NUI Galway
OÉ Gaillimh

Autumn Examinations 2018-2019

Course Instance Code(s)	4BCT1, 1EM1, 1CSD1, 1CSD2, 4BS2
Exam(s)	4th B.Sc. (Computer Science and Information Technology) Erasmus M.Sc. Data Analytics 4th B.Sc. (Bachelor of Science)
Module Code(s)	CT422
Module(s)	Modern Information Management
Paper No.	1
External Examiner(s)	Dr Jacob Howe
Internal Examiner(s)	Professor Michael Madden *Dr. Colm O'Riordan

Instructions: Candidates should answer any **THREE** questions.
All questions will be marked equally.

Duration	2 hours
No. of Pages	3
Discipline(s)	Information Technology
Course Co-ordinator(s)	Dr D Chambers

Requirements:

Release in Exam Venue	Yes	☐	No	☐
MCQ Answersheet	Yes	☐	No	☐
Handout	None			
Statistical/ Log Tables	None			
Cambridge Tables	None			
Graph Paper	None			
Log Graph Paper	None			
Other Materials	None			
Graphic material in colour	Yes	☐	No	☐

PTO

Q.1.

(a) Learning mechanisms has been used successfully to search for suitable means to combine sources of evidence in information retrieval. Discuss such an approach applied to a problem of your choice in information retrieval. Your answer should also identify the strengths and weaknesses of this approach. (10)

(b) Many modern web-based search engines attempt to take into account the web link structure in addition to the content of the pages. Describe the Page Rank algorithm that uses information embedded in the web link structure to return relevant documents to a user. (10)

(c) Illustrate, with a suitable example, how the rank of the pages may be calculated according to the PageRank algorithm. (5)

Q.2

(a) The vector space model for Information Retrieval is one of the most commonly adopted models. Outline the model explaining both the representation of queries and documents and a means to calculate similarity. Discuss the advantages and disadvantages of such a model. (9)

(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 of a good weighting scheme. (10)

(c) Prior to calculating the similarity between a query and a document, documents and queries are often pre-processed using stemming and stopword removal prior to assigning weights to terms. Explain what is meant by pre-processing and the possible effects it may have on the retrieval process. (6)

Q.3.

(a) Explain what is meant by collaborative filtering. Explain briefly the main stages involved in generating a prediction/recommendation for users. (5)

(b) Discuss what you consider to be the main limitations of a collaborative filtering approach and suggest, with suitable examples, approaches to overcome these limitations. (10)

(c) With reference to a clustering algorithm of your choice, describe suitable approaches to measuring the quality of the clustering algorithm. Your answer should distinguish between internal and external criteria. (10)

Q.4.

(a) Explain the terms precision and recall and discuss their suitability as a means of measuring the performance of information retrieval system. (6)

(b) Query modification is often used by systems to attempt to improve precision and recall for a given information need. Discuss an approach, given user feedback on the returned answer set, to improve the performance of the query. (10)

(c) 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)