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 2019**
Duration:	**60 Minutes**	**Total Marks:**	**18**
Paper Date:	**7-Nov-19**	**Weight**	**13%**
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	4	Total
Marks	/ 4	/ 4	/6	/ 4	**/18**

Q1) Consider following collection of 3 documents. [4 Marks]

Document	Words
D1	a b a c b b c
D2	a b a b a
D3	b c b b b a a b b

Query = < b c >

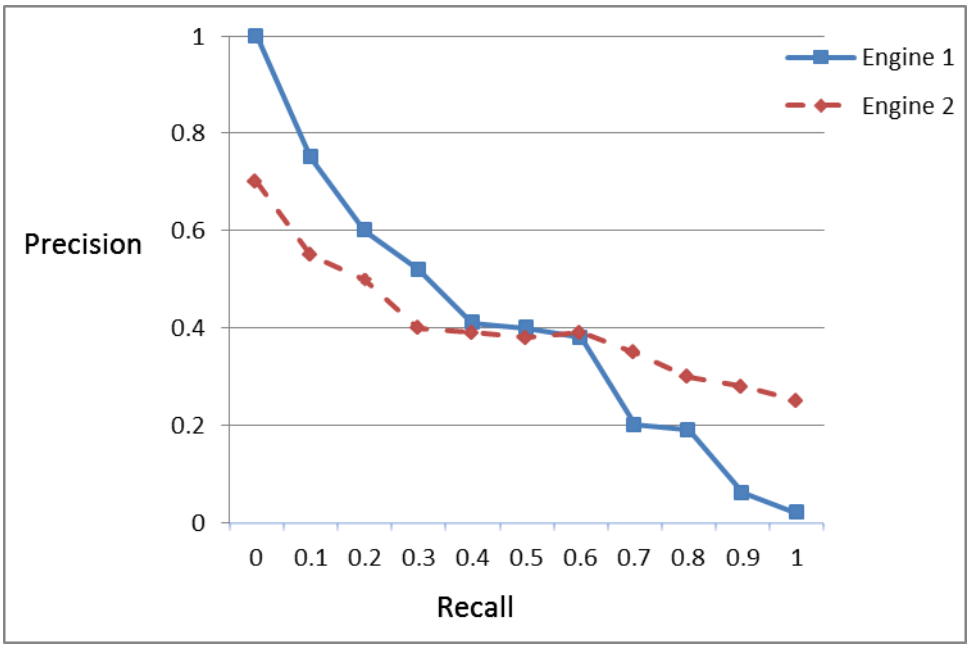
Use Dirichlet smoothing to find similarity of document D2 with query (mu = 5).

Q2) (a) Compute average precision of following list of documents for 1 query. Leftmost is top ranked document. Total relevant for query 1 = 5. [2 Marks]

NR NR R NR NR R NR R

Name _____ Roll No _____ Section _____

Q2) (b) The figure below depicts precision-recall curves for two search engines that index research articles. There is no difference between the engines except in how they score documents. Imagine you're a scientist looking for all published work on some topic. You don't want to miss any citation. Which engine would you prefer and why? [2 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. [4 Marks]

Q3 (b) Why do we use teleportation in page rank algorithm? [2 Marks]

Q4) (a) Write down similarities and differences of the following two methods of word vector representations. [2 Marks]

i. WordToVec
ii. Tf.IDF based Vector

Similarities:

Differences:

Q4) (b) The WordToVec training method discussed in class is practically inefficient. Which part of the training makes it inefficient? [1 Mark]

Q4) (c) Briefly describe some solution for the problem identified in part (a) for efficient implementation. [1 Mark]