

國立台灣大學電機資訊學院資訊網路與多媒體研究所

碩士論文

Graduate Institute of Networking and Multimedia

College of Electrical Engineering and Computer Science

National Taiwan University


Master Thesis

影片字幕檢索系統

以臺大文學講座系列影片為例

Retrieval System for Video Subtitles

with Videos of Literature Seminar at NTU



傅泓翊

Fu Hung-I

指導教授：項潔 教授

Advisor: Jieh Hsiang, Professor

中華民國 101 年 7 月

July, 2012

誌謝

從進入臺大校園開始，也已經過了四個年頭，在研究所這個階段，受到了許多人的指導及幫助，才能將學業完成，感謝你們。

首先感謝項潔教授的教導，精準的想法及思考邏輯，是我學習到最多的地方。感謝老師在研究所生涯給予的建議及指導，不斷的導正我的方向，教導我正確的觀念，而使我最終能將論文完成。感謝杜老大、詩沛學姊、浩洋學長、筱盈學姊，給我學業上的幫助實在太多了，無論什麼情況下都願意給我指點及意見，讓我能夠克服一路上的困難直到完成。

感謝已畢業的屹靈、國延、家慶、承恩、思靜、于鳴、小玉、雋凡、嘉翔等學長姊，你們在課業上的建議還有實驗室事務的幫助，讓我在實驗室中是輕鬆快樂的。感謝唯一同屆的同學韋翰，在實驗室能夠互相激勵。感謝豐恩和稷安，你們的專業知識總是給實驗室很多的衝擊。感謝博士班的宋浩、農堯學長，在後來的研究所階段也給我引導及幫助。感謝一同畢業的嘉文、鈺淳、柏淳、士綱，做好許多考試的準備，使得整個考試順利。

感謝我的父母，讓我任性的在學校裡待了這麼久的一段時間，給我在求學過程中的所有支持。感謝大頭、rainmann、夙吟、小當、卡以及陪在我身邊的 Emma，沒有你們的幫忙本論文不會完成。

感謝你、妳，以及你們。

摘要

一般在使用影音光碟時，只能按照章節來觀看，而不能對影片內容作檢索，來找到想要看的片段。於是我們建立一個影片字幕檢索系統，希望能對影片內容做一些搜尋，使用的影片是臺大文學講座系列影片。

臺大文學講座系列影片為臺大出版中心將近代文學的作家，如白先勇、葉維廉、葉嘉瑩、高行健...等，於臺大演講的情況錄製成影音光碟，主要內容為大師們文學創作的經歷，以及對文學、美學的想法。此系列光碟大部分含有演講手冊，為了讓使用者在看到演講手冊中有興趣的部分時，能快速找到影片中的該片段，而設計了此套字幕檢索系統。由於影片內容皆是演講，因此對字幕檢索也就是對影片內容做檢索。

我們首先利用 esrXP 取出包含字幕的圖片，並利用 Microsoft Office Document Imaging 中的 OCR 功能來辨識字幕圖片，將辨識結果送回 esrXP 製作成字幕檔，來取得字幕文字與字幕時間；並且利用最長共同子序列計算字幕與演講手冊句子的相似度，來知道字幕與句子的對應關係，進而得到字幕的發言者以及字幕對應到的演講手冊句子。

接著建立一個網站系統，利用 HTML5 的 video 標籤，讓使用者只要使用支援 HTML5 的瀏覽器即可觀看影片；在搜尋字幕以及觀看影片的時候，也可以看到當下字幕所對應的演講手冊句子，而給予使用者更多資訊。另外，我們還引入多維度的後分類導覽方式，幫助使用者能對搜索結果做更進一步的篩選。

關鍵字：臺大文學講座、字幕檢索、影音檢索、資料庫、數位人文

Abstract

When we watch videos with video discs like DVDs or VCDs, we can only watch by chapters. We cannot do some search on the content of video. So we provide a retrieval system for video subtitles, and hopefully do some progress on searching the video content.

