'fox', u'jumped')), 1), ((u'doc1', (u'jumped', u'fox', u'ran')), 1), ((u'doc1', (u'fox', u'jumped', u'fox')), 1), ((u'doc1', (u'of', u'crazy', u'fox')), 1), ((u'doc1', (u'over', u'fence', u'too')), 1), ((u'doc1', (u'fast', u'over', u'fence')), 1), ((u'doc1', (u'ran', u'fast', u'over')), 1), ((u'doc1', (u'fence', u'too', u'high')), 1), ((u'doc1', (u'fox', u'ran', u'fast')), 1), ((u'doc2', (u'crazy', u'fox', u'jumped')), 1), ((u'doc2', (u'a', u'crazy', u'fox')), 1), ((u'doc2', (u'a', u'crazy', u'fox')), 1), ((u'doc2', (u'jumped', u'over', u'the')), 1), ((u'doc2', (u'the', u'fence', u'again')), 1), ((u'doc2', (u'fox', u'jumped', u'over')), 1), ((u'doc2', (u'crazy', u'fox', u'jumped')), 1), ((u'doc2', (u'over', u'the', u'fence')), 1), ((u'doc3', (u'fast', u'over', u'fence')), 1), ((u'doc3', (u'ran', u'fast', u'over')), 1), ((u'doc3', (u'fox', u'ran', u'fast')), 1), ((u'doc1', (u'high', u'on', u'sugar')), 1), ((u'doc1', (u'fox', u'is', u'high')), 1), ((u'doc1', (u'jumped', u'but', u'failed')), 1), ((u'doc1', (u'is', u'high', u'on')), 1), ((u'doc1', (u'sugar', u'but', u'jumped')), 1), ((u'doc1', (u'on', u'sugar', u'but')), 1), ((u'doc1', (u'but', u'jumped', u'but')), 1), ((u'doc2', (u'crazy', u'fox', u'ran')), 1), ((u'doc2', (u'ran', u'ran', u'ran')), 1), ((u'doc2', (u'a', u'crazy', u'fox')), 1), ((u'doc2', (u'ran', u'ran', u'fast')), 1), ((u'doc2', (u'fox', u'ran', u'ran')), 1), ((u'doc3', (u'jumped', u'jumped', u'jumped')), 1), ((u'doc3', (u'crazy', u'fox', u'jumped')), 1), ((u'doc3', (u'a', u'crazy', u'fox')), 1), ((u'doc3', (u'fox', u'jumped', u'jumped')), 1), ((u'doc4', (u'fox', u'ran', u'jumped')), 1), ((u'doc4', (u'crazy', u'fox', u'ran')), 1), ((u'doc4', (u'ran', u'jumped', u'ran')), 1), ((u'doc4', (u'a', u'crazy', u'fox')), 1), ((u'doc4', (u'jumped', u'ran', u'jumped')), 1), ((u'doc4', (u'crazy', u'fox', u'jumped')), 1), ((u'doc4', (u'a', u'crazy', u'fox')), 1), ((u'doc4', (u'fox', u'jumped', u'jumped')), 1), ((u'doc4', (u'crazy', u'fox', u'jumped')), 1), ((u'doc4', (u'a', u'crazy', u'fox')), 1), ((u'doc4', (u'fox', u'jumped', u'jumped')), 1), ((u'doc5', (u'a', u'crazy', u'fox')), 1), ((u'doc5', (u'fence', u'very', u'high')), 1), ((u'doc5', (u'over', u'fence', u'very')), 1), ((u'doc5', (u'jumped', u'over', u'fence')), 1), ((u'doc5', (u'fox', u'jumped', u'over')), 1), ((u'doc5', (u'crazy', u'fox', u'jumped')), 1), ((u'doc5', (u'a', u'crazy', u'fox')), 1), ((u'doc5', (u'jumped', u'over', u'the')), 