2002 Paper 7 Question 11

Natural Language Processing

An open-domain question-answering system should be able to answer questions such as the following by searching a database of documents and returning an appropriate text snippet. For example:

User: What debts did Quintex leave?

System 1: About $1.4 billion (Australian)

extracted from:

Quintex Australia Ltd. and Quintex Ltd. together have debt of about $1.4 billion (Australian), according to two analysts at Australian brokerage firms.

and:

System 2: Around ADollars 1.5bn (Pounds 680m)

from:

Quintex group collapsed yesterday. The failure left corporate debts of around ADollars 1.5bn (Pounds 680m) and additional personal debts.

(a) What natural language processing techniques would be needed to build such a system? [8 marks]

(b) What problems would arise as a consequence of the open-domain requirement? [4 marks]

(c) What problems would arise selecting an appropriate text snippet from a matching sentence? [4 marks]

(d) How feasible would it be to have the system reliably rank such answers (in this case preferring the second)? [4 marks]