「工人智慧」到「人工智慧」

-從中文資料庫數據軌跡看台灣學術研究發展新趨勢

2023/11/16

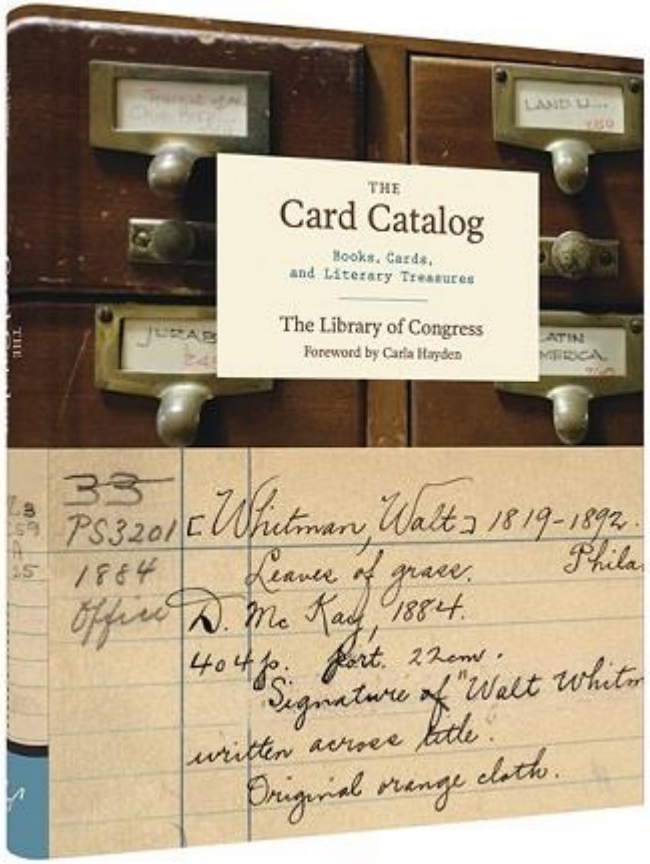
OUTLINE

- 圖書館服務的演變
- 中文資料庫的數據軌跡
- 人工智慧的影響與技術應用未來展望


圖書館服務的演變

請以50年為一個階段，說明從有圖書館到現在，在不同的時間點中，圖書館扮演什麼樣的角色？提供什麼樣的服務？

ChatGPT can make mistakes. Verify important information.