1), ((u'doc5', (u'the', u'fence', u'again')), 1), ((u'doc5', (u'fox', u'jumped', u'over')), 1), ((u'doc5', (u'crazy', u'fox', u'jumped')), 1), ((u'doc5', (u'over', u'the', u'fence')), 1), ((u'doc6', (u'fence', u'jumped', u'but')), 1), ((u'doc6', (u'about', u'fox', u'and')), 1), ((u'doc6', (u'reading', u'about', u'fox')), 1), ((u'doc6', (u'fox', u'and', u'fence')), 1), ((u'doc6', (u'jumped', u'but', u'failed')), 1), ((u'doc6', (u'and', u'fence', u'jumped')), 1), ((u'doc6', (u'book', u'reading', u'about')), 1), ((u'doc1', (u'fox', u'jumped', u'but')), 1), ((u'doc1', (u'crazy', u'fox', u'jumped')), 1), ((u'doc1', (u'jumped', u'but', u'failed')), 1)]  
  
  
**r13 = r12.reduceByKey(lambda x,y:(x+y))  
>>> r13.collect()**[((u'doc3', (u'crazy', u'fox', u'jumped')), 1), ((u'doc6', (u'fence', u'too', u'high')), 1), ((u'doc5', (u'the', u'fence', u'again')), 1), ((u'doc4', (u'ran', u'jumped', u'ran')), 1), ((u'doc6', (u'of', u'crazy', u'fox')), 1), ((u'doc1', (u'fast', u'over', u'fence')), 1), ((u'doc1', (u'over', u'the', u'fence')), 1), ((u'doc1', (u'sugar', u'but', u'jumped')), 1), ((u'doc6', (u'book', u'reading', u'about')), 1), ((u'doc4', (u'crazy', u'fox', u'jumped')), 2), ((u'doc1', (u'fox', u'jumped', u'over')), 1), ((u'doc2', (u'ran', u'ran', u'fast')), 1), ((u'doc1', (u'ran', u'fast', u'over')), 1), ((u'doc5', (u'a', u'crazy', u'fox')), 2), ((u'doc6', (u'fast', u'over', u'fence')), 1), ((u'doc5', (u'crazy', u'fox', u'jumped')), 2), ((u'doc1', (u'jumped', u'over', u'the')), 1), ((u'doc3', (u'fox', u'jumped', u'jumped')), 1), ((u'doc6', (u'jumped', u'fox', u'ran')), 1), ((u'doc4', (u'a', u'crazy', u'fox')), 3), ((u'doc1', (u'fox', u'is', u'high')), 1), ((u'doc6', (u'fox', u'ran', u'fast')), 1), ((u'doc2', (u'a', u'crazy', u'fox')), 3), ((u'doc4', (u'fox', u'ran', u'jumped')), 1), ((u'doc2', (u'fox', u'ran', u'ran')), 1), ((u'doc2', (u'the', u'fence', u'again')), 1), ((u'doc6', (u'fox', u'and', u'fence')), 1), ((u'doc5', (u'over', u'the', u'fence')), 1), ((u'doc6', (u'jumped', u'but', u'failed')), 1), ((u'doc6', (u'ran', u'fast', u'over')), 1), ((u'doc1', (u'fox', u'ran', u'fast')), 1), ((u'doc1', (u'fox', u'jumped', u'fox')), 1), ((u'doc6', (u'reading', u'about', u'fox')), 1), ((u'doc2', (u'jumped', u'over', u'the')), 1), ((u'doc5', (u'jumped', u'over', u'fence')), 1), ((u'doc3', (u'ran', u'fast', u'over')), 1), ((u'doc4', (u'jumped', u'ran', u'jumped')), 