2003 Paper 7 Question 11

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:

$$\text{``John did not fire Jim''} \longrightarrow \left[\begin{array}{l} \text{Relationship: Boss} \\ \text{Person 1: John} \\ \text{Person 2: Jim} \end{array} \right.$$

$$\text{``James and Jill visited their parents''} \longrightarrow \left[\begin{array}{l} \text{Relationship: Sibling} \\ \text{Person 1: James} \\ \text{Person 2: Jill} \end{array} \right.$$

[5 marks]