2011 Paper 8 Question 9

Information Retrieval

The figure below shows the output of two information retrieval systems on the same two queries in a competitive evaluation. The top 15 ranks are shown. Crosses correspond to a document which has been judged relevant by a human judge; dashes correspond to irrelevant documents. There are no relevant documents in lower ranks.

System 1				System 2		
Rank	Q1	Q2		Rank	Q1	Q2
1	–	X		1	X	X
2	X	–		2	X	–
3	X	–		3	X	–
4	X	–		4	–	X
5	–	–		5	X	X
6	–	–		6	X	–
7	–	–		7	–	–
8	X	–		8	–	–
9	X	–		9	–	–
10	X	–		10	–	–
11	X	–		11	X	–
12	–	–		12	X	–
13	–	X		13	–	–
14	–	X		14	–	–
15	X	–		15	X	–

(a) Explain the following evaluation metrics and give results for query Q1 for both systems.

    (i) Precision at rank 10.   [2 marks]

    (ii) Recall at precision 0.5.   [2 marks]

(b) The metrics in part (a) above are not adequate measures of system performance for arbitrary queries. Why not? What other disadvantages do these metrics have?   [3 marks]

(c) Give the formula for mean average precision (MAP), and illustrate the metric by calculating System 1's MAP.   [6 marks]

(d) For each system, draw a precision-recall curve. Explain how you arrived at your result. How could one create more informative curves?   [7 marks]