1), ((u'doc6', (u'fence', u'jumped', u'but')), 1), ((u'doc4', (u'fox', u'jumped', u'jumped')), 2), ((u'doc1', (u'jumped', u'but', u'failed')), 2), ((u'doc1', (u'crazy', u'fox', u'jumped')), 2), ((u'doc1', (u'fence', u'too', u'high')), 1), ((u'doc5', (u'fence', u'very', u'high')), 1), ((u'doc1', (u'over', u'fence', u'too')), 1), ((u'doc6', (u'fox', u'jumped', u'fox')), 1), ((u'doc2', (u'fox', u'jumped', u'over')), 1), ((u'doc5', (u'jumped', u'over', u'the')), 1), ((u'doc2', (u'crazy', u'fox', u'jumped')), 2), ((u'doc1', (u'of', u'crazy', u'fox')), 1), ((u'doc1', (u'high', u'on', u'sugar')), 1), ((u'doc3', (u'fox', u'ran', u'fast')), 1), ((u'doc2', (u'over', u'the', u'fence')), 1), ((u'doc1', (u'is', u'high', u'on')), 1), ((u'doc1', (u'but', u'jumped', u'but')), 1), ((u'doc3', (u'jumped', u'jumped', u'jumped')), 1), ((u'doc2', (u'crazy', u'fox', u'ran')), 1), ((u'doc6', (u'crazy', u'fox', u'jumped')), 1), ((u'doc1', (u'jumped', u'fox', u'ran')), 1), ((u'doc3', (u'fast', u'over', u'fence')), 1), ((u'doc6', (u'over', u'fence', u'too')), 1), ((u'doc5', (u'fox', u'jumped', u'over')), 2), ((u'doc4', (u'crazy', u'fox', u'ran')), 1), ((u'doc6', (u'and', u'fence', u'jumped')), 1), ((u'doc1', (u'on', u'sugar', u'but')), 1), ((u'doc1', (u'fox', u'jumped', u'but')), 1), ((u'doc3', (u'a', u'crazy', u'fox')), 1), ((u'doc2', (u'ran', u'ran', u'ran')), 1), ((u'doc6', (u'about', u'fox', u'and')), 1), ((u'doc5', (u'over', u'fence', u'very')), 1)]  
>>>   
  
  
**r14 = r13.map(lambda x:(x[0][0], (x[0][1],x[1])))  
>>> r14.collect()**[(u'doc3', ((u'crazy', u'fox', u'jumped'), 1)), (u'doc6', ((u'fence', u'too', u'high'), 1)), (u'doc5', ((u'the', u'fence', u'again'), 1)), (u'doc4', ((u'ran', u'jumped', u'ran'), 1)), (u'doc6', ((u'of', u'crazy', u'fox'), 1)), (u'doc1', ((u'fast', u'over', u'fence'), 1)), (u'doc1', ((u'over', u'the', u'fence'), 1)), (u'doc1', ((u'sugar', u'but', u'jumped'), 1)), (u'doc6', ((u'book', u'reading', u'about'), 1)), (u'doc4', ((u'crazy', u'fox', u'jumped'), 2)), (u'doc1', ((u'fox', u'jumped', u'over'), 1)), (u'doc2', ((u'ran', u'ran', u'fast'), 1)), (u'doc1', ((u'ran', u'fast', u'over'), 1)), (u'doc5', ((u'a', u'crazy', u'fox'), 2)), (u'doc6', ((u'fast', u'over', u'fence'), 1)), (u'doc5', ((u'crazy', u'fox', u'jumped'), 2)), (u'doc1', ((u'jumped', u'over', u'the'), 1)), (u'doc3', ((u'fox', u'jumped', u'jumped'), 1)), (u'doc6', ((u'jumped', u'fox', u'ran'), 