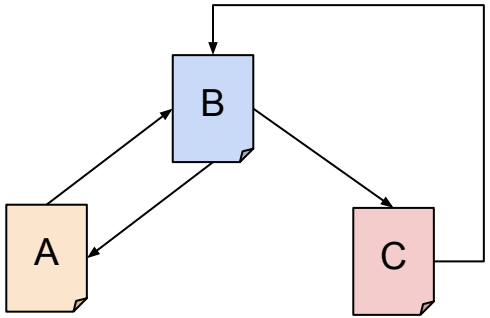
Homework 1

A search query “Word1 Word2” is being scored against the following documents (see table). The document corpus (as shown in the table) only contains five documents. The number of times the words “Word1” and “Word2” appear in each of the documents is given in the table. The length of each document is also given. Assume k = 1.2 and b = 0.75. Write a python program to calculate the BM25 score for the query against all the documents and rank the documents by their BM25 score. Your program should read the table from a file, i.e. do not hard code these values into your program.

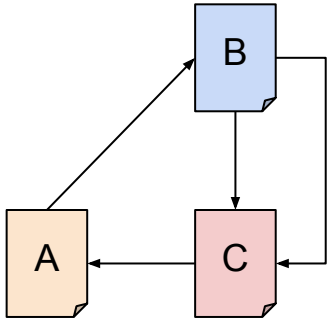
DocumentID	DocumentL	FrequencyOfWord1	FrequencyOfWord2
1	50	0	78
2	80	56	6
3	20	45	89
4	200	89	23
5	10	76	0

Homework 2

Given the number of web pages N = 3, and the damping parameter d = 0.7. For the two networks shown below, calculate the PageRank of the pages A, B, and C. Links between the pages are shown in the graph itself.



(a)



(b)