

數位人文專題

課程資訊

- 科號(Course No.)：10620CL 534400
- 科目名稱(Course Title)：數位人文專題(Topics on Digital Humanities)
- 上課時間(Time)：W3W4WN (星期三 10:10-13:00)
- 授課老師(Teacher)：祝平次(ptchu@mx.nthu.edu.tw)
- 學分(Credit)：3
- 教室(Room)：計中電腦教室(2)
- 助教(Teaching Assistant)：許漢祥(中文所博四) (popaloha@gmail.com)
- 辦公室時間：星期一、二下午1:30-3:30或事先約定。
- 辦公室：人社院B518
- 辦公室電話：42742

一、課程說明(Course Description)

近二三十年來，數位科技的進步大大促進了各種不同知識學門在研究上的可能性。由於電腦比對、計算能力可以快速地處理大量資料，使得之前很多光靠人力難以進行的研究議題，可以藉著電腦來完成。在這一波新的數位工具風潮中，台灣人文學科的跟進並沒有比較慢，但在研究工具的開發上則比較遲緩，直到近十年，所謂的數位人文才越來越興盛。本課程的目的，在於讓沒有從事數位人文學習的修課同學有初步的體驗，了解文字、數字、表格和圖表之間的轉換；另外一個重點則在於，利用Google Earth來進行資料整理。前面的一個重點，在於文本分析；後面的一個重點，則在於簡單地理資訊系統(Geographical Information System, GIS)的使用。修習完本課程的同學，可以建立自己的全文檢索、進行初步的文本分析，並將分析結果轉換成數字與圖表，以視覺化分析結果。Google Earth的使用，則可以將空間的思考帶入人文學的學習中，立體化文字資料與空間的關係，並且可以和田野調查與圖片(包括舊照片與舊地圖)結合。

藉由數位工具進行人文學的探勘，我們可以快速地掌握一本書、或幾本書的大致外貌；也可以將由這樣的初步的處理，再進一步把相關的文獻快速地檢索出來，進行更細緻的閱讀。

二、指定用書(Text Books)

- 本課程所使用的數位工具/程式可以在下列網站取得：
 - Freeplane：
<https://www.freeplane.org/wiki/index.php/Home>
 - Notepad++：
<https://notepad-plus-plus.org/zh/>
 - AntConc：
<http://www.laurenceanthony.net/software.html>
 - Corpro庫博中文語料庫-分析工具：
<http://dh.lis.ntu.edu.tw/resource.html>
 - MSExcel (校園授權軟體)
 - Google Earth：
<https://www.google.com/earth/download/gep/agree.html>
 - Excel2Earth：
<https://ceiba.ntu.edu.tw/course/51f7ca/download.html>

三、參考書籍(References)

- 本課程的輔助教學錄影，可以在Youtube透過搜尋檢得。

四、教學方式(Teaching Method)

- 本課程在電腦教室進行，基本上每個單元，會以教師說明、示範，然後同學實做的順序進行。

五、教學進度(Syllabus)

週數	日期	
1	02-21	<ul style="list-style-type: none">• 02-22才開始上課。
2	02-28	<ul style="list-style-type: none">• 和平紀念日。(休假)
3	03-07	<ul style="list-style-type: none">• 課程介紹

		<ul style="list-style-type: none"> 本週課程介紹，會很快地瀏覽這學期會用到數位工具，粗略地介紹它們能夠幫助我們做到的事情。也會介紹一些其它人文學科常用到的數位工具，具體地讓修課同學了解數位人文的可能性。
4	03-14	<ul style="list-style-type: none"> Freeplane 心智圖軟體：示範、說明與實做。 <ul style="list-style-type: none"> 心智圖裏面蘊涵了分析、綜合的兩種研究方式，也幫助我們對於所要分析、論述的課題，可以有一種全景式的掌握。Freeplane除了做樹狀式的展開，也可以重新綜彙分支，進行總結；再繼續展開。對於研究議題的腦力激盪、議題整理都很有幫助。 在這一節課程裏，同學將被要求一個多功能的心智圖，能夠呈現出資料的結構、分析綜合的關係、內部以及外部的超連結，並能夠利用心智圖進行類似報告。 標記語言（一）：認識網頁的真相。 <ul style="list-style-type: none"> 標記語言，是目前處理數位資料幾種最常用的方式之一。本節課程的目的，一方面讓同學對網頁的超文本格式(HyperText Markup Language, HTML)有一基本的認識，一方面也藉由實做，讓同學了解網頁連結的機制。 在這一節課程裏，同學將被要求一個資料來裏，做出五個互相連結的網頁。
5	03-21	<ul style="list-style-type: none"> NotePad++：說明與示範。 <ul style="list-style-type: none"> Notepad++是一款非常受歡迎的文字編輯軟體，我們將利用它來學會 <ol style="list-style-type: none"> 1)如何利用Notepad++來轉換中文檔案的內碼， 2)如何建立自己的全文檢索， 3)利用Markdown表示法來格式化文字， 4)進行「規則運算式」(Regular Expression)的複雜檢索， 5)利用「規則運算式」的取代功能來清理資料。 概念：迴避字元(escape character)。
6	03-28	<ul style="list-style-type: none"> NotePad++：實做。 <ul style="list-style-type: none"> 在這一節課程裏，同學將被要求完成說明示範時所教授的各種功能，以具備清理數位文本的能力，以便進行進一步的文本分析。 並且進一步利用Excel來統計分析處理過後的資料。
7	04-04	<ul style="list-style-type: none"> 兒童節。（休假）
8	04-11	<ul style="list-style-type: none"> CText.org線上平台與數位分析工具（一） <ul style="list-style-type: none"> 資料比對是文史工作者必須常常面對的問題、從事的工作，也與我們種種對文本的判斷有所關聯。而電腦的特長正在於快速的資料比對，全文檢索就是一個很好的例子。在這兩週的課程中，我們將利用Notepad++處理資料，然後利用CText.org來統計、分析資料，並將統計分析的結果視覺化為圖表或網絡關係。視覺化不但會影響到我們的研究視角，也是教學上的利器，更是文史科系與其它學科進行交流最好的橋樑。而如何藉由處理資料、分析資料、統計資料，完成從文字到數字、表格與圖表的轉換，正是本次工作坊要完成的目標。 概念：N-Gram、文本異同比較。
9	04-18	<ul style="list-style-type: none"> CText.org線上平台與數位分析工具（二） <ul style="list-style-type: none"> 這個禮拜我們將實做CText.org平台的數位工具。
10	04-25	<ul style="list-style-type: none"> 期中測試
11	05-02	<ul style="list-style-type: none"> CBDB說明與示範。 <ul style="list-style-type: none"> CBDB (China Biographical Database 中國歷代人物傳記資料庫)由美國哈佛大學費正清中國研究中心、台灣中央研究院歷史語言研究所及北京大學中國古代史研究中心三個單位合作開發。目前收有約35萬筆的人物資料，主要集中在唐到清的人物。藉由這個資料庫，可以進行集體傳記學、地理空間、社會網絡等不同型態的研究。從本週開始，連續三週，我們一方面要藉著CBDB了解關聯式資料庫的原理，一方面也要熟悉CBDB的基本操作，以便在自己的研究中，可以利用它。 相關連結：https://projects.iq.harvard.edu/chinesebdb/home
12	05-09	<ul style="list-style-type: none"> CBDB實做練習。 <ul style="list-style-type: none"> 本週我們利用幾個題目來熟悉CBDB的操作。