1)), (u'doc4', ((u'a', u'crazy', u'fox'), 3)), (u'doc1', ((u'fox', u'is', u'high'), 1)), (u'doc6', ((u'fox', u'ran', u'fast'), 1)), (u'doc2', ((u'a', u'crazy', u'fox'), 3)), (u'doc4', ((u'fox', u'ran', u'jumped'), 1)), (u'doc2', ((u'fox', u'ran', u'ran'), 1)), (u'doc2', ((u'the', u'fence', u'again'), 1)), (u'doc6', ((u'fox', u'and', u'fence'), 1)), (u'doc5', ((u'over', u'the', u'fence'), 1)), (u'doc6', ((u'jumped', u'but', u'failed'), 1)), (u'doc6', ((u'ran', u'fast', u'over'), 1)), (u'doc1', ((u'fox', u'ran', u'fast'), 1)), (u'doc1', ((u'fox', u'jumped', u'fox'), 1)), (u'doc6', ((u'reading', u'about', u'fox'), 1)), (u'doc2', ((u'jumped', u'over', u'the'), 1)), (u'doc5', ((u'jumped', u'over', u'fence'), 1)), (u'doc3', ((u'ran', u'fast', u'over'), 1)), (u'doc4', ((u'jumped', u'ran', u'jumped'), 1)), (u'doc6', ((u'fence', u'jumped', u'but'), 1)), (u'doc4', ((u'fox', u'jumped', u'jumped'), 2)), (u'doc1', ((u'jumped', u'but', u'failed'), 2)), (u'doc1', ((u'crazy', u'fox', u'jumped'), 2)), (u'doc1', ((u'fence', u'too', u'high'), 1)), (u'doc5', ((u'fence', u'very', u'high'), 1)), (u'doc1', ((u'over', u'fence', u'too'), 1)), (u'doc6', ((u'fox', u'jumped', u'fox'), 1)), (u'doc2', ((u'fox', u'jumped', u'over'), 1)), (u'doc5', ((u'jumped', u'over', u'the'), 1)), (u'doc2', ((u'crazy', u'fox', u'jumped'), 2)), (u'doc1', ((u'of', u'crazy', u'fox'), 1)), (u'doc1', ((u'high', u'on', u'sugar'), 1)), (u'doc3', ((u'fox', u'ran', u'fast'), 1)), (u'doc2', ((u'over', u'the', u'fence'), 1)), (u'doc1', ((u'is', u'high', u'on'), 1)), (u'doc1', ((u'but', u'jumped', u'but'), 1)), (u'doc3', ((u'jumped', u'jumped', u'jumped'), 1)), (u'doc2', ((u'crazy', u'fox', u'ran'), 1)), (u'doc6', ((u'crazy', u'fox', u'jumped'), 1)), (u'doc1', ((u'jumped', u'fox', u'ran'), 1)), (u'doc3', ((u'fast', u'over', u'fence'), 1)), (u'doc6', ((u'over', u'fence', u'too'), 1)), (u'doc5', ((u'fox', u'jumped', u'over'), 2)), (u'doc4', ((u'crazy', u'fox', u'ran'), 1)), (u'doc6', ((u'and', u'fence', u'jumped'), 1)), (u'doc1', ((u'on', u'sugar', u'but'), 1)), (u'doc1', ((u'fox', u'jumped', u'but'), 1)), (u'doc3', ((u'a', u'crazy', u'fox'), 1)), (u'doc2', ((u'ran', u'ran', u'ran'), 1)), (u'doc6', ((u'about', u'fox', u'and'), 1)), (u'doc5', ((u'over', u'fence', u'very'), 1))]  
>>>   
  
  
  
**>>> r15 = r14.cartesian(r14)  
>>> r15.collect()**Output too huge!!!!!

**r16 = r15.filter(lambda x: x[0][0] < x[1][0])  
>>> r16.collect()**Output too huge!!!!!

