

Solution for Bonus Question --- Find Coocur

Ans-

```
def find_coocur(w1, w2, dtm, words, paragraphs):

    result = []
    # add your code here
    # find location of the words having both words
    w1_index = np.where(words == w1)[0]
    w2_index = np.where(words == w2)[0]

    #get the dtm matrix of both locations
    w1_dtm = dtm[:,w1_index]
    w2_dtm = dtm[:,w2_index]

    #get arrays of satisfying conditions
    para_position = np.where((w1_dtm!=0) & (w2_dtm!=0))[0]

    #get the paragraphs location with matching para_position wher eboth words occur
    result = np.array(paragraphs)[para_position]
    return result
```

Explanation

1. Input params - w1, w2, dtm, words, paragraphs
W1 & w2 are the words that we need to find to coocur in multiple paragraphs
2. For this we need to find the position of the words to search in the paragraphs so we can use to identify their occurrence from the dtm matrix.
3. We get location of w1 & w2 using np.where which returns the location of the entry element in the nparray.
4. We get the np vector for each word position for all the paragraphs from the dtm matrix. W1_dtm for w1 dtm matrix per paragraph and w2_dtm for w2 dtm matrix per paragraph.
5. Again, we can use the np where to check if either of the dtm matrixes are not equal to 0 and get the result as an array list to match the position of paragraphs in the document.
6. Since paragraphs is a list, it does not allow us to retrieve positional values in one sentence, hence we have to convert or typecast it to an np.array and then retrieve the positional values from the paragraphs list.