

2007 Paper 7 Question 5

Information Retrieval

- (a) Information retrieval systems are commonly compared against each other in large-scale evaluative conferences. Which difficulties do designers of such evaluations face? Describe *three* difficulties and how they could be addressed. [5 marks]
- (b) 11-pt avg. precision is an area-based evaluation measure which combines precision and recall.
- (i) Why are simpler evaluation metrics insufficient for comparative IR evaluation exercises? [3 marks]
- (ii) Give the formula for 11-pt avg. precision, defining all variables used. [2 marks]
- (iii) The following table shows the output of an information retrieval system on two queries. You can assume that there are no relevant documents in ranks lower than the top 10 ranks shown. Calculate 11-pt avg. precision for these two queries, using an appropriate interpolation method if necessary, and sketch it in a precision-recall graph. [5 marks]

Rank	Q1	Q2
1	X	–
2	–	X
3	–	–
4	X	X
5	X	–
6	–	X
7	–	X
8	–	–
9	–	X
10	–	X

Output of an IR system on two queries Q1 and Q2; top 10 ranks. Crosses correspond to relevant documents, dashes to irrelevant documents.

- (c) Large-scale evaluations of information extraction systems use a different setup from IR evaluations, and different evaluation metrics. Describe in detail how you would organise the evaluation of information extraction systems on the task that SNOWBALL solves, i.e. the determination of pairs of company names and headquarters. Which data would you prepare, and which evaluation metrics would you use? Justify your decisions. [5 marks]