**r17 = r16.filter(lambda x: x[0][1][0] == x[1][1][0])  
>>> r17.collect()**[((u'doc3', ((u'crazy', u'fox', u'jumped'), 1)), (u'doc4', ((u'crazy', u'fox', u'jumped'), 2))), ((u'doc3', ((u'crazy', u'fox', u'jumped'), 1)), (u'doc5', ((u'crazy', u'fox', u'jumped'), 2))), ((u'doc3', ((u'crazy', u'fox', u'jumped'), 1)), (u'doc6', ((u'crazy', u'fox', u'jumped'), 1))), ((u'doc1', ((u'fast', u'over', u'fence'), 1)), (u'doc6', ((u'fast', u'over', u'fence'), 1))), ((u'doc1', ((u'over', u'the', u'fence'), 1)), (u'doc5', ((u'over', u'the', u'fence'), 1))), ((u'doc1', ((u'fast', u'over', u'fence'), 1)), (u'doc3', ((u'fast', u'over', u'fence'), 1))), ((u'doc1', ((u'over', u'the', u'fence'), 1)), (u'doc2', ((u'over', u'the', u'fence'), 1))), ((u'doc4', ((u'crazy', u'fox', u'jumped'), 2)), (u'doc5', ((u'crazy', u'fox', u'jumped'), 2))), ((u'doc1', ((u'ran', u'fast', u'over'), 1)), (u'doc6', ((u'ran', u'fast', u'over'), 1))), ((u'doc4', ((u'crazy', u'fox', u'jumped'), 2)), (u'doc6', ((u'crazy', u'fox', u'jumped'), 1))), ((u'doc1', ((u'fox', u'jumped', u'over'), 1)), (u'doc2', ((u'fox', u'jumped', u'over'), 1))), ((u'doc1', ((u'fox', u'jumped', u'over'), 1)), (u'doc5', ((u'fox', u'jumped', u'over'), 2))), ((u'doc1', ((u'ran', u'fast', u'over'), 1)), (u'doc3', ((u'ran', u'fast', u'over'), 1))), ((u'doc2', ((u'the', u'fence', u'again'), 1)), (u'doc5', ((u'the', u'fence', u'again'), 1))), ((u'doc4', ((u'a', u'crazy', u'fox'), 3)), (u'doc5', ((u'a', u'crazy', u'fox'), 2))), ((u'doc2', ((u'a', u'crazy', u'fox'), 3)), (u'doc5', ((u'a', u'crazy', u'fox'), 2))), ((u'doc2', ((u'a', u'crazy', u'fox'), 3)), (u'doc4', ((u'a', u'crazy', u'fox'), 3))), ((u'doc1', ((u'fox', u'ran', u'fast'), 1)), (u'doc6', ((u'fox', u'ran', u'fast'), 1))), ((u'doc5', ((u'crazy', u'fox', u'jumped'), 2)), (u'doc6', ((u'crazy', u'fox', u'jumped'), 1))), ((u'doc1', ((u'jumped', u'over', u'the'), 1)), (u'doc2', ((u'jumped', u'over', u'the'), 1))), ((u'doc1', ((u'jumped', u'over', u'the'), 1)), (u'doc5', ((u'jumped', u'over', u'the'), 1))), ((u'doc3', ((u'fox', u'jumped', u'jumped'), 1)), (u'doc4', ((u'fox', u'jumped', u'jumped'), 2))), ((u'doc1', ((u'fox', u'ran', u'fast'), 1)), (u'doc3', ((u'fox', u'ran', u'fast'), 1))), ((u'doc2', ((u'a', u'crazy', u'fox'), 3)), (u'doc3', ((u'a', u'crazy', u'fox'), 1))), ((u'doc1', ((u'crazy', u'fox', u'jumped'), 2)), (u'doc3', ((u'crazy', u'fox', u'jumped'), 