

7 Information Retrieval (HY)

- (a) (i) Describe the difference between relevance feedback and query expansion in terms of user interaction. [2 marks]
- (ii) Explain what we mean by equivalence classing of terms and why it is useful. Give one example of equivalence classing that can fail to retrieve the right documents. [2 marks]
- (iii) Give an example of how asymmetric expansion of query terms can usefully model users' expectations. Is asymmetric expansion of query terms more or less efficient than equivalence classing? Justify your answer. [2 marks]
- (iv) How would you evaluate an Information Retrieval task for which there is high tolerance for overlooked relevant information items? How might you modify the *F measure* for such a task? Justify your answer. [3 marks]
- (b) (i) Given the query “**elvis music**” and the following term frequencies for the three documents *doc1*, *doc2* and *doc3*:

	elvis	presley	mississippi	pop	music	life
<i>doc1</i>	3	4	0	6	0	0
<i>doc2</i>	4	0	4	0	0	3
<i>doc3</i>	5	3	0	4	4	0

calculate the cosine similarity between the query and each document (you can ignore the *idf* term) in order to rank these documents in order of relevance. Show your workings. [2 marks]

- (ii) The Rocchio algorithm is a classic algorithm for implementing relevance feedback. Use Rocchio to compute the new query vector for “**elvis music**” using *doc3* for relevance feedback (i.e., *doc3* has been marked as relevant). Give suitable values for Rocchio's weight parameters. As above, calculate cosine similarity (you can ignore the *idf* term) in order to rank the documents in order of relevance. Show your workings. [4 marks]
- (iii) In Rocchio's algorithm, positive feedback is typically more useful than negative feedback. Give two example cases of negative documents being fed back that can decrease the retrieval effectiveness of the Rocchio re-formulated query. Propose one way in which you can incorporate negative feedback more effectively, and explain why this helps with the two examples you provided. Motivate your answer. [5 marks]