



檔案樂活情報
Archives LOHAS

書籤分享：[f](#) [t](#) [p](#) [g](#)

No.102 發刊日:104年12月16日



臺灣瘧疾根除半世紀



檔案知識+

威尼斯時光機器數位典藏計畫簡介



© creative commons: Veduta del Palazzo Ducale di Venezia

圖片來源：Venice Time Machine網站

本計畫由瑞士洛桑聯邦理工學院與義大利威尼斯卡福斯卡里大學發起，希望將圖書、檔案及地圖等這些獨特的歷史資源數位化，轉化成為虛擬的時光機器，就像是一個有時間軸的Google地圖，讓我們可用數位的方式體驗，並探索這一千年來威尼斯的人文及社會風貌。

國家發展委員會檔案管理局檔案典藏組助理員 蔣佳蓉

壹、計畫簡介

威尼斯時光機器 (Venice Time Machine) 是歐洲近年的跨國數位典藏計畫，由隆奧基金會 (Fondation Lombard Odier) 資助，瑞士洛桑聯邦理工學院 (École Polytechnique Fédérale de Lausanne, EPFL) 與義大利威尼斯卡福斯卡里大學 (Università Ca' Foscari Venezia) 共同執行，計畫主持人是洛桑聯邦理工學院數位人文實驗室 (Digital Humanities Laboratory) 的創辦人卡普蘭教授 (Prof. Dr. Frédéric Kaplan)，他先於巴黎高科電信學院 (Ecole Nationale Supérieure des Telecommunications) 獲得工程學位，之後在巴黎第六大學 (Paris VI) 獲得人工智慧博士學位；曾於巴黎幾所大學進行有關機器人技術與人工智慧的研究，也曾於Sony的電腦科學實驗室工作過10年之久。

卡普蘭教授指出，關於過去10年，我們擁有大量的資訊，但10年前甚至越往前，我們所擁有的資訊就越少。如果要建立可以回到過去一千年的Google地圖、Facebook或社群網路等，就必須擴大所擁有的資訊。除了將現有的檔案數位化，還需透過歷史學的推斷 (extrapolate) 或電腦科學所稱的模擬 (simulation)，加以擴展手邊的資料。例如，一本航海日誌，我們可以不只把它看成是一位威尼斯船長的特定航程，還可以把它視為當時普遍的航程模式；又，如果有一幅建築外觀的繪畫，除了把它看成特定的建築，還可推論這棟建築可能與其它資料已佚失的建築有著相同的風格。

本計畫係與威尼斯檔案館 (Archivio di Stato di Venezia) 合作，致力於建立一個模型以呈現威尼斯過去一千餘年發展歷程與多面向的風貌，以及一個大型開放資料庫可用於研究與教學。威尼斯檔案館典藏大量手稿，其中使用的語言及方言的年代橫跨中世紀至20世紀，一千年來的行政文書檔案計約80公里長，內容涵蓋威尼斯生活的每個面向，包括出生登記、死亡證明、稅務狀況、地圖及都市計畫等等。本計畫旨在將這些數十公里長的檔案予以數位化、轉錄、建立索引，以利外界搜尋，成為威尼斯人文史上最大資料庫的基礎。另外，從這些資料萃取出的資訊將進一步組織成相連的語意圖 (semantic graph)，在歷史地理資訊系統中呈現時間與空間。

