

國立中興大學教學大綱

社會領域－資訊與傳播學群（統合領域－專業實作）

課程名稱 (course name)	文本探勘：小技術大應用 Text Mining: Small Skill and Big Application				
開課系所班級 (dept. & year)	通識教育中心	學分 (credits)	3	規劃教師 (teacher)	教研所 許健將
課程類別 (course type)	必修	授課語言 (language)	中文或英文	開課學期 (semester)	上或下
課程簡述 (course description)	<p>本課程《文本探勘小技術大應用》，係以建構理論之「過程導向引導探究學習」(Process Oriented Guided Inquiry Learning)為基礎，以互動式講授(interactive lecture)與反思(reflection)與為策略，藉以培養本校人文、社會與法政相關系所學生運用 R 語言之「文字探勘」(text mining)技術進行相關小說、著作、新聞、文件或網路輿情之分析能力，促進其對「大數據」在各該領域未來之發展與應用有更深層之認識，促發其對培養跨領域能力的重視，增加其就業競爭力。</p> <p>This course "Text Mining: Small Skill & Big Application" is based on the "Process Oriented Guided Inquiry Learning" of construction theory, with interactive lectures and reflection and strategies. To cultivate the students' ability to analyze novels, books, news, documents or online public opinion by R "text mining" technology and promote their deeper understanding in "big data" and the future development and application of this field, prompting its emphasis on cultivating cross-disciplinary capabilities and increasing its employment competitiveness.</p>				
先修課程 (prerequisites)	無				
課程目標與核心能力關聯配比(%) (relevance of course objectives and core learning outcomes)					
課程目標	course objectives			核心能力 core learning outcomes	配比 合計 100%
培養本校人文、社會與法政相關系所學生運用 R 語言之「文字探勘」(text mining)技術進行相關小說、著作、新聞、文件或網路輿情之分析能力，促進其對「大數據」在各該領域未來之發展與應用有更深層之認識，促發其對培養跨領域能力的重視，增加其就業競爭力。	To cultivate the students' ability to analyze novels, books, news, documents or online public opinion by R "text mining" technology and promote their deeper understanding in "big data" and the future development and application of this field, prompting its emphasis on cultivating cross-disciplinary capabilities and increasing its employment competitiveness.			人文素養	20%
				科學素養	40%
				溝通能力	10%
				創新能力	30%
				國際視野	0%
				社會關懷	0%

課程目標之教學方法與評量方法
(teaching and assessment methods for course objectives)

教學方法 (teaching methods)	學習評量方式 (evaluation)
互動式講授(interactive lecture) 反思教學(reflective teaching) 上機實作	利用「過程導向的探索式學習」評量表(Process Oriented Guided Inquiry Learning, POGIL) (Loertscher, 2010)，將 R 語言與「文字探勘」知識建構拆解成認知(cognitive)、社會性(social)、情意(affective)、後設認知(meta-cognitive)等四個面向，使知識成為可觀察、能測量的能力。 各面向依據其評量指標，分別給予極優良(90 分以上)、優良(89~80)、通過(79~70)與待改進(69 分以下)四個評量等第。

授課內容 (單元名稱與內容、習作 / 考試進度、備註)
(course content and homework / tests schedule)

- 單元 1：介紹文本探勘的意義與其應用
- 單元 2：R 之字符串處理之基本操作
- 單元 3：R 之 stringr 套件進行字符串處理
- 單元 4：R 之 tm 套件進行分析資料之準備
- 單元 5：R 之 jiebaR 套件進行中文斷詞
- 單元 6：R 之 Wordcloud2 套件製作文字雲
- 單元 7：介紹網絡爬蟲與 R 語言之相關套件
- 單元 8：R 之 rvest 套件進行網頁爬蟲
- 單元 9：R 之 rvest 套件進行新聞爬蟲
- 單元 10：介紹 R 之 RSelenium 套件
- 單元 11：示範 R 之 RSelenium 套件的網頁爬蟲
- 單元 12：網路新聞與文本探勘
- 單元 13：認識 Gutenberg Project 所提供之上萬本免費中英文世界名著電子書之內容
- 單元 14：介紹 tidy 文本格式與符號化，以七言絕句進行示範
- 單元 15：介紹 tidytext 套件的三種英文情感詞典
- 單元 16：示範英文小說《Little Women》之情感分析
- 單元 17：示範中文小說《西遊記》之情感分析

教科書與參考書目 (書名、作者、書局、代理商、說明)
(textbook & other references)

1. 許健將、劉福蓀(2020)。文本探勘：小技術大應用。台北：全華圖書。
2. Dhawan, V., & Zanini, N. (2014). Big data and social media analytics. *Research Matters: A Cambridge Assessment Publication*, 18, 36–41.
3. Thanasuan, K., Chaisangmongkon, W., & Wongviriyawong, C. (2017). *Emerging Patterns in Student's learning Attributes through Text Mining*. Proceedings of the 10th International Conference on Educational data Mining. http://educationaldatamining.org/EDM2017/proc_files/papers/paper_83.pdf

課程教材 (教師個人網址請列在本校內之網址)
(teaching aids & teacher's website)

另行公告

課程輔導時間
(office hours)

另行公告