Source: www.chroniclebooks.com/products/the-card-catalog

1970年代至1990年代：

1. 資源中心：

   主要作為紙本書籍和期刊的存儲和借閱中心。

2. 傳統參考服務：

   圖書館員提供參考服務，幫助讀者查找信息。

3. 圖書館卡目錄：

   讀者使用卡目錄來查找圖書館的收藏。


Source: forbeslibrary.org/help/use-a-library-computer/

1990年代至2000年代：

1. 數位轉型開始：

   引入電子資源，如電子期刊和數位資料庫。

2. 網絡接入：

   提供公共互聯網接入，幫助讀者瀏覽網絡。

3. 教育和培訓：

   圖書館提供培訓，幫助讀者利用電子資源。

Source: world.edu/5-reasons-digital-learning-future-education/

2000年代至2020年代：

1. 電子資源擴展：

   數位內容變得更為豐富，包括電子書和多媒體資源。

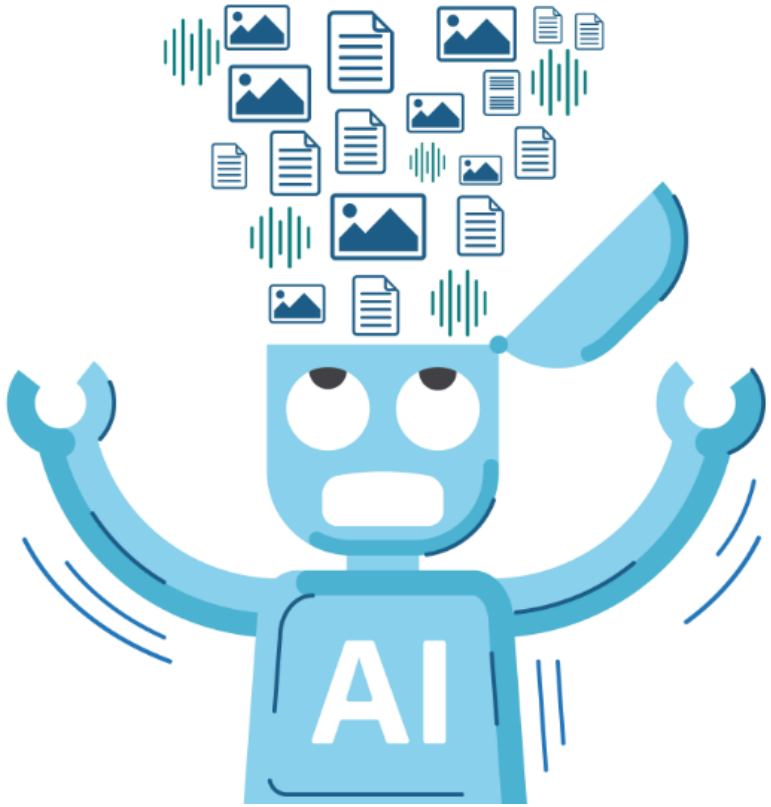
2. 資源管理系統：

   引入自動化圖書館管理系統，提高效率。

3. 社區參與：

   舉辦活動、講座和工作坊，增進社區參與。

2020年代至今：



Source: DEANLONG.io

1. 遠程服務：

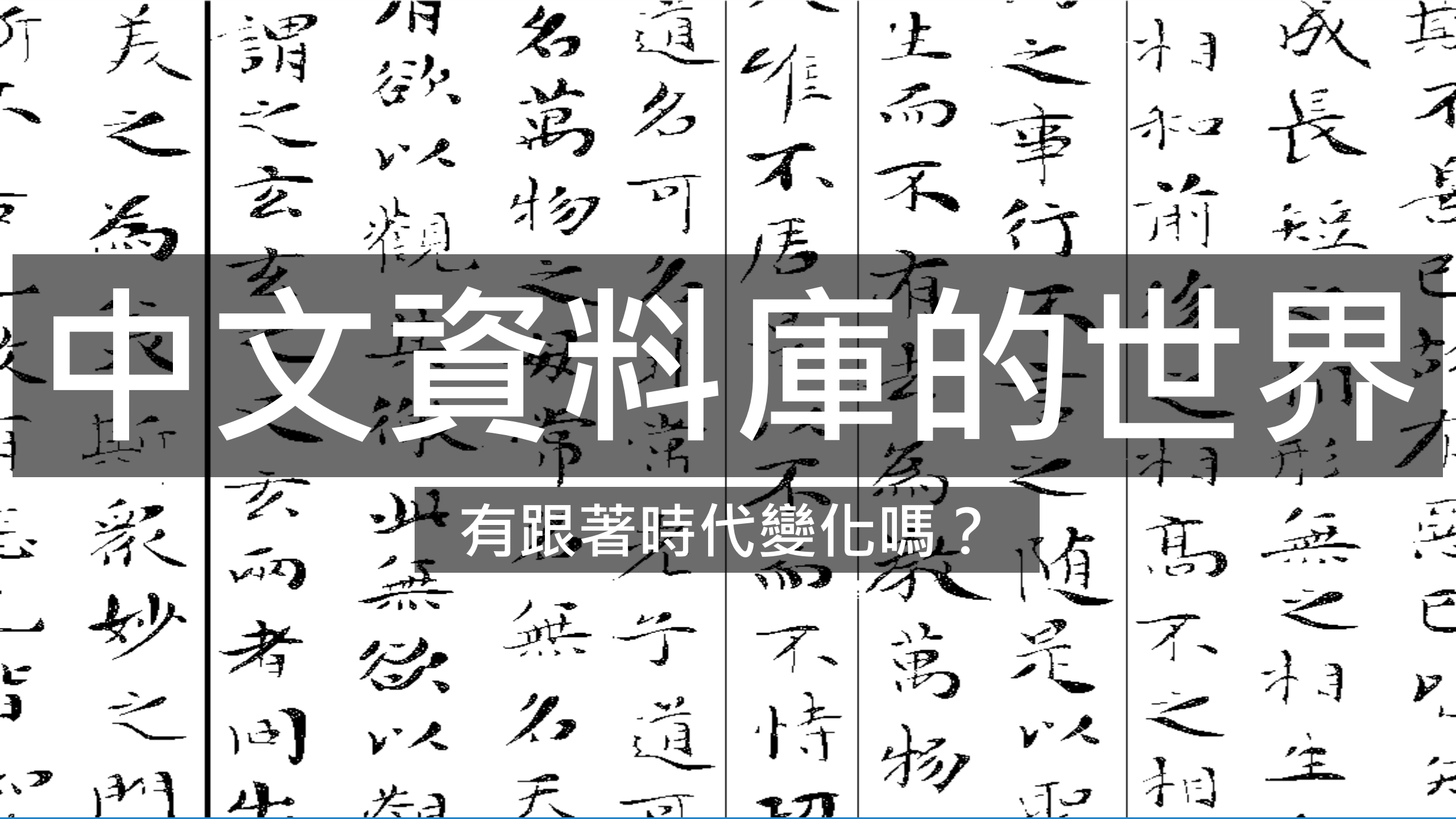
   強調虛擬參考服務和數位資源的遠程訪問。

2. 開放數位內容：

   提供開放存取資源，如學術期刊和教育課程。

3. 創新技術：

   採用AI、虛擬現實等技術改善服務，如自助借書機和智能圖書推薦。



中文資料庫的世界

有跟著時代變化嗎？


airiti Library 華藝線上圖書館
Language
華藝數位股份有限公司，您好！
瀏覽
進階檢索
登入 | 個人化服務 | 客服中心 | 使用說明 | 網站地圖
一般檢索 | 全文檢索 | 出版品檢索
所有欄位
更多選項
查詢歷史
2022年度最佳傳播獎獲獎論文
訊息動態
2022年度最佳傳播獎，獎...
Airiti學界快訊：帶著...
Airiti學界快訊：IN...
Airiti學界快訊：DO...
Airiti學界快訊：道別...
>>更多
推薦閱讀
預刊文章
最新上線
精選期刊
學界快訊
研討會訊息
11/18-11/18 2023 台灣資訊社會研究學會年會暨...
我要張貼訊息
華藝DOI熱門解析期刊
護理雜誌
寵物治療在護理專業領域之臨床應用

一些來自airiti Library華藝線上圖書館的觀察：

2005 2006 2007 2008 2009 ..... 2021 2022 2023

資料來源

研究主題

閱讀行為

搜尋習慣

人工智慧

的影響與技術應用未來展望

近年來，隨著人工智慧的發展，在各式各樣的產業中，都可以看到有趣的變化在持續發生。人工智慧可以幫助我們更好地理解使用者需求，提供個性化建議，並加速數據處理過程。

理解資料查找需求：Learning to Rank

