



Semester 1 Examinations 2020-2021

Course Instance 4BCT1, 1CSD1, 1CSD2, 1MAI1, 4BS2
Code(s)
Exam(s) 4th B.Sc. Computer Science and IT
M.Sc. in Computer Science (Data Analytics)
M.Sc. in Computer Science (Artificial Intelligence)
B.Sc. (Hons)

Module Code(s) CT4100
Module(s) Information Retrieval

Paper No. 1

External Examiner(s) Professor J. Howe
Internal Examiner(s) Professor M. Madden
*Dr. C. O’Riordan

Instructions: Answer any three questions. All questions will be marked equally.

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

Requirements:

Release in Exam Venue	Yes	<input type="checkbox"/>	No	<input checked="" type="checkbox"/>
MCQ Answer sheet	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

CT4100 Information Retrieval

Q.1.

(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 with a suitable example. (10)

(b) In many domains, in addition to ratings, extra information is often present and may prove useful in the provision of useful recommendations. For example, in the domain of movie recommendations, we may have information on each movie such as: year, producers, actors, genre. Outline an approach where this extra information can be used to hopefully improve the performance of the system. (8)

(c) Suggest an approach to attempt to ensure a diverse set of recommendations to the user. (7)

Q.2.

(a) A user submits a query, q , to an information retrieval system which returns a ranked list to the user. Given the top k ranked documents, explain how the query q could be modified using evidence from the returned set to improve performance. Outline any potential limitations of this approach. (8)

(b) Given the original query q and the modified query q' , discuss, in your own words, how you might predict which query is likely to perform better. (9)

(c) In a collaborative filtering scenario, given a user profile (their ratings) and a set of recommendations made for that user, suggest measures that could be adopted to identify whether these recommendations are likely to be good. (8)

Q.3.

- (a) Describe, in your own words, with reference to any well-known term weighting scheme, the main constituents of a good weighting scheme. (7)
- (b) Axiomatic approaches have been proposed which specify how the similarity of document to a query should change given the addition of terms to a document. Discuss, in your own words, the advantages of such an approach. (6)
- (c) Consider the following scenario: a company search engine is employed to allow people to search a large repository. All queries submitted to the system are recorded. A record that contains the id of the user and the terms in the query is stored. The timestamp of every query is recorded.

Each entry is: a user id, a timestamp and a set of terms. Any duplicate terms in a query are ignored. The designers of the search engine decide to use this information to develop an approach to make query term suggestions for users, i.e., at run time once a user has entered their queries terms, the system will suggest potential extra terms to add to the query.

Given the data available, outline an approach that could be adopted to generate these suggested terms. Identify advantages and disadvantages of your approach (briefly).

(12)

Q.4.

- (a) A company has a large set of scientific articles (each of which contains a title, abstract, authors, key words, year of release, main body of the paper and a bibliography)

Suggest a means to measure the similarity between two documents based on:

- i) Content of the document (4)
- ii) Authors listed (2)
- iii) Bibliographies (4)
- iv) Content, authors and bibliographies (2)

- (b) The company wishes to rank all papers that are relevant to a given query and to then re-order the papers in the answer set according to how authoritative or influential the papers are. Outline an approach that could be used to give a suitable solution for this requirement. (13)