國立東華大學

教學計劃表 Syllabus

課程名稱(中文) Course Name in Chinese	文字探勘			學年/學期 Academic Year/Semester	111/2
課程名稱(英文) Course Name in English	Text mining				
科目代碼 Course Code	FIN_21060	系級 Department & Year	學三	開課單位 Course-Offering Department	財務金融學系
修別 Type	學程 Program	學分數/時間 Credit(s)/Hour(s)	3.0/3.0		
授課教師 Instructor	/林金龍				
先修課程 Prerequisite					

課程描述 Course Description

As there exist voluminous text information over the Internet and are accumulating at an amazing speed,
mining useful information from text becomes increasingly more important. Knowledge extracted from text are essential for marketing,
finance, social behavior analysis and many other fields.
This course will cover important topics in text mining including basic natural language processing techniques, document representation,
text categorization and clustering, document summarization, sentiment analysis, probabilistic topic models, text visualization and financial applications.
I shall emphasize applying textming for financial analysis. Real examples will be provided.

I choose R as the main software as it is free, powerful and very popular for text analytics. Yet, as Python has gained increasing popularity
in text mining and in other fields of big data analytics, I shall also used it as the second programming software.

課程目標 Course Objectives

教導學生文字探勘的理論與演算法，內容包含關鍵字抽取、關鍵字權重計算、向量空間模型、二元獨立模型、貝氏分類器、Rocchio分類器、nearest neighbor、k-means分群法與階層式分群法，學生得以用之分析財經資訊。

圖示說明Illustration ：● 高度相關 Highly correlated ○中度相關 Moderately correlated

授 課 進 度 表 Teaching Schedule & Content

週次Week	內容 Subject/Topics	備註Remarks
1	Overview of Text Mining (Silge and Robinson chap 1; Kwartler chap 1; Weiss, Indurkhya, and Zhang chap 1;)	
2	Software for text mining: R, Python, tmsk, nktext (Silge and Robinson chap 1; Kwartler chaps 2,3;)	
3	From Textual Information to Numerical Vectors (I)	
4	From Textual Information to Numerical Vectors (II)	
5	Using Text for Prediction	
6	Information Retrieval and Text Mining	
7	Finding Structure in a Document Collection	

8	Looking for Information in Documents	
9	期中考試週 Midterm Exam	
10	Text clustering	
11	Text mining for Chinese (I)	
12	Text mining for Chinese (II)	
13	Text mining for Chinese (III)	
14	\item Applying text mining for financial prediction	
15	Case Studies (I)	
16	Case Studies (II)	
17	Presentation	
18	期末考試週 Final Exam	

教 學 策 略 Teaching Strategies

☑ 課堂講授 Lecture ☑ 分組討論Group Discussion ☐ 參觀實習 Field Trip

☐ 其他Miscellaneous:

教 學 創 新 自 評 Teaching Self-Evaluation

創新教學(Innovative Teaching)

☐ 問題導向學習(PBL) ☐ 團體合作學習(TBL) ☐ 解決導向學習(SBL)

☐ 翻轉教室 Flipped Classroom ☐ 磨課師 Moocs

社會責任(Social Responsibility)

☐ 在地實踐Community Practice ☐ 產學合作 Industy-Academia Cooperation

跨域合作(Transdisciplinary Projects)

☐ 跨界教學Transdisciplinary Teaching ☐ 跨院系教學Inter-collegiate Teaching

☐ 業師合授 Courses Co-taught with Industry Practitioners

其它 other:

學期成績計算及多元評量方式 Grading & Assessments									
配分項目 Items	配分比例 Percentage	多元評量方式 Assessments							
		測驗會考	實作觀察	口頭發表	專題研究	創作展演	卷宗評量	證照檢定	其他
平時成績 General Performance			✓						
期中考成績 Midterm Exam									
期末考成績 Final Exam									
作業成績 Homework and/or Assignments	50%								
其他 Miscellaneous (期末報告)	50%								

評量方式補充說明
Grading & Assessments Supplemental instructions

教科書與參考書目（書名、作者、書局、代理商、說明）
Textbook & Other References (Title, Author, Publisher, Agents, Remarks, etc.)

Main textbooks: Julia Silge and David Robinson (2017) Text Mining with R: A Tidy Approach}, 1st Edition
manuscript available at https://github.com/dgrtwo/tidy-text-mining. Codes are available at https://github.com/PBrad/text-mining

Reference books:

Ted Kwartler (2017), Text Mining in Practice with R, John Wiley: available at https://onlinelibrary.wiley.com/doi/book/10.1002/9781119282105. Github Repo available at https://github.com/kwartler/text_mining

Sholom M. Weiss, Nitin Indurkhya, and Tong Zhang (2015) Fundamentals of Predictive Text Mining, 2nd ed. Springer. E-book available at \url{http://www.springer.com/us/book/9781447125655}

Christopher D. Manning, Prabhakar Raghavan and Hinrich Schutze (2009), Introduction to Information Retrieval, Cambridge University Press, E-bok available at \url{http://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf.

Sanjiv Ranjan Das (2014) Text and Context: Language Analytics in Finance, Now Publishers. E-book available at http://srdas.github.io/Papers/Das_TextAnalyticsInFinance.pdf

Ronen Feldman and James Sanger (2007), \textbf{The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data}, Cambridge University Press. E-book available at http://srdas.github.io/Papers/Das_TextAnalyticsInFinance.pdf

Michael W. Berry and Jacob Kogan, \textbf{Text Mining Applications and Theory}, John Wiley 2010.

Jiawei Han, Micheline Kamber, and Jian Pei, \textbf{Data Mining: Concepts and Techniques}, 3rd edition, Morgan Kaufmann Publishers, , 2012.

課程教材網址(含線上教學資訊,教師個人網址請列位於本校內之網址)
Teaching Aids & Teacher's Website(Including online teaching information. Personal website can be listed here.)

其他補充說明 (Supplemental instructions)