Martyna Stasiak id. 156071

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):

k=	1	2	3	4	5	6	7	8	9	10
	R	N	R	N	N	N	N	N	R	R
P@k	1/1	1/2	2/3	2/4	2/5	2/6	2/7	2/8	3/9	4/10

$P@k = \frac{1}{k}\sum_{i=1}^{k} r_i$

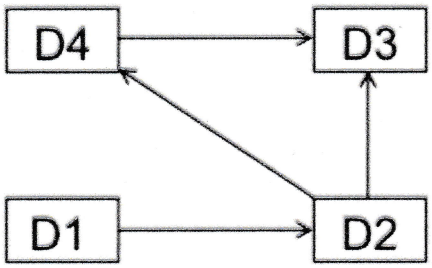
$R@k = \frac{1}{\#rel}\sum_{i=1}^{k} r_i$

$MAP(q) = \frac{1}{\#rel}\sum_{i=1}^{\#rel} P@i$

What is the recall at 6 (R@6)? Answer: 2/4

What is the Mean Average Precision? Answer: $\frac{1}{4}\left(\frac{1}{1}+\frac{2}{3}+\frac{3}{9}+\frac{4}{10}\right) = \frac{1}{4}\cdot\frac{24}{10} = \frac{6}{10} = 3/5$

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	1/2	0	1
D4	0	1/2	0	0

Write the equation for PR(D3) without dumping factor *q*? Answer: PR(D3) = $\frac{1}{2}PR(D2) + PR(D4)$

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 = [D1: 0 , D2 ✓: 1 , D3: 0 , D4 ✗: 0]