1))), ((u'doc2', ((u'crazy', u'fox', u'jumped'), 2)), (u'doc3', ((u'crazy', u'fox', u'jumped'), 1))), ((u'doc1', ((u'fence', u'too', u'high'), 1)), (u'doc6', ((u'fence', u'too', u'high'), 1))), ((u'doc1', ((u'of', u'crazy', u'fox'), 1)), (u'doc6', ((u'of', u'crazy', u'fox'), 1))), ((u'doc1', ((u'crazy', u'fox', u'jumped'), 2)), (u'doc4', ((u'crazy', u'fox', u'jumped'), 2))), ((u'doc2', ((u'crazy', u'fox', u'jumped'), 2)), (u'doc4', ((u'crazy', u'fox', u'jumped'), 2))), ((u'doc3', ((u'fast', u'over', u'fence'), 1)), (u'doc6', ((u'fast', u'over', u'fence'), 1))), ((u'doc3', ((u'ran', u'fast', u'over'), 1)), (u'doc6', ((u'ran', u'fast', u'over'), 1))), ((u'doc1', ((u'jumped', u'but', u'failed'), 2)), (u'doc6', ((u'jumped', u'but', u'failed'), 1))), ((u'doc1', ((u'crazy', u'fox', u'jumped'), 2)), (u'doc5', ((u'crazy', u'fox', u'jumped'), 2))), ((u'doc2', ((u'crazy', u'fox', u'jumped'), 2)), (u'doc5', ((u'crazy', u'fox', u'jumped'), 2))), ((u'doc3', ((u'fox', u'ran', u'fast'), 1)), (u'doc6', ((u'fox', u'ran', u'fast'), 1))), ((u'doc2', ((u'over', u'the', u'fence'), 1)), (u'doc5', ((u'over', u'the', u'fence'), 1))), ((u'doc1', ((u'jumped', u'fox', u'ran'), 1)), (u'doc6', ((u'jumped', u'fox', u'ran'), 1))), ((u'doc1', ((u'fox', u'jumped', u'fox'), 1)), (u'doc6', ((u'fox', u'jumped', u'fox'), 1))), ((u'doc2', ((u'jumped', u'over', u'the'), 1)), (u'doc5', ((u'jumped', u'over', u'the'), 1))), ((u'doc1', ((u'crazy', u'fox', u'jumped'), 2)), (u'doc2', ((u'crazy', u'fox', u'jumped'), 2))), ((u'doc1', ((u'crazy', u'fox', u'jumped'), 2)), (u'doc6', ((u'crazy', u'fox', u'jumped'), 1))), ((u'doc1', ((u'over', u'fence', u'too'), 1)), (u'doc6', ((u'over', u'fence', u'too'), 1))), ((u'doc2', ((u'fox', u'jumped', u'over'), 1)), (u'doc5', ((u'fox', u'jumped', u'over'), 2))), ((u'doc2', ((u'crazy', u'fox', u'jumped'), 2)), (u'doc6', ((u'crazy', u'fox', u'jumped'), 1))), ((u'doc2', ((u'crazy', u'fox', u'ran'), 1)), (u'doc4', ((u'crazy', u'fox', u'ran'), 1))), ((u'doc3', ((u'a', u'crazy', u'fox'), 1)), (u'doc5', ((u'a', u'crazy', u'fox'), 2))), ((u'doc3', ((u'a', u'crazy', u'fox'), 1)), (u'doc4', ((u'a', u'crazy', u'fox'), 3)))]  
>>>   
  
  
