University College Dublin

An Coláiste Ollscoile, Baile Átha Cliath

SEMESTER I EXAMINATIONS

ACADEMIC YEAR 2017/2018

COMP 47600

Using Text Analytics to Discover Meaning

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.

Use of calculators is prohibited.

Instructions for Invigilators

Use of calculators is prohibited.

1. Several different approaches have been used to find temporal regularities in text data. Describe three of the approaches that have been used, illustrating each with an example from the literature, with a critical evaluation of each.

[25 marks]

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 how they are used. Also critically evaluate the adequacy of each of these methods in sentiment identification.

[25 marks]

3. List three supervised, machine-learning techniques that are commonly used in text analytics? Describe each of these supervised techniques, in detail, providing an example of each from the literature. Explain when it is appropriate to use each technique.

[25 marks]

4. Describe a few of the ways in which Log-Likelihood Ratios (LLRs) have been used to find patterns of significant words in comparisons between different texts or text-corpora. Provide a numeric example showing how LLRs are computed. Describe one study from the text analytics literature that used LLRs and critically evaluate this study.

[25 marks]

5. Text analytics typically begins with the pre-processing of each text-item, in some selected corpus, to prepare it for subsequent processing. Describe five of the typical pre-processing steps that are carried out during this pre-processing stage and show how each step modifies a given text fragment. In describing each pre-processing step, explain why it is used and the benefits that follow from its use.

[25 marks]