

2006 Paper 7 Question 11

Information Retrieval

Information retrieval systems vary in the expressivity of the query languages they employ. For instance, some systems support proximity search: if two query terms are connected by the “Next” operator, then only those documents are retrieved where the query terms appear close together (i.e., within a certain number of words of each other).

- (a) List and briefly describe other ways in which the syntax and the interpretation of query languages may vary. [8 marks]
- (b) Describe with an example how the “Next” operator described above is implemented efficiently in modern information retrieval systems. Your answer should include a description of the data structure(s) necessary to support it. [5 marks]
- (c) A search engine supports error correction in the following way: If an error is suspected in a query term, the system provides a link labelled “Did you mean X?”, where X is the corrected term, in addition to its normal results. The link leads to a list of retrieved documents, corresponding to a variant of the original query, with X replacing the misspelled term.
 - (i) Explain why it is non-trivial to implement this feature efficiently. [3 marks]
 - (ii) Discuss methods for implementing this feature in a realistic setting. [4 marks]