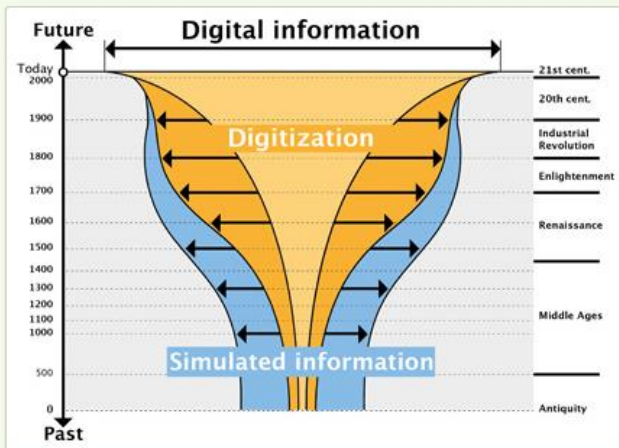


圖1：模擬資訊後的資訊蘑菇 (information mushroom)
圖片來源：Frédéric Kaplan網站



圖2：威尼斯檔案館一景
圖片來源：Venice Time Machine網站

貳、作業流程

本數位典藏計畫實為具有自動文字識別功能的大規模數位化作業，作業流程包括掃描、轉錄、文本處理以及資料連結，說明如下：

一、掃描 (scanning)

將紙本文件掃描成高解析度的數位影像，不同類型的文件在使用掃描器時有不同的限制及速度，目標是1天完成450冊圖書掃描。洛桑聯邦理工學院正與業界合作研發一種每小時可以掃描1,000頁的半自動機器人掃描設備，結合前述設備，將可創造因應古代文獻的高效數位化流程。另外，也探索利用粒子加速器製造的X光同步輻射，希望可以不用翻頁就能掃描整本書。

二、轉錄 (transcription)

所謂轉錄，是將不同形式的資料轉化為文字；手寫檔案的圖像複雜度及多樣性使轉錄作業相當困難，其中又包含了拉丁文、托斯卡尼、威尼斯等各種語言及方言。目前正在開發將圖像轉化為可能的文字的新演算法，即將圖像自動拆解成一個個可能為單字的小圖像，每個小圖像將與另一個小圖像比對，並根據其可能代表的單字來分類。再配合語言模型、文件模型以及機器視覺辦法，使此步驟可以部分自動化，至少能辨識出一些有標準格式的官方文件。每當有一個新的字被轉錄出來時，便可讓資料庫中的一些其他被歸屬在同類的字也被辨識出來。

三、文本處理 (text processing)

文本處理是將這些辨識出的文字利用文本處理器變成一個個句子，以便汲取該文件中的資訊。估計從這些檔案中能汲取到1百億個事件。

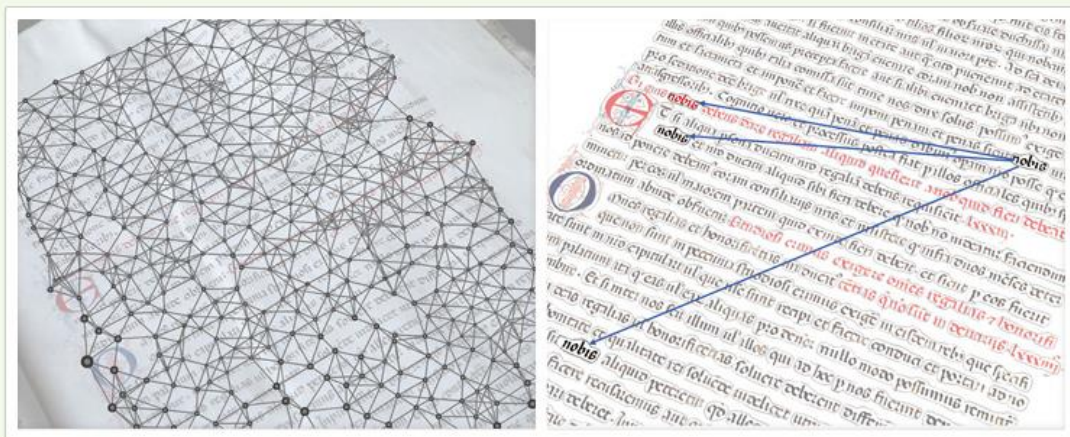


圖3：分析圖像結構，以及辨識出類似的小圖像
圖片來源：Venice Time Machine網站

四、資料連結 (connecting data)

