



Autumn Examinations 2018

Course Instance Code(s) Exam(s)	4BCT1, 1EM1, 1CSD1, 4BS2 4th B.Sc. (Information Technology) Erasmus M.Sc. Data Analytics 4 th B.Sc. (Bachelor of Science)
Module Code(s) Module(s)	CT422 Modern Information Management
Paper No.	1
External Examiner(s)	Dr. Jacob Howe
Internal Examiner(s)	Professor Michael Madden *Dr. Colm O’Riordan

Instructions: Answer 3 questions. All questions carry equal marks.

Duration	2 hours
No. of Pages	2
Discipline(s)	Information Technology

Requirements:

Release in Exam Venue	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
MCQ Answersheet	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
Handout	None			
Statistical/ Log Tables	None			
Cambridge Tables	None			
Graph Paper	None			
Log Graph Paper	None			
Other Materials	None			
Graphic material in colour	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>

PTO

Q.1.

- (a) Describe the vector space model approach to information retrieval. Your answer should include a description of the query and document representations and also the comparison approach used. (8)
- (b) The accuracy of the vector space model depends on the quality of the weighting of the terms in both the query and documents. Discuss, with reference to well-known weighting schemes, the main components and properties of a good weighting scheme. (9)
- (c) The Extended Boolean model has been used to allow users express their queries in Boolean algebra and have documents ranked in order of relevancy by adopting a more expressive weighting scheme than afforded in the traditional Boolean model. Explain such an approach. Discuss the advantages and limitations of adopting such a model. (8)

Q.2.

- (a) Empirical evaluation of information retrieval systems plays an important role in information retrieval research. Define and discuss the following metrics that can be used to measure the performance of an Information Retrieval system: precision, recall, novelty and coverage. (9)
- (b) Discuss potential uses of clustering algorithms in the domain of information retrieval. Outline any clustering algorithm and discuss any limitations associated with that algorithm. (8)
- (c) Discuss suitable approaches to evaluating the usefulness of a clustering algorithm. (8)

Q.3.

- (a) Link analysis has been used in many modern web search engines. Describe, with a suitable example, how link analysis can be applied to the answer to obtain a new ranking over the returned set for a specific query. Outline any difficulties or limitations of this approach. (8)
- (b) Define what is meant by collaborative filtering. Describe, with a suitable example, the main stages involved in generating a recommendation for a user via collaborative filtering. (10)
- (c) Classical collaborative filtering considers ratings provided by users for items. In many domains, extra information is also available regarding users and items; discuss how this extra information could be used to increase the coverage of a collaborative filtering system. (7)

Q.4.

- (a) User feedback is often used to modify a user's query with the aim of improving retrieval performance. Outline such an approach for the vector space model. (8)
- (b) Query augmentation can also take place without explicit user feedback. Outline an approach to automatically generate suggested keywords for a user to augment their query. (9)
- (c) Describe the term query difficulty and outline measures to predict the difficulty of a query. (8)