

淡江大學 107 學年度第 2 學期課程教學計畫表

課程名稱	文字探勘	授課 教師	陳景祥 CHEN CHING-HSIANG
	□TEXT MINING		
開課系級	大數據碩一 A	開課 資料	選修 單學期 3學分
	TLXDM1A		
系 (所) 教育目標			
<p>一、培育學生具研究大數據的能力。</p> <p>二、培育學生具大數據程式設計的能力。</p>			
系 (所) 核心能力			
<p>A. 具研究大數據分析理論的能力。</p> <p>B. 具大數據分析的能力。</p> <p>C. 具整合各領域之知識的能力。</p> <p>D. 具邏輯推理的能力。</p>			
課程簡介	本課程介紹文字探勘所需的相關理論以及R語言文字探勘相關程式技巧		
	This course will introduce text mining theories and coding techniques with R language.		

本課程教學目標與目標層級、系(所)核心能力相關性

一、目標層級(選填)：

- (一)「認知」(Cognitive 簡稱C)領域：C1 記憶、C2 瞭解、C3 應用、C4 分析、C5 評鑑、C6 創造
- (二)「技能」(Psychomotor 簡稱P)領域：P1 模仿、P2 機械反應、P3 獨立操作、P4 聯結操作、P5 自動化、P6 創作
- (三)「情意」(Affective 簡稱A)領域：A1 接受、A2 反應、A3 重視、A4 組織、A5 內化、A6 實踐

二、教學目標與「目標層級」、「系(所)核心能力」之相關性：

- (一)請先將課程教學目標分別對應前述之「認知」、「技能」與「情意」的各目標層級，惟單項教學目標僅能對應C、P、A其中一項。
- (二)若對應「目標層級」有1~6之多項時，僅填列最高層級即可(例如：認知「目標層級」對應為C3、C5、C6項時，只需填列C6即可，技能與情意目標層級亦同)。
- (三)再依據所訂各項教學目標分別對應其「系(所)核心能力」。單項教學目標若對應「系(所)核心能力」有多項時，則可填列多項「系(所)核心能力」。(例如：「系(所)核心能力」可對應A、AD、BEF時，則均填列。)

序號	教學目標(中文)	教學目標(英文)	相關性	
			目標層級	系(所)核心能力
1	中文斷詞技術	Chinese word segmentation techniques	P3	ABCD
2	文字探勘相關理論, 包含 TF-IDF、主題模型分析等	Text mining related theories, including TF-IDF, topic modeling etc.	C2	ABCD
3	用於文字探勘的資料探勘理論與技術	Applications of data mining theories and techniques in text mining.	P4	ABCD

教學目標之教學方法與評量方法

序號	教學目標	教學方法	評量方法
1	中文斷詞技術	講述、討論、實作	實作、報告、上課表現
2	文字探勘相關理論, 包含 TF-IDF、主題模型分析等	講述、討論	實作、報告、上課表現
3	用於文字探勘的資料探勘理論與技術	講述、討論、實作、問題解決	實作、報告、上課表現

本課程之設計與教學已融入本校校級基本素養

淡江大學校級基本素養	內涵說明
◆ 全球視野	培養認識國際社會變遷的能力，以更寬廣的視野了解全球化的發展。
◆ 資訊運用	熟悉資訊科技的發展與使用，並能收集、分析和妥適運用資訊。
◆ 洞悉未來	瞭解自我發展、社會脈動和科技發展，以期具備建構未來願景的能力。
◇ 品德倫理	了解為人處事之道，實踐同理心和關懷萬物，反省道德原則的建構並解決道德爭議的難題。
◆ 獨立思考	鼓勵主動觀察和發掘問題，並培養邏輯推理與批判的思考能力。
◇ 樂活健康	注重身心靈和環境的和諧，建立正向健康的生活型態。
◆ 團隊合作	體察人我差異和增進溝通方法，培養資源整合與互相合作共同學習解決問題的能力。
◆ 美學涵養	培養對美的事物之易感性，提升美學鑑賞、表達及創作能力。

授課進度表

週次	日期起訖	內容 (Subject/Topics)	備註
1	108/02/18~ 108/02/24	文字探勘簡介	
2	108/02/25~ 108/03/03	文字探勘與資料探勘之間的關聯性	
3	108/03/04~ 108/03/10	文字資料的收集：網路爬文(1)	
4	108/03/11~ 108/03/17	文字資料的收集：網路爬文(2)	
5	108/03/18~ 108/03/24	文字資料的資料清理方法	
6	108/03/25~ 108/03/31	中文斷詞理論、技術、與斷詞詞典製作(1)	
7	108/04/01~ 108/04/07	中文斷詞理論、技術、與斷詞詞典製作(2)	
8	108/04/08~ 108/04/14	新生詞彙的偵測與彙整	
9	108/04/15~ 108/04/21	文件語料庫(text corpus)處理	
10	108/04/22~ 108/04/28	DTM 矩陣與TDM矩陣處理與重要分析變數的萃取	
11	108/04/29~ 108/05/05	Big Data 狀況下的語料庫、DTM/TDM 矩陣處理	
12	108/05/06~ 108/05/12	文章集群分析與詞彙集群分析	

13	108/05/13~ 108/05/19	主題模型分析(Topic Modeling)(1)	
14	108/05/20~ 108/05/26	主題模型分析(Topic Modeling)(2)	
15	108/05/27~ 108/06/02	詞向量理論與技術(1)	
16	108/06/03~ 108/06/09	詞向量理論與技術(2)	
17	108/06/10~ 108/06/16	情感分析(Sentiment Analysis) (1)	
18	108/06/17~ 108/06/23	情感分析(Sentiment Analysis) (2)	
修課應 注意事項	(1) 期中考與期末考均採報告繳交方式 (2) 本課程使用 R 語言作為分析工具, 修課學生需具備 R 語言基本能力		
教學設備	電腦、投影機		
教材課本	教師自編講義		
參考書籍			
批改作業 篇數	4 篇 (本欄位僅適用於所授課程需批改作業之課程教師填寫)		
學期成績 計算方式	◆出席率： 20.0 % ◆平時評量：20.0 % ◆期中評量：30.0 % ◆期末評量：30.0 % ◆其他〈 〉： %		
備 考	「教學計畫表管理系統」網址： http://info.ais.tku.edu.tw/csp 或由教務處 首頁〈網址： http://www.acad.tku.edu.tw/CS/main.php 〉業務連結「教師教學 計畫表上傳下載」進入。 ※不法影印是違法的行為。請使用正版教科書，勿不法影印他人著作，以免觸法。		