"Learning to Rank"（LTR）是一種機器學習技術，它用於優化搜尋結果的排序，以便提供更有用和相關的結果給用戶。這可以用以下簡單的例子來解釋：

假設你是一家電子商務網站的搜索引擎工程師，用戶在你的網站上搜索 "智能手機"。當一個用戶輸入這個搜索詞時，你的搜索引擎必須決定如何排列顯示結果。你有數百個手機產品，並且希望用戶看到最有可能符合他們需求的產品。

如果過去用戶更常點擊某個品牌的手機，或者更常在價格範圍內購買手機，這些信息都可以被 "Learning to Rank" 模型納入考慮。然後，當新的用戶進行搜索時，這個模型會根據這些因素重新排列結果，使最有可能符合用戶需求的產品出現在前面。

簡而言之，"Learning to Rank" 是一種利用機器學習來優化搜索結果排序的技術，以確保用戶更容易找到他們想要的內容或產品。这个模型通过分析大量的用戶數據，學習如何以最有效的方式呈現結果，從而提高了搜索引擎的質量和用戶滿意度。

提供個性化建議：Deep Dive Discovery

“Deep Dive Discovery” (DDD) 是一項AI工具，旨在協助使用者更有效地尋找他們需要的學術資訊。它包含兩個關鍵部分，讓我們用簡單的例子來解釋：

- 關聯詞推薦：當你想查找特定主題的學術文獻時，通常會輸入一些關鍵詞。假設你對環境保護感興趣，所以你輸入了 “環境保護” 這個關鍵詞。 DDD 會分析大量的學術文本並做「詞間距計算」， 然後推薦你可以添加到你的查詢中的其他主題詞。這樣一來，你的查詢變得更具針對性，幫助你更好地縮小搜索範圍，就像在網上商店中選購時，系統會建議相關的商品一樣。

- 延伸閱讀推薦：當你開始閱讀一篇文章時，可能會想了解更多相關信息。舉個例子，你正在閱讀一篇有關氣候變化的文章，你可能會想看看其他相關的研究。 DDD 使用人工智慧模型，理解你正在閱讀的文章以及不同文章之間的關聯。然後，它會推薦給你其他五篇它認為值得繼續深入閱讀的文章。這就像當你在看一本書時，書店店員會告訴你其他可能會喜歡的書籍。

總之，DDD 通過利用AI對大量文獻的理解，幫助你更輕鬆地找到你需要的學術知識，不僅提供更具針對性的查詢建議，還通過推薦相關文章，讓你更深入地了解你感興趣的主題。

文獻閱讀：Document Image Understanging

“Document Image Understanding”是華藝因應機器中文文本閱讀兩大困難(排版格式與閱讀順序)而開發的解決方案，它包含了三個主要部分：Document Layout Analysis, Reading Order Detection, 和Hybrid OCR。

與大多數的英文文件不同，中文的檔案、期刊、論文、書籍都存在著非常多樣性的排版格式，許多早期文獻會大量使用對電腦來說較難以理解的直書。這樣的排版特性，造成在訓練AI進行中文文獻理解時，閱讀順序的問題也會因應而生。

為了解決這個眼前的大石頭，我們將問題拆解，優先處理文件版面的分析，將整個文件圖像分為小塊，並未每個區塊分配類型標籤，如：表格、圖片、或章節標題。再透過閱讀順序檢測，識別每個區塊的閱讀方向，與所有區塊之前的合理順序。最後，出於”排版具其意義”的信念，建立自己的混合式OCR模型，在數位化過程中保留盡可能多的資訊，如：字體及座標，來維持盡量完整的資訊轉譯。