



University College Dublin
An Colaiste Ollscoile, Baile Átha Cliath

Autumn, 23/24 TRIMESTER EXAMINATIONS

COMP41730

Text Analytics (5 credits)

Module Coordinator: Professor Mark Keane

Student Number

--	--	--	--	--	--	--	--

Seat Number

--	--	--	--

Time Allowed: 120 minutes

Materials Permitted in the Exam Venue:

None

Materials to be Supplied to Students:

12 Page Answer Booklets

Instructions to Students:

Answer any FOUR questions.
All questions carry equal marks.
Total marks available 100.
Use of calculators is prohibited.

Instructions to Invigilators:

Use of calculators is prohibited.

1. Three main methods are used to identify sentiments in texts: human ratings, sentiment lexicons and sentiment classifiers. Describe each of these methods in detail and describe how they are used. Also critically evaluate the adequacy of each of these methods as solutions to the sentiment-identification problem. **[3 x 8.33 marks]**

[25 marks overall]

2. Latent Methods (e.g., Latent Semantic Indexing and Latent Semantic Analysis) are used in text analytics to find hidden or latent associations between words. Describe how LSI/LSA are computed when applied to a term-document matrix. **[10 marks]** Evaluate these methods on some of the issues that arise in their use. **[5 marks]** How do these methods differ from more recent attempts to define word embeddings (e.g., in word2vec or in BERT)? **[10 marks]**

[25 marks overall]

3. There are three main approaches to similarity that group the techniques used in text analytics. Describe each approach and illustrate each with an example, identifying when it is appropriate to use one approach-technique rather than another, and any issues that can arise in their use. **[3 x 8.33 marks]**

[25 marks overall]

4. In machine learning, a fundamental distinction is often made between supervised and unsupervised methods. Describe the main differences between these two broad classes of methods. **[5 marks]** Then, give detailed accounts of one example of each class (i.e., provide two specific technique descriptions, one that is supervised and one that is unsupervised). **[2 x 5 marks]** Finally, illustrate each of these techniques with an example from the text analytics literature. **[2 x 5 marks]**

[25 marks overall]

5. Four key ideas -- Ground Truth, Precision, Recall and the F-measure -- are typically used to evaluate text-analytics systems. Why do we need to evaluate these systems, at all ? Describe these four key ideas, giving an example of their application and explaining their role in system evaluation. **[4 x 6.25 marks]**

[25 marks overall]

oOo