

OLLSCOIL NA hÉIREANN GAILLIMH  
NATIONAL UNIVERSITY OF IRELAND GALWAY

---

SEMESTER II  
SUMMER EXAMINATIONS 2003

---

Bachelor of Science in Information Technology (4IF1)

**Modern Information Management (CT422)**

Professor P. Nixon  
Professor G. Lyons  
Mr. C. O'Riordan

Time allowed: *three* hours.

**Attempt any *four* questions**

1. (a) Describe the main components of an Information Retrieval system. Outline the differences between information retrieval and traditional data retrieval. Discuss the main differences and similarities between information retrieval and information filtering. (8)
- (b) 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 representation, comparison approaches and the relative strengths and weaknesses of each model. (9)
- (c) Write short notes on two of the following approaches to Information Retrieval:
  - Latent Semantic Indexing.
  - Neural Networks.
  - Probabilistic networks.
  - Fuzzy Set representations.
 (8)
2. (a) Discuss the role of relevance feedback in information retrieval systems. Discuss techniques for query modification in both the vector based models and probabilistic models. Your answer should include query term re-weighting and query expansion. (10)
- (b) Outline the difficulties associated with obtaining explicit feedback from users. Suggest alternative mechanisms that could be used in an information retrieval or filtering system to obtain feedback. (7)
- (c) The usefulness of an information retrieval system is often measured using the metrics of precision and recall. Describe, with an example, these metrics. Discuss the suitability of these metrics for different types of information retrieval and filtering systems. (8)
3. (a) For the requirements specified in *either* (a) *or* (b) 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:

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

(25)

4. (a) Discuss suitable indexing strategies and algorithms to deal with the following types of queries: single term, prefix, Boolean expressions. (9)
- (b) 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)
- (c) With respect to compression, outline, with an example, Huffman based compression. Discuss the relative merits and shortcomings of word-based Huffman encoding over character-based Huffman encoding. (8)
5. (a) Discuss techniques that can be adopted for retrieval of non-textual objects (e.g image, music). (9)
- (b) Many different activities may be involved in an information retrieval/filtering system: 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. (7)
- (c) Outline possible approaches to aid the user in visualising both the relationship between the given query and the document set, and the relationship between returned documents and the remainder of the document set. Include in your answer a discussion on interface design issues and suitable underlying algorithms. (9)
6. (a) An online retail outlet wishes to adopt some mechanism to identify groups of items which 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)
- (b) 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. (8)
- (c) Describe the concept of collaborative filtering (recommender systems) outlining the main approaches that have been adopted. Discuss any limitations inherent in these approaches. (9)