



網路爬蟲與資料分析

Instructor: 馬豪尚

授課教師

- › 馬豪尚 (Hao-Shang Ma)
- › Office: 2504
- › Office Hours: 週二14:00-17:00, 週三10:00-12:00
- › Email: hsma@nutc.edu.tw

研究領域

- › 資料探勘 (Data Mining) / 機器學習(Machine Learning)
/ 人工智慧 (Artificial Intelligence)
 - 社群網路分析(Social Network Analysis)
 - 推薦系統(Recommender Systems)
 - 自然語言處理(Natural Language Processing)

課程大綱

- › 簡介
 - 網路通訊協定
 - HTML, JavaScript, JSON, CSV
- › Python資料處理函數
- › Python網路爬蟲和分析
- › Web API應用
- › 網路資料探勘

Python資料處理函數

› Numpy基礎操作

- 一維陣列操作
- 二維陣列操作
- 多維陣列操作

› Pandas

- 資料清理
- 資料操作
- 資料讀取與儲存
- 資料視覺化

Python網路爬蟲和分析

- › 爬取HTML網頁資料
- › 爬取動態網頁資料
- › 爬取網路論壇資料
- › 爬取股市金融資料
- › 爬取google Map資料

網路資料探勘

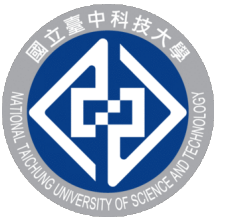
- › 全文檢索
 - 將全部的文字訊息儲存起來
 - 使用者必須詳細的規劃自己的查詢
- › 關鍵字查詢
 - 字詞切割
 - 關鍵字定義與比對
- › 自然語言處理

評分方式

- › 出席和上課態度 30%
- › 期中報告 30%
- › 期末專題 40%

參考用書

- › 文淵閣工作室, Python Programing Bible自學聖經, 碁峯出版, 2021第二版
- › 陳會安, Python從網路爬蟲到生活應用超實務, 博碩出版, 2022初版
- › 陳惠貞, HTML5、CSS3、JavaScript、jQuery、Vue.js、RWD網頁設計, 碁峯出版, 2022第八版



Teams團隊代碼

› 3t5skkg