Introduction of Text Mining

2020 fall

CHIEN CHIN CHEN

Course Information (1/5)

Instructor:
- 陳建錦 (Chien Chin Chen)
- E-mail: patonchen@ntu.edu.tw
- Office: Mgmt II 414.
- Office hours: by appointment

Teaching assistant:
- 賴柏霖, r08725022@ntu.edu.tw
- 羅良瑋, r08725024@ntu.edu.tw
- 陳芝妘, r08725031@ntu.edu.tw

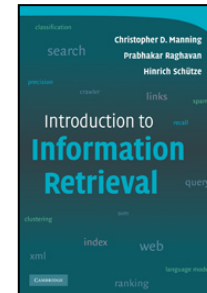Class time: 9:10~12:10, every Friday.

Location: Mgmt II 104
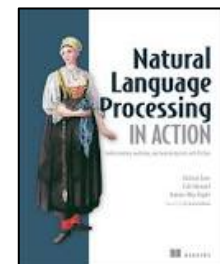
Course website: NTU COOL

Course Information (2/5)

Text book:

◦ Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press, 2008.

References:

◦ Hobson Lane, Cole Howard, and Hannes Max Hapke, Natural Language Processing in Action, Manning, 2019.

◦ Christopher D. Manning and Hinrich Schütze, Foundations of Statistical Natural language Processing, The MIT Press, 1999.

Course Information (4/5)

This course **would** cover the following contents:

- Term Vocabulary
- Term Weighting and Vector Space Model
- Text Classification and Naive Bayes
- Vector Space Classification
- Flat Clustering
- Hierarchical Clustering
- Latent Semantic Analysis
- Language Models
- Link Analysis
- Concept of Neural Network
- Word2Vec & Bert

Course Information (5/5)

Prerequisite:
- Programming
  - We have a lot of text mining code examples using Python.
- Probability
  - Many text mining model are based on term frequency.

Grading:
- Programming assignment (at least 4 assignments): 30%
- Final: 35%, on 1/8/2021
- Term project: 35%, demo on 1/15/2021

Programming Assignment Policy

Please <u>zip and submit your source code</u> and a <u>brief note</u> via email to TA.

- ◦ Format:
  - ◦ Plain A4 size page
  - ◦ ≥ 2 pages
  - ◦ Font size: 12 point
  - ◦ Single column
  - ◦ Single space

Late homework would not get any credit!!

Students who copy (or copied) other's code will not get any credit!!

Term Project Policy (1/2)

Each team can have at most 5 members

Please prepare a one-page proposal and send it to TA on **12/11/2020** (font size: 12 point, single column, single space)

◦ The name of your project
◦ Members
◦ A brief description of your project and possible solutions

Each team will have an **5**-minute presentation to show your system

◦ An 1-minute Q&A

Term Project Policy (2/2)

Each team also needs to submit a project report to TA before **1/22/2021**.

- ◦ Format: http://www.acm.org/sigs/publications/proceedings-templates
- ◦ At least 4 pages
- ◦ The purpose of your project
- ◦ Your solution
- ◦ System outcomes
- ◦ Conclusions
- ◦ Member's workload
- ◦ References

Schedule Note

Sorry that I have to manage 碩士推甄 on 11/6 ☹

NO CLASS on 11/6!!

BUT … we do have class on the midterm day, that is, 11/13!!