這些檔案的最大價值在於其間的關聯性。幾個關鍵字連結了不同類型的檔案，讓資料變得可以搜尋。這些大量資料間透過關鍵字而可以進行交互對照，組織出了巨大的關聯圖，因此讓新面向的資訊得以浮現。

參、結語

在西方史料中，像威尼斯檔案這樣既為數可觀，且又具備多樣性及準確性的情形相當少見，本計畫希望創建類似具有時間軸的Google地圖，透過操作簡易的模型，以視覺化方式呈現並可探索威尼斯的變遷，令人不僅能在其中穿越空間，更能在其中穿越時間，更加認識威尼斯的過往。同時佐以建立Facebook，用來發布古時候威尼斯居民的生活，例如「西元1434年里亞托（Rialto）市場裡一條鯛魚的價格」或「1544年在穆拉諾島（Murano）玻璃工匠的薪水」等資訊，增加能見度。開放的數位化檔案將提供全新的研究環境，若再將這些檔案產生出的關聯放入當時歐洲的情境之中，將有機會重現一千年來威尼斯在地中海沿岸的繁華與衰敗。

卡普蘭教授認為，威尼斯時光機器是以長期的觀點來看待資訊，利用科技融合歷史，甚具價值，未來應訓練歷史學家利用這項科技來重建其他地區，並培育數位人文學家（digital humanist）利用大數據進行研究，以迎接人文學典範轉移（paradigm shift）時代的來臨。

參考資料：

1. Venice Time Machine. <http://vtm.epfl.ch/page-109337.html> (accessed 22 Oct, 2015)
2. Frederic Kaplan. <http://www.linkedin.com/in/frederickaplan> (accessed 22 Oct, 2015)
3. Lancement de la Venice Time Machine. <http://fkaplan.wordpress.com/2013/03/14/lancement-de-la-venice-time-machine/> (accessed 23 Oct, 2015)
4. Frederic Kaplan: How I built an information time machine. <http://www.youtube.com/watch?v=2-Ev4rU27HY> (accessed 23 Oct, 2015)

本文檔案下載：[知識No102.pdf](#)

本頁最後更新日期：104.12.16

本頁點閱瀏覽次數：7

102期專區	資訊專區	會員專區	下載專區
檔案瑰寶 檔案知識 ⁺ 檔案搶先報 回本期首頁	徵稿訊息 各期電子報查詢 回最新一期首頁	訂閱電子報 取消電子報	簡報及桌布下載 精華版下載
歡迎您對檔案樂活情報提出寶貴建議，請聯絡： alohas@archives.gov.tw			

[列印](#) [上一頁](#) [回頂端](#)

認識我們

- 本局概況
- 局長
- 施政計畫與成果
- 大事紀
- 顧客服務白皮書
- 出版品及報告
- 聯絡我們
- 政府資訊公開

國家檔案

- 國家檔案管理制度
- 典藏概要及精華檔案
- 國家檔案應用
- 國家檔案保存維護中心
- 檔案樂活情報
- 檔案支援教學網

機關服務

- 文書服務
- 機關檔案管理
- 金檔獎暨金質獎
- 教育訓練
- 諮詢溝通

便民服務

- 檔案申辦服務
- 參訪申請
- 電子檔案技術服務
- 過時資訊設備捐贈
- 應用系統
- 訂閱電子報
- 徵稿訊息
- 諮詢溝通
- 志工團地
- 相關網站

文檔法規

- 文檔法規架構圖
- 檔案法(含解釋函)
- 相關子法
- 作業規定
- 相關法令網站



[隱私權、著作權及資訊安全政策宣告](#) | [政府資訊公開](#) | [徵稿訊息](#) | [訂閱電子報](#)

24220 新北市新莊區中平路439號(北棟)9樓
本網站最佳瀏覽解析度為 1024*768

總機：(02)8995-3700

傳真：(02)8995-6469

