



COMP30810

Intro to Text Analytics

Dr. Binh Thanh Le

thanhbinh.le@ucd.ie

Insight Centre for Data Analytics

School of Computer Science

University College Dublin

Today Goals:

- Module Overview
 - Learning Objectives
 - Module Topics
 - Marking Process
 - Late Submission Policy
 - Marking to grades mapping used in this module
- Questions & E-mail

Overview

- Module materials/announcements on Moodle
- Moodle Website:
 - <https://csmoodle.ucd.ie/moodle/course/view.php?id=701>
 - Self-enroll with your user name and password: **comp30810181**

Schedule:

| | Monday (LAB) | Wednesday (LEC) | Thursday (LEC) |
|------|-----------------|-----------------|----------------|
| Time | 10:00 (110mins) | 12:00 (50mins) | 10:00 (50mins) |
| Room | A109 ART | 129 SCN | G80 NEWST |

laptop required!

- Course Lecturer:
Dr. Binh Thanh Le (<https://www.insight-centre.org/users/thanh-binh-le>)
- Office hours: by appointment; @UCD- Insight C17 (O'Brien Centre)
- Support team: **Iulia Ilie** (iulia.ilie@insight-centre.org)
John Sloan (john.sloan.1@ucdconnect.ie)

Module: **COMP30810 Introduction to Text Analytics**

Module Coordinator: **Dr. Binh Thanh Le**

Semester: **1**

Delivery:

22 x 1h lectures

10 x 2h labs

2h topic seminar

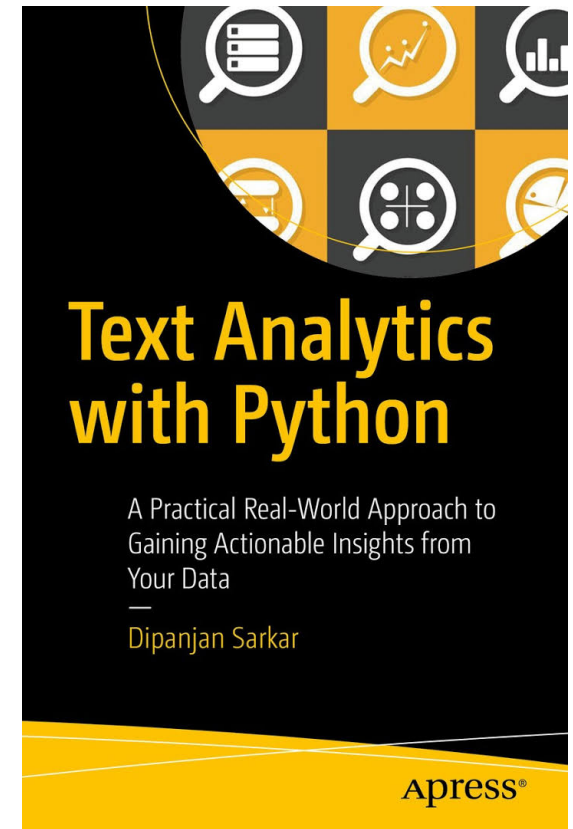
Reference Material:

Lecture Notes

Recommended Books

Notebooks

Worked Examples



Learning Objectives

1. Have a clear understanding of the **core concepts** in text analytics and be familiar with a range of **common analytics techniques** in this area.
2. Be familiar with **common representations** used to represent textual data.
3. Have developed a **knowledge of the main application areas** in which these techniques prove useful.
4. Be able to **collect, pre-process and analyze** textual datasets using Python.

Marking Process

- Marks distributed over 4 components with different weight (40% H1, 45% H2, 5% attendance, 5% discussion)
- **Homework: 85%**
 - H1: Document Understanding; out in week 3 , due in week 6
 - H1: individual homework
 - H2: Model & Evaluation; out in week 7 , due in week 10
 - H2: competition from groups.
 - 2 presentations for the top results
- **Discussion: 10%**
 - Good questions, good answers (moodle forum)
 - Researching and discussing in front of class
- **Attendance: 5%**

Late Submission Policy

- Up to 1 week late, 10% marks lost
- Between 1-2 weeks late, 10% - 20% marks lost
- 2 weeks cut-off (no submissions accepted after 2 weeks)

Marking to grades mapping used in this module

| Grade | Lower | Upper | Calculation Point |
|-------|-------|-------|-------------------|
| A+ | 95 | 100 | 97.5 |
| A | 90 | 95 | 92.5 |
| A- | 85 | 90 | 87.5 |
| B+ | 80 | 85 | 82.5 |
| B | 75 | 80 | 77.5 |
| B- | 70 | 75 | 72.5 |
| C+ | 65 | 70 | 67.5 |
| C | 60 | 65 | 62.5 |
| C- | 55 | 60 | 57.5 |
| D+ | 50 | 55 | 52.5 |
| D | 45 | 50 | 47.5 |
| D- | 40 | 45 | 42.5 |

| Grade | Lower | Upper | Calculation Point |
|-------|-------|-------|-------------------|
| E+ | 35 | 40 | 37.5 |
| E | 30 | 35 | 32.5 |
| E- | 25 | 30 | 27.5 |
| F+ | 20 | 25 | 22.5 |
| F | 15 | 20 | 17.5 |
| F- | 10 | 15 | 12.5 |
| G+ | 8 | 10 | 9 |
| G | 5 | 8 | 6.5 |
| G- | 2 | 5 | 3.5 |
| NG | 0 | 0 | 0 |

Questions & E-mail

- If you have a technical question, discuss in the lecture/lab or post to Moodle
- Only send e-mail to TA/lecturer if of personal nature (e.g., health or grade related)
- If you need to discuss with the TA/lecturer outside the lecture/lab hours, please set an appointment