INFORMATION RETRIEVAL – SHORT EXERCISES III – EVALUATION IN INFORMATION RETRIEVAL AND PAGERANK

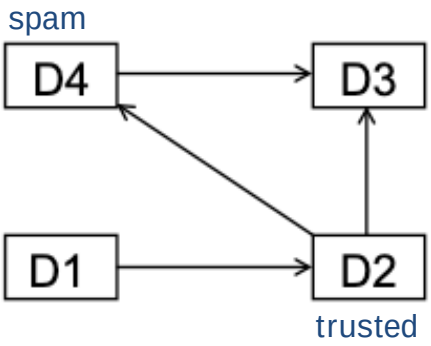
I. Consider an information need for which there are 4 relevant documents in the collection. A system run on this collection returned the top 10 results for which the relevance is judged as follows (R – relevant; N – non-relevant):

R N R N N N N N R R

What is the recall at 6 (R@6)? Answer: (1/4) * (1 + 0 + 1 + 0 + 0 + 0) = 2/4 = 1/2

What is the Mean Average Precision? Answer: (1/4) * (P@1 + P@3 + P@9 + P@10) =
= (1/4) * ((1/1) * 1 + (1/3) * 2 + (1/9) * 3 + (1/10) * 4) = 3/5

II.Consider the web graph presented below to the left. It involves four pages D1-D4 and four links. Fill in the stochastic matrix M given to the right.



	D1	D2	D3	D4	
D1	0	0	0	0	
D2	1	0	0	0	
D3	0	1	0	1	
D4	0	1	0	0	

Write the equation for PR(D3) without dumping factor *q*? Answer: PR(D3) = 0*PR(D1) + 1*PR(D2) + 0*PR(D3) + 1*PR(D4)

Which page has the greatest PageRank (without computing the exact PR values)? Answer: D3

An oracle has evaluated D2 as trusted and D4 as spam. What is the starting vector d for TrustRank?

Answer: d = [0 , 1 , 0 , 0]

(Of course, apart from d = [0, 0, 0, 0])