University College Dublin
An Coláiste Ollscoile, Baile Átha Cliath

SEMESTER I EXAMINATIONS

ACADEMIC YEAR 2017/2018

Module Code XXXX

Module Title

Prof. J. Pitt

Prof. P. Cunningham

Prof. M Keane *

Time Allowed: 2 Hours

Instructions for Candidates

Answer any FOUR questions.
All questions carry equal marks.
Total marks available 100.

Instructions for Invigilators

Use of calculators is prohibited.

1. Several different methods have been used to find temporal regularities in textual data. Describe three of the main methods that have been used in this area, illustrating each with an example from the literature, while critically evaluate each method.

[3 x 8.33]
[25 overall]

2. 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 8.333%]
[25% overall]

3. What are unsupervised, machine learning techniques predominantly used for in Text Analytics? Describe three specific, unsupervised techniques and explain when it is appropriate to use each technique.

[5 x 5%]
[25% overall]

4. In Text Analytics, Log-likeihood Ratios (LLRs) are often used to find patterns of significant words in comparisons between different corpora. Describe how LLRs are computed using a numeric example. Elaborate two different cases of corpus comparison where LLRs have been used to find word patterns and critically evaluated this work.

[3 x 8.333%]
[25% overall]

5. Text Analytics typically begins with the pre-processing of each text, in some selected corpus of texts, to prepare it for subsequent processing.

   Describe five of the typical pre-processing steps that are carried out during in this initial stage of processing and show how each pre-processing step might modify a text fragment. In describing each pre-processing step, explain why it is used and describe some of the benefits it brings to subsequent processing.

[5 x 5%]
[25% overall]