

Summer Examinations 2004
Front Page Template

Exam Code(s) 4IF1, 1MF1

Exam(s) 4th BSc. In Information Technology
1st MSc. Software Design and Development

Module Code(s) CT422

Module(s) Modern Information Management

Paper No. _____
Repeat Paper _____

External Examiner(s) Prof. P. Nixon

Internal Examiner(s) Prof. G. Lyons

Mr. C. O'Riordan

Instructions Answer any 4 questions
All questions carry equal marks.

Duration _____ 3 hours

No. of Answer books _____ 1

Requirements:

Handout _____

MCQ _____

Statistical Tables _____

Graph Paper _____

Log Graph Paper _____

Other Material _____

No. of Pages _____ 4

Department(s) Information Technology

SUMMER EXAMINATIONS 2004

CT422 Modern Information Management

Time allowed: **Three hours**

Answer any 4 questions

All questions carry equal marks

- Q.1**
- i) The Boolean and vector space models are information retrieval models in widespread use. Compare and contrast the two approaches. Your answer should include a description of query and document representations, comparison approaches and the relative strengths and weaknesses of each model. (9)
 - ii) Discuss the role of relevance feedback in information retrieval systems. Discuss techniques for query modification in the vector-based models. Your answer should include descriptions of query term re-weighting and query expansion. (8)
 - iii) Write short notes on two of the following approaches to Information Retrieval:
 - a) Extended Boolean Model.
 - b) Neural Network Models.
 - c) Probabilistic Models.
 - d) Fuzzy Set Approaches.

Your answer should include descriptions of how queries and documents are represented and how comparison is achieved.

(8)

- Q.2** For the requirements specified in *either* (i) *or* (ii) below, provide a high level design. For all components of the system, outline the algorithm and technique used. Outline any limitations/problems with your designed system.

- i) An information provider (articles, news-feeds, online magazines, online books etc.) needs to develop a system to provide a suitable means for customers to find the information they require. Notes: It should be possible for a user to provide queries on structured fields and to provide natural language queries representing their information needs. Users should be able to browse related items. Any evidence that can be used to guide/aid the user should be integrated into your solution.

- ii) An information provider has access to a set of distributed collections of documents. Each site has their own search engine to handle queries. The information vendor wishes to accept queries from his/her customers and retrieve suitable results from the set of distributed collections. Suggest approaches and algorithms that could be used to:

- a) select suitable collections to handle a submitted query.
- b) combine results from more than one source.

(25)

- Q.3** i) Many modern web-based search engines attempt to take into account the web link structure in addition to the content of pages. Outline an approach that could be adopted to use information embedded in the web link structure. Discuss any limitations associated with this approach. (11)

- ii) Write a note on current and future trends in web-search engines. Your answer should briefly review new features/techniques currently being explored in a particular area— e.g. personalisation, visualisation of results, exploiting multiple sources of information, etc. (14)

- Q.4** i) Discuss suitable indexing strategies and algorithms to deal with the following types of queries: *single term*, *prefix*. Include examples in your answer. (9)

- ii) Stop-word removal and stemming algorithms are common components of the pre-processing of document collections. Describe these processes and outline approaches to both stop-word removal and stemming. Discuss the advantages and possible disadvantages of these approaches. (8)

- iii) With respect to compression, outline, with an example, techniques that may be adopted to compress an inverted index. (8)

- Q.5**
- i) Many different activities may be involved in information retrieval/filtering systems: browsing, querying, viewing results, reformulating query etc. Explain, with reference to existing systems, the issues involved in designing an interface to aid the user with the involved activities. (9)
 - ii) The usefulness of an IR system is often measured using the metrics of precision and recall. Describe, with an example, these metrics. Discuss other metrics that have also been adopted. (8)
 - iii) Discuss techniques that can be adopted for retrieval of non-textual objects (e.g. image, music). (8)
- Q.6**
- i) Discuss the main steps involved in collaborative filtering. Discuss the main limitations associated with the approach. (9)
 - ii) An online retail outlet wishes to adopt some mechanism to identify groups of items that are regularly purchased together. Outline a solution that could be adopted. Your answer should include a description of algorithms and any associated limitations of the approaches. (8)
 - iii) Explain how a decision tree could be developed from a set of tuples of the form $\langle \text{attribute}_1, \dots, \text{attribute}_n, \text{category} \rangle$ such that future tuples of the form $\langle \text{attribute}_1, \dots, \text{attribute}_n \rangle$ can be accurately placed in a correct category. (8)