

重新查詢

友善列印

### 0982學期 課程基本資料

系所 / 年級	資工系碩士班 1年級	課號 / 班別	65M00090 / A
學分數	3學分	選 / 必修	選修
科目中文名稱	文件探勘	科目英文名稱	Text minig
主要授課老師	王經篤	開課期間	一學年之下學期
人數上限	40 人	已選人數	24人

### 起始週 / 結束週 / 上課地點 / 上課時間

第1週 / 第18週 / I627 / 星期3第04節  
第1週 / 第18週 / I627 / 星期3第05節  
第1週 / 第18週 / I627 / 星期3第06節

請各位同學遵守智慧財產權觀念；請勿非法影印。

### 教學綱要

- 一、教學目標(Objective) 瞭解Text Mining的應用範圍與相關實作技術
- 二、先修科目(Pre Course) (1) Programming (2) Database
- 三、教材內容(Outline) Introduction What is text mining? Why use Perl? Pattern Matching Regular Expression Web Page Extraction Web crawlers, also known as Spiders or robots, are programs that automatically download Web pages.(Agent) Parser Quantitative Text Summaries Zipf's Law Probability and Text Sampling Probabilistic Models for Information Extraction. Hidden Markov Models (HMM) Information Retrieval (IR) Term-Document Matrix Term Weighting Document Similarity Text Categorization (TC) Given a set of categories (subjects, topics) and a collection of text documents, the process of finding the correct topic (or topics) for each document. (Ronen Feldman & Hame Sanger, 2007) Text Mining Application corporate finance patent research life sciences
- 四、教學方式(Teaching Method) (1)課堂講解 (2)上機 (3)專題實作與報告
- Text Book  
Practical Text Mining with Perl  
Roger Bilisoly  
Wiley (August 2008) ISBN: 978-0-470-17643-6  
Text Mining Application Programming  
Manu Konchady,  
Thomson (2006), ISBN: 1-58450-460-9  
Install  
The Text Mining Handbook  
Ronen Feldman & Hame Sanger, Cambridge(2007)ISBN:0-521-83657-3  
Reference Book

五、參考書目(Reference)

Search Engines: Information Retrieval in Practice  
W. Bruce Croft, Donald Metzler and Trevor Strohman  
Addison Wesley, (2010) ISBN-13: 978-0-13-607224-9  
Introduction to Information Retrieval  
Christopher D. Manning et al. Cambridge(2008)ISBN:978-0-521-86571-5  
<http://www-csli.stanford.edu/~hinrich/information-retrieval-book.html>  
投影片ppt  
Building Search Applications  
Lucene, LingPipe, and Gate  
Manu Knochady  
Mustru Publishing, (2008) ISBN: 978-0-61520-425-3

Web Data Mining  
Exploring Hyperlinks, Contents and Usage Data  
Bing Liu, Springer, December, 2006, ISBN:978-540-37881-5  
Data mining resource site: KDnuggets Directory  
MDR: Mining Data Records from Web Pages  
(CBA - Association rules based classification)  
<http://www.cs.uic.edu/~liub/WebMiningBook.html>

六、教學進度(Syllabi)

2010/2/24	Introduction(簡介) Text Mining Projects perl install	王經篤
2010/3/3	Text Patterns	王經篤
2010/3/10	Quantitative Text Summaries (+Agents for Web Download)	王經篤
2010/3/17	Quantitative Text Summaries (Zipf's Law)	王經篤
2010/3/24	Probability and Text Sampling	王經篤
2010/3/31	Probability and Text Sampling (Bag-of-Words)	王經篤
2010/4/7	Project Announcement (Auto Web Page Download)	王經篤
2010/4/14	Project 1 presentation: Quantitative and Probability	王經篤
2010/4/21	Project 1 Report	王經篤
2010/4/28	Information Retrieval to Text Mining(Term-Document Matrix)	王經篤
2010/5/5	Information Retrieval to Text Mining(Document Similarity)	王經篤
2010/5/12	Text Classification and Clustering	王經篤
2010/5/19	Project 2 Presentation	王經篤
2010/5/26	Project 2 Report	
2010/6/2	Trend Analysis	
2010/6/9	Project 3 presentation	王經篤
2010/6/23	Project 3 Report	王經篤

七、評量方式(Evaluation)

(1) English Text Processing&Statistics(20%) (2) Chinese Text Processing&Statistics(20%) (3) Text Vectorization(20%) (4) Text Classification and Clustering(40%)

八、講義位址(http://)

[http://dns2.asia.edu.tw/~jdwang/TeachingCourses/TextMining/98\\_2\\_TextMining.htm](http://dns2.asia.edu.tw/~jdwang/TeachingCourses/TextMining/98_2_TextMining.htm)

九、教育目標

重新查詢