**r18 = r17.map(lambda x:((x[0][0],x[1][0]),(x[0][1][1],x[1][1][1])))  
>>> r18.collect()**[((u'doc3', u'doc4'), (1, 2)), ((u'doc3', u'doc5'), (1, 2)), ((u'doc3', u'doc6'), (1, 1)), ((u'doc1', u'doc6'), (1, 1)), ((u'doc1', u'doc5'), (1, 1)), ((u'doc1', u'doc3'), (1, 1)), ((u'doc1', u'doc2'), (1, 1)), ((u'doc4', u'doc5'), (2, 2)), ((u'doc1', u'doc6'), (1, 1)), ((u'doc4', u'doc6'), (2, 1)), ((u'doc1', u'doc2'), (1, 1)), ((u'doc1', u'doc5'), (1, 2)), ((u'doc1', u'doc3'), (1, 1)), ((u'doc2', u'doc5'), (1, 1)), ((u'doc4', u'doc5'), (3, 2)), ((u'doc2', u'doc5'), (3, 2)), ((u'doc2', u'doc4'), (3, 3)), ((u'doc1', u'doc6'), (1, 1)), ((u'doc5', u'doc6'), (2, 1)), ((u'doc1', u'doc2'), (1, 1)), ((u'doc1', u'doc5'), (1, 1)), ((u'doc3', u'doc4'), (1, 2)), ((u'doc1', u'doc3'), (1, 1)), ((u'doc2', u'doc3'), (3, 1)), ((u'doc1', u'doc3'), (2, 1)), ((u'doc2', u'doc3'), (2, 1)), ((u'doc1', u'doc6'), (1, 1)), ((u'doc1', u'doc6'), (1, 1)), ((u'doc1', u'doc4'), (2, 2)), ((u'doc2', u'doc4'), (2, 2)), ((u'doc3', u'doc6'), (1, 1)), ((u'doc3', u'doc6'), (1, 1)), ((u'doc1', u'doc6'), (2, 1)), ((u'doc1', u'doc5'), (2, 2)), ((u'doc2', u'doc5'), (2, 2)), ((u'doc3', u'doc6'), (1, 1)), ((u'doc2', u'doc5'), (1, 1)), ((u'doc1', u'doc6'), (1, 1)), ((u'doc1', u'doc6'), (1, 1)), ((u'doc2', u'doc5'), (1, 1)), ((u'doc1', u'doc2'), (2, 2)), ((u'doc1', u'doc6'), (2, 1)), ((u'doc1', u'doc6'), (1, 1)), ((u'doc2', u'doc5'), (1, 2)), ((u'doc2', u'doc6'), (2, 1)), ((u'doc2', u'doc4'), (1, 1)), ((u'doc3', u'doc5'), (1, 2)), ((u'doc3', u'doc4'), (1, 3))]  
>>>   
  
  
**>> r19 = r18.map(lambda x:(x[0], x[1][0] if(x[1][0]<x[1][1]) else x[1][1]))  
>>> r19.collect()**[((u'doc3', u'doc4'), 1), ((u'doc3', u'doc5'), 1), ((u'doc3', u'doc6'), 1), ((u'doc1', u'doc6'), 1), ((u'doc1', u'doc5'), 1), ((u'doc1', u'doc3'), 1), ((u'doc1', u'doc2'), 1), ((u'doc4', u'doc5'), 2), ((u'doc1', u'doc6'), 1), ((u'doc4', u'doc6'), 1), ((u'doc1', u'doc2'), 1), ((u'doc1', u'doc5'), 1), ((u'doc1', u'doc3'), 1), ((u'doc2', u'doc5'), 1), ((u'doc4', u'doc5'), 2), ((u'doc2', u'doc5'), 2), ((u'doc2', u'doc4'), 3), ((u'doc1', u'doc6'), 1), ((u'doc5', u'doc6'), 1), ((u'doc1', u'doc2'), 1), ((u'doc1', u'doc5'), 1), ((u'doc3', u'doc4'), 1), ((u'doc1', u'doc3'), 1), ((u'doc2', u'doc3'), 1), ((u'doc1', u'doc3'), 1), ((u'doc2', u'doc3'), 1), ((u'doc1', u'doc6'), 1), ((u'doc1', u'doc6'), 1), ((u'doc1', u'doc4'), 2), ((u'doc2', u'doc4'), 2), ((u'doc3', u'doc6'), 1), ((u'doc3', u'doc6'), 1), ((u'doc1', u'doc6'), 1), ((u'doc1', u'doc5'), 2), ((u'doc2', u'doc5'), 2), ((u'doc3', u'doc6'), 1), ((u'doc2', u'doc5'), 1), ((u'doc1', u'doc6'), 1), ((u'doc1', u'doc6'), 1), ((u'doc2', u'doc5'), 1), ((u'doc1', u'doc2'), 2), ((u'doc1', u'doc6'), 1), ((u'doc1', u'doc6'), 1), ((u'doc2', u'doc5'), 1), ((u'doc2', u'doc6'), 1), ((u'doc2', u'doc4'), 1), ((u'doc3', u'doc5'), 1), ((u'doc3', u'doc4'), 1)]  
