

文科生在資訊時代的挑戰與機會

2015/12/7 東海大學中文系

中華開放古籍協會

葉健欣

大綱

- 自介
- 數位科技和數位內容的性質
- 近二十年來中文資料庫發展的過程
- 文科生有優勢的資訊相關領域

自介

馬來西亞華裔，1991年來台。

曾參與之計劃：

佛光大辭典、漢文大藏經 1995年~

中研院漢籍全文資料庫、印順法師佛學著作 1999年

朱邦復中文電子書計劃 2001年

巴利文大藏經 2006年

康熙字典、說文解字注數位化 2008年

藏文大藏經 2013年~

數位科技和數位內容的性質

未來只有兩種工作可高枕無憂：

一) 指揮、配合電腦工作。

二) 電腦無法取代的工作。

必須了解數位資訊科技(Digital Information Technology), 即數化(Digitization)、數位內容(Digital Content)、數位處理(Digital processing)的性質。

掌握基本性質，理解及預測變化

資訊科技日新月異，學不勝學。

資訊科技的分工越來越細，專業人士也面臨選擇的困難。

以不變應萬變，萬變不離其宗。

資訊科技的「宗」，即基本性質。

資訊設備的三大部份

輸入：將自然界的資訊轉為數位內容/訊號。(數位化)

鍵盤、滑鼠、手勢、麥克風、相機、掃瞄器、三軸陀螺儀...

處理：計算機，軟體程式。

輸出：將數位內容/訊號，轉為人類可以讀取的形式。

螢幕、喇叭、印表機、機器人...

數位化的特質

數位化：從物質媒材到能量媒材 (從原子到電子)

- 一) 非破壞性。不失真 (例子：影印機)
- 二) 資訊增加，重量不變。(無篇幅之限制)
- 三) 複製的邊際成本接近零 (例子：偷書及盜拷)

在數位世界，何為賤？何為貴？

物以稀為貴。

注意力。

在數位世界，時間和空間的隔離越來越小，什麼變遙遠了？

資訊發展的幾個趨勢

一) 摩爾定律。每18個月晶片上的電晶體密度加倍。

二) 資訊服務雲端化。(自來水喻)

三) 設備行動化。

和你賽跑的不是人

一) 電腦擅長什麼？(回顧：工業革命機械取代了什麼？)

只要電腦開始能做，就會越來越便宜(摩爾定律)

避免進入電腦「即將」能做的領域。

二) 電腦不擅長什麼？(電腦能開車嗎？)

電腦也是一種機械，它是腦力的延伸。

近二十年來中文資料庫發展的過程

中央研究院漢籍電子文獻。1984~ 特色：正確、權威

CBETA 中華電子佛典協會。1998~ 特色：提供離線閱讀

海外

CTEXT 中國哲學書電子化計劃。2006~ 特色：「相似段落」等新技術

[開放的數位古籍平台](#) 2008

漢文資料庫的發展方向

一)更精密的資料:Diplomatic edition, Parallel-corpus, Critical edition

二)結合計算語言學(Computational Linguistic)的各種工具。機械翻譯。

三)更完整的資料。

以下台灣或有優勢。

四)人文與科技協作。

五)新概念的雛型製作。如[校勘平台](#)

兩個預測

一) 還原創作情境。

由於沒有篇幅的限制，創作之際所引用的所有文獻，在讀者端重現。

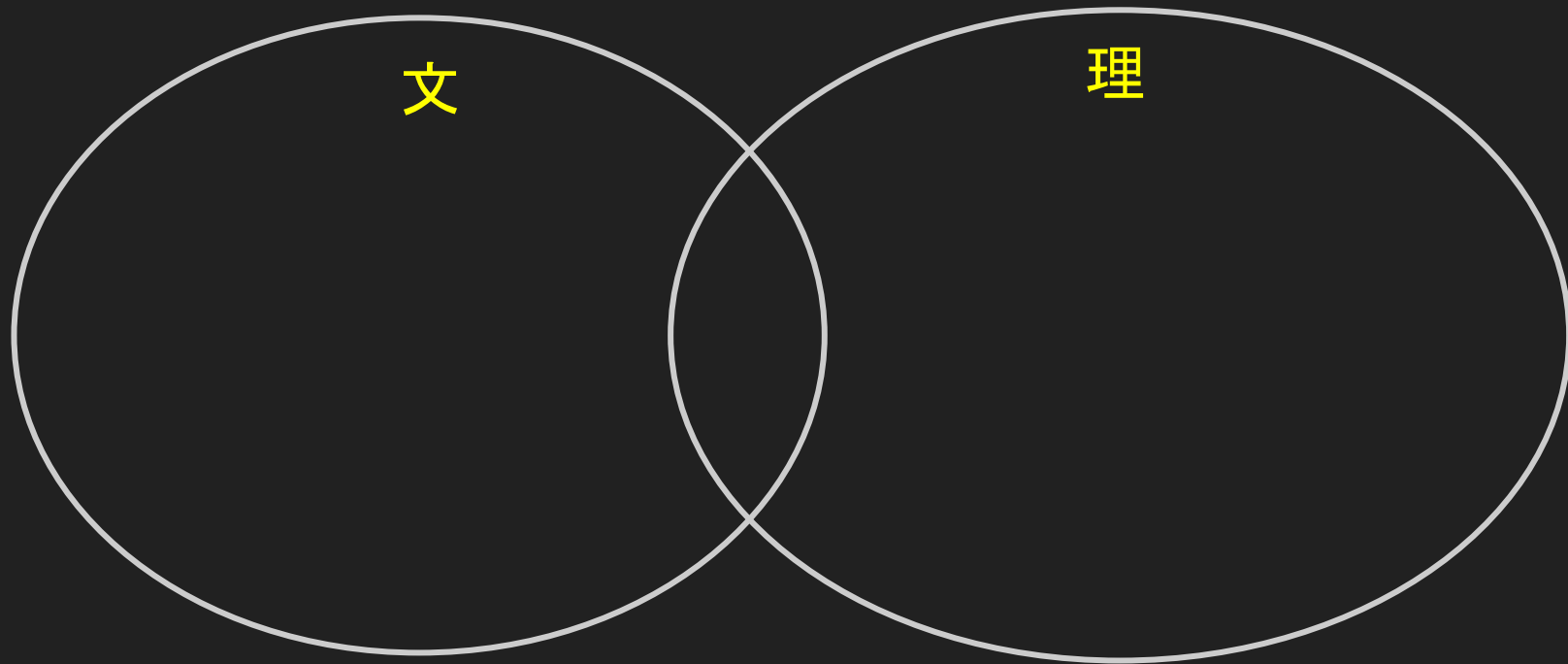
二) 整合讀者心得。

A) 大量的古書文字輸入電腦之後，接下來需要輸入文字之間的「關係」。

例如：看到一段古文，電腦給出古代相關的註解。

B) 今人讀書心得筆記注解的自動合併。

文科生有優勢的資訊相關領域



文武雙全

- 一) 選修計算機概論。掌握基本詞彙。
- 二) 學習一種程式語言。推薦 Javascript。(三個月)。
- 三) 和資訊相關科系的朋友協作。

可能的方向

一) APP 創業, 組隊打Boss。(語言教學)

二) 網路講書。(羅輯思惟)

三) 寫故事, 寫文案...

跨界合作例子: 臉部辨識 + 領養 ([Twin Souls](#))

創造屬於你們時代的第一

新的工作不斷的產生，舊工作不斷消亡。

挑戰不斷而機會無窮。

致力於「代表作」，打造品牌及信譽。

謝謝

yapcheahshen@gmail.com