



University College Dublin  
An Coláiste Ollscoile, Baile Átha Cliath

---

**SEMESTER I EXAMINATIONS**  
**ACADEMIC YEAR 2018/2019**

---

**COMP 47600**

**Using Text Analytics to Discover Meaning**

Assoc. Prof. V. Dimitrova

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. Most text analytics systems have a pipeline of pre-processing steps that are applied to each text-item in a corpus, to prepare it for subsequent analysis. Describe five of the main pre-processing steps that are carried out in a typical pipeline, showing how each step modifies a sample text fragment. In describing each pre-processing step, explain why it is used and the benefits that follow from processing the textual data in this way.

[25 marks]

2. 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.

[25 marks]

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

[25 marks]

4. Log-Likelihood Ratios (LLRs) are used in text analytics to find patterns of significant words in comparisons between different texts or text-corpora. What formula is used to compute LLRs? Describe four different ways in which LLRs have been applied to textual data, giving a sample study for each application, along with a critical evaluation of that study.

[25 marks]

5. Five main approaches have been taken in the text analytics literature to the problem of finding temporal regularities in textual data. Describe each of these five approaches, illustrating each one with an example of its application in the literature, and provide a critical evaluation of each study mentioned.

[25 marks]