>>>   
  
  
**r20 = r19.reduceByKey(lambda x,y:(x+y))  
>>> r20.collect()**[((u'doc5', u'doc6'), 1), ((u'doc2', u'doc4'), 6), ((u'doc3', u'doc4'), 3), ((u'doc3', u'doc5'), 2), ((u'doc1', u'doc2'), 5), ((u'doc1', u'doc3'), 4), ((u'doc2', u'doc3'), 2), ((u'doc4', u'doc6'), 1), ((u'doc4', u'doc5'), 4), ((u'doc3', u'doc6'), 4), ((u'doc1', u'doc4'), 2), ((u'doc1', u'doc5'), 5), ((u'doc2', u'doc6'), 1), ((u'doc1', u'doc6'), 10), ((u'doc2', u'doc5'), 8)]  
>>>   
  
**rFinal = r20.join(r10)  
>>> rFinal.collect()**[((u'doc1', u'doc2'), (5, 35)), ((u'doc2', u'doc4'), (6, 24)), ((u'doc1', u'doc3'), (4, 29)), ((u'doc4', u'doc6'), (1, 27)), ((u'doc5', u'doc6'), (1, 28)), ((u'doc3', u'doc4'), (3, 18)), ((u'doc3', u'doc5'), (2, 19)), ((u'doc2', u'doc3'), (2, 20)), ((u'doc1', u'doc4'), (2, 33)), ((u'doc1', u'doc6'), (10, 38)), ((u'doc3', u'doc6'), (4, 23)), ((u'doc1', u'doc5'), (5, 34)), ((u'doc4', u'doc5'), (4, 23)), ((u'doc2', u'doc5'), (8, 25)), ((u'doc2', u'doc6'), (1, 29))]  
>>>   
  
  
**rFinalV1 = rFinal.map(lambda x:(x[0],(x[1][0],(x[1][1]-x[1][0]))))  
>>> rFinalV1.collect()**[((u'doc1', u'doc2'), (5, 30)), ((u'doc2', u'doc4'), (6, 18)), ((u'doc1', u'doc3'), (4, 25)), ((u'doc4', u'doc6'), (1, 26)), ((u'doc5', u'doc6'), (1, 27)), ((u'doc3', u'doc4'), (3, 15)), ((u'doc3', u'doc5'), (2, 17)), ((u'doc2', u'doc3'), (2, 18)), ((u'doc1', u'doc4'), (2, 31)), ((u'doc1', u'doc6'), (10, 28)), ((u'doc3', u'doc6'), (4, 19)), ((u'doc1', u'doc5'), (5, 29)), ((u'doc4', u'doc5'), (4, 19)), ((u'doc2', u'doc5'), (8, 17)), ((u'doc2', u'doc6'), (1, 28))]  
>>>   
  
  
**rSuperFinal = rFinalV1.map(lambda x: (x[0],(x[1][0]\*100/x[1][1])))  
>>> rSuperFinal.collect()**[((u'doc1', u'doc2'), 16), ((u'doc2', u'doc4'), 33), ((u'doc1', u'doc3'), 16), ((u'doc4', u'doc6'), 3), ((u'doc5', u'doc6'), 3), ((u'doc3', u'doc4'), 20), ((u'doc3', u'doc5'), 11), ((u'doc2', u'doc3'), 11), ((u'doc1', u'doc4'), 6), ((u'doc1', u'doc6'), 35), ((u'doc3', u'doc6'), 21), ((u'doc1', u'doc5'), 17), ((u'doc4', u'doc5'), 21), ((u'doc2', u'doc5'), 47), ((u'doc2', u'doc6'), 3)]  
>>> 