Name _____ Roll No _____ _____
Section

National University of Computer and Emerging Sciences, Lahore Campus

Course:	Information Retrieval	Course Code:	CS317
Program:	BS(Computer Science)	Semester:	Fall 2018
Duration:	60 Minutes	Total Marks:	17
Paper Date:	15-Nov-18	Weight	12%
Section:	ALL	Page(s):	4
Exam:	Midterm-2	Roll No:	

Instruction/Notes: ***Attempt the examination on this question paper.***. *You can use extra sheets for rough work but do not attach extra sheets with this paper. Do not fill the table titled Question/marks*

Question	1	2	3	Total
Marks	/ 4	/ 11	/ 4	**/17**

Q1) Suppose a word w has occurred 10 times in a document D with 100 words. Assume that the probability of the word according to the collection (background) language model, p(w|C) is 0.01, and the Dirichlet prior smoothing parameter (mu) is 300. What is the estimated probability of this word in the document language model p(w|D) if we use Dirichlet prior smoothing? If we increase the smoothing parameter, would the estimated probability p(w|D) become larger or smaller? Why? [4 Marks]

Solution:
(100/400)*(10/100) + (300/400)*(0.01)
=0.03

Smaller, because we are giving more weight to background model.

Q2) (a) Suppose the relevance status of the top-8 ranked results from a system is

R R R N N N N R.

Suppose there are in total 10 relevant documents in the collection. Compute the following evaluation measures for this result:
(1) Precision at 5 documents. [1 Mark]
(2) Average Precision. [2 Marks]

Solution:

P@5 = 3/5
AP = 1 + 1 + 1 +

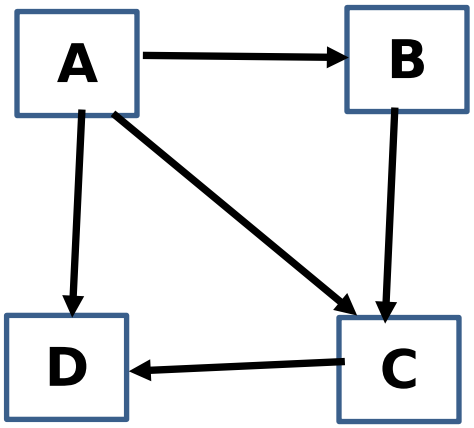
Q2) (b) Compute NDCG@10 of following rank list of documents.

Name _____ Roll No _____ _____
Section

3, 0, 1, 0, 0, 2, 0, 0, 2, 0

Grades of relevant documents for this query in collections are 3, 3, 2, 2, 2, 1
Clearly write the DCG formula you have used. [4 Marks]

Q3) (a) Compute page rank of all nodes of following graph. Damping factor d = 0.8. Perform only one iteration of page rank algorithm. [2 Marks]



Solution:

A = 0.2/4 + 0.8*(0) + 0.8*(0.25/4) = 0.1
B = 0.2/4 + 0.8*(0.25/3) + 0.8*(0.25/4) = 0.167
C = 0.2/4 + 0.8*(0.25/3 + 0.25) + 0.8*(0.25/4) = 0.37
D = 0.2/4 + 0.8*(0.25/3 + 0.25) + 0.8*(0.25/4) 0.37

A = 0.2/4 + 0.8*(0) + 0.8*(0.37/4) = 0.124
B = 0.2/4 + 0.8*(0.1/3) + 0.8*(0.37/4) = 0.151
C = 0.2/4 + 0.8*(0.1/3 + 0.167) + 0.8*(0.37/4) = 0.284
D = 0.2/4 + 0.8*(0.1/3 + 0.37) + 0.8*(0.37/4) = 0.446

A = 0.2/4 + 0.8*(0) + 0.8*(0.446/4) = 0.139
B = 0.2/4 + 0.8*(0.124/3) + 0.8*(0.446/4) = 0.172
C = 0.2/4 + 0.8*(0.124/3 + 0.15) + 0.8*(0.446/4) = 0.29
D = 0.2/4 + 0.8*(0.124/3 + 0.284) + 0.8*(0.446/4) = 0.399

A = 0.2/4 + 0.8*(0) + 0.8*(0.4/4) = 0.13
B = 0.2/4 + 0.8*(0.14/3) + 0.8*(0.4/4) = 0.167
C = 0.2/4 + 0.8*(0.14/3 + 0.172) + 0.8*(0.4/4) = 0.31

D = 0.2/4 + 0.8*(0.14/3 + 0.29) + 0.8*(0.4/4) = 0.399

Q3) (b) Suppose that P, Q, and R are different web pages. Explain how it can happen that adding a link from P to Q can raise the PageRank of R. Explain how it can happen that adding a link from P to Q can lower the PageRank of R. In both cases, you should show a specific graph where this happens, though you need not work out the actual numerical values. [3 Marks]

Solution:

First case: Initial graph: P Q --> R Adding a link from P to Q raises the PageRank of Q and thus indirectly the PageRank of R. First case: Initial graph: P ---> R Q If you add a link from P to Q, then P's "contribution of importance" is divided between Q and R rather than going exclusively to R, so the PageRank of R decreases.

Q4) Compare language modeling with vector space model. Give some similarities and differences. [2 Marks]

Solution:

Similar in some ways
–Term weights based on frequency

–Terms often used as if they were independent

–Inverse document/collection frequency used

–Some form of length normalization useful

•Different in others
–Based on probability rather than similarity
•Intuitions are probabilistic rather than geometric
–Details of use of document length and term, document, and collection frequency differ