13	05-16	<ul style="list-style-type: none"> • CBDB實做測試： <ul style="list-style-type: none"> ◦ 本週將由同學展現利用同學實做幾個測試的題目，來驗收大家對於CBDB的熟悉程度。並同有一個研究題目的測試，亦即設計一個適合利用CBDB來進行研究的題目。 ◦ 研究題鉤設計：請想出一個適合利用CBDB來進行研究的題目；並說明這個題目是什麼，以及如何利用CBDB來進行研究。並描述可能的研究成果。
14	05-23	<ul style="list-style-type: none"> • Corpro庫博文本分析案例（闕河嘉） <ul style="list-style-type: none"> ◦ 本節將請台灣大學闕河嘉老師介紹她利用文本分析所做出來的研究例，以讓同學了解文本分析在現今社會可能的研究對象。 ◦ 同學將被要求，完成一分瀏覽網頁後，所設想出來的研究題目。
15	05-30	<ul style="list-style-type: none"> • Corpro庫博中文語料庫分析工具。（闕河嘉） <ul style="list-style-type: none"> ◦ 本節將請台灣大學闕河嘉老師介紹她的團隊研發出來的語料庫分析工具—Corpro庫博。目前媒體喜歡解析各大國領袖重要談話的即時語頻分析，即是利用類似工具所做出來的結果。庫博則具有很多為台灣客製化的功能，比起AntConc更適合在台灣同學使用。 ◦ 同學將被要求利用Corpro來完成上週所設想出來的研究題目，以做為期末報告的主要內容之一。
16	06-06	<ul style="list-style-type: none"> • 標記語言（二）：TEI/XML。（洪振洲） <ul style="list-style-type: none"> ◦ 標記語言對於人文學科，不管是語言內容的分析、結構的觀察，乃至於田野調查資料的整理都是最常用的手段。之前我們介紹了處理網頁呈現的超文本標記語言HTML，這節我們則將檢視做為網路資料庫最常見的可擴展標記語言(eXensible Markup Language, XML)中的一種西方人文學科標準 TEI (Text Encoding Initiative)。我們將請到長期主持法鼓山中華電子佛典CBETA 標記的洪振洲老師來為我們講解 TEI。 ◦ 同學將被要求使用以TEI為文本標記建置起來CBETA數位研究平台 (http://cbeta-tp.dila.edu.tw/)，並寫成觀察報告500字以上。
17	06-13	<ul style="list-style-type: none"> • 標記語言（二）：TEI/XML的實做。 <ul style="list-style-type: none"> ◦ 在這一節課，我們將藉由oXygen Editor的編輯軟體來實做一個TEI檔案。 ◦ 同學將被要求，完成一個TEI標記檔案。
18	06-20	<ul style="list-style-type: none"> • 期末發表會。 <ul style="list-style-type: none"> ◦ 本次發表會將以同學所想到的文本分析計畫為主。若同學能夠再這個過程中，利用本學期所學到的其它技能來增加內容的豐富性，將在分數上，給予較優異的考量。

六、成績考核(Evaluation)

平時成績(50%)：含課堂報告、討論、期中測試。

期末報告(50%)：含期末發表及書面報告。

出缺席另計，缺席三次者成績降一級（如由B+變成B，以下同），四次降兩級，五次降三級，缺席達三分之一者、作業報告抄襲者學期成績以E級分計。若因故缺席，可在授課教授指導下，進行與授課內容相關的修補作業，修補作業通過者可以抵銷缺席記錄。本科報告，除非經過授課教師書面同意，不得將同一分報告做為本課程與其它課程的共同作業使用。

七、可連結之網頁位址

- <http://tinyurl.com/pingtzuchu>
- <http://lms.nthu.edu.tw/course/>
- <https://sites.google.com/a/ptc.cl.nthu.edu.tw/courses/my-forms>（本網頁的子網頁有和一般學習相關的資料，以及和研究論文學作相關的簡報檔。）