Course Instance Code(s)	4BCT, 1CSD1, 1CSD2, 4BS2
Exam(s)	4th B.Sc. Computer Science and IT
	M.Sc. Computer Science (Data Analytics)
	M.Sc. Computer Science (Artificial Intelligence)
	B.Sc. (Hons)
Module Code(s)	CT4100
Module(s)	Information Retrieval

Paper No.	1

External Examiner(s)	Dr. Ramona Trestian, Dr. John Woodword
Internal Examiner(s)	Professor M. Madden
	*Dr. C. O'Riordan

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

Duration	2 hours
No. of Pages	3
Discipline(s)	Computer Science
Course Co-ordinator(s)	Dr. Colm O'Riordan,
	Dr. Frank Glavin,
	Dr. Mattias Nickels

Requirements:

Release in Exam Venue   Yes

CT4100 Information Retrieval

Question 1   (25 marks)

(a) Recommender systems are used to generate recommendations for users on unseen items. Collaborative filtering is one such approach. Explain, in your own words, the main stages of collaborative filtering and illustrate how this approach can be used to generate a recommendation.   (10)

(b) Suggest approaches to evaluate the usefulness of a collaborative filtering system. Discuss any limitations of these approaches   (8)

(c) Outline a suitable approach to incorporating content information about items into a recommender system.   (7)

Question 2   (25 marks)

(a) Precision and recall are often used to evaluate the performance of an IR system. Explain these terms. Given a ranked answer set and evaluation judgements, explain how you would generate a precision-recall graph.   (8)

(b) Outline any limitations associated with using precision and recall and suggest alternative evaluation measures.   (7)

(c) Learning mechanisms has been used successfully in information retrieval. Using an approach of our choice, suggest a learning mechanism to identify good weights to apply to terms in documents. Discuss any limitations of your approach   (10)

Question 3   (25 marks)

(a) Describe, in your own words, with reference to any well-known term weighting scheme, the main constituents of a good weighting scheme. (9)

(b) Local analysis of the returned answer set is often used to select a number of terms to add to the current query. Describe an approach that may be used to select candidate terms to add to the query. Discuss the advantages and potential limitations of your approach. (8)

(c) Suggest an approach to extend local analysis in order to ensure a diverse set of candidate terms. (8)

Question 4   (25 marks)

(a) Discuss approaches to identify the difficulty of a user provided query. Your answer should include features that are used to measure the level of difficulty and you should distinguish between pre-retrieval and post-retrieval approaches. (8)

(b) Given a set of scientific articles (which contain title, abstract, authors, key words, main body of the paper and a bibliography), suggest a suitable approach to measure the similarity between these documents in the collection into useful sub clusters that may be of use in user search tasks. (8)

(c) Given the collection described in (b) and using the similarity measure(s), you have defined, describe an approach to clustering this collection in sub collections that may be of use to users engaging in search tasks. (9)