NUI Galway
OÉ Gaillimh

Semester 1 Examination 2018/2019

Exam 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
Repeat Paper	
External Examiner(s)	Dr. Jacob Howe
Internal Examiner(s)	Prof. Michael Madden Dr. Colm O'Riordan*

Instructions: Candidates should answer any **Three** questions

All questions carry equal marks.

Duration: 2 hours

No. of Pages 3

Discipline(s) Information Technology

Requirements:
MCQ
Handout
Statistical/ Log Tables
Cambridge Tables
Graph Paper
Log Graph Paper
Other Materials

Release to Library: No

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. Illustrate how the page rank of the pages in the following graph might be calculated.

The graph contains three nodes a,b,c with the following directed edges (a,b),(a,c),(b,c),(c,b) (10)

c) Discuss any limitations associated with the PageRank approach and suggest how one may overcome such limitations. (5)

Q.2.

a) Prior to calculating the similarity between a query and a document, documents and queries are often pre-processed 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)

b) 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)

c) 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)

Q.3.

a) 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. Outline how your approach could include expressive feedback where users can assign a numeric score representing the relevance of a returned document. (10)

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) Suggest a means to evaluate the performance of the query modification techniques discussed (using explicit feedback or pseudo feedback). (6)

Q.4.

a) Explain what is meant by *collaborative filtering*. Explain, with a suitable example, the main stages involved in generating a prediction/recommendation for users. (10)

b) Discuss what you consider to be the main limitations of a collaborative filtering approach and suggest approaches to overcome these limitations. (5)

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)