Information Retrieval

(a) Why are query terms and document terms weighted in modern information retrieval systems? Describe a method for weighting terms. [6 marks]

(b) Describe two challenges in recognising arbitrary person names in unrestricted text. [4 marks]

(c) Describe how bootstrapping can be used in information retrieval. [5 marks]

(d) Which of the following information retrieval tasks is bootstrapping best suited for, and why?

(i) Recognition of person names.

(ii) Association of a person with his or her job title.

(iii) Identification of arbitrary relationship between two people, as in the following examples:

“John did not fire Jim” →  
Relationship: Boss  
Person 1: John  
Person 2: Jim

“James and Jill visited their parents” →  
Relationship: Sibling  
Person 1: James  
Person 2: Jill

[5 marks]