NTU Literary Lecture Series published by National Taiwan University Press are videos of speech giving by some modern literature writers in Taiwan. There are videos on DVDs and a speech manual for every video in NTU Literary Lecture Series. People may read speech manuals to scan the content of videos quickly. When people find an interesting paragraph and want to watch the part of video, they cannot easily do that.

To solve this problem, we create the subtitle files of videos by esrXP which captures pictures of subtitles and Microsoft Office Document Imaging which does OCR on pictures to get the text of subtitles. Additionally, we match subtitles to the sentences in speech manual for giving more information to users.

Then we access videos through web. By using video tag of HTML5 on webpage, users can easily watch the videos without any plug-in if they use HTML5-supported browsers like Google Chrome and Mozilla Firefox. When users watch videos, the sentences correspond to the subtitle will be displayed below the player. It will provide more information to users on selecting subtitles. We also provide the function of post-classification to users for filtering the retrieval results.

Keywords: NTU Literary Lecture Series, subtitle retrieval, text match, database, digital humanity

目 錄

誌謝.....	i
中文摘要.....	ii
英文摘要.....	iii
目錄.....	iv
圖目錄.....	v
表目錄.....	vi
第一章 緒論.....	1
1.1 研究背景.....	1
1.2 研究對象及動機.....	4
1.3 研究目的.....	5
1.4 研究方法與流程.....	6
1.5 論文架構.....	8
第二章 擷取影片字幕.....	9
2.1 使用工具介紹.....	9
2.2 材料分析.....	14
2.3 擷取流程.....	15
第三章 字幕與演講手冊的對應.....	17
3.1 演講手冊說明.....	17
3.2 對應方法.....	18
3.3 使用此對應方法的理由.....	26
第四章 系統實作與功能.....	28
4.1 系統概述.....	28
4.2 系統功能與操作實例.....	31
第五章 結論與未來工作.....	36
5.1 結論.....	36
5.2 未來工作.....	37
參考文獻.....	38
附錄一 臺大文學講座影音光碟列表.....	39
附錄二 MODI 常見錯字取代列表.....	40
附錄三 資料處理流程.....	43

圖目錄

圖 1-1 研究流程圖.....	6
圖 2-1 esrXP 過濾畫面.....	10
圖 2-2 esrXP 過濾器設定.....	10
圖 2-3 字幕圖片.....	11
圖 2-4 esrXP 主視窗.....	12
圖 2-5 Microsoft Office Document Imaging 畫面.....	13
圖 2-6 字幕與影像畫面獨立.....	14
圖 2-7 字幕與影像畫面重疊.....	15
圖 2-8 擷取字幕流程圖.....	16
圖 2-9 srt 字幕檔格式.....	17
圖 3-1 對應錯誤範例.....	26
圖 4-1 系統架構圖.....	28
圖 4-2 檢索結果畫面.....	31
圖 4-3 檢索結果條目.....	31
圖 4-4 後分類索引.....	32
圖 4-5 後分類階層展開.....	33
圖 4-6 點擊分類反白提醒.....	33
圖 4-7 影片播放視窗.....	35

表目錄

表 3-1 LCS 分數矩陣 X.....	21
表 3-2 紀錄 LCS 分數最大值的矩陣 X'.....	24
表 3-3 字幕與句子對應表格 (例).....	25



第一章 緒論

1.1 研究背景

在數位人文的領域裡面，除了對全文資料進行檢索及分析之外，影音檢索也是其中重要的一部份。在許多資料都已經數位化的現在，文字資料相較於影音資料是比較好分析而且進行檢索的，因此對於影音資料內容的檢索，則是另一個重要的問題。

影音資料約可分成幾種檢索方式[4]，例如影像視訊檢索，先將視訊經過分割之後，將每個片段進行分析，如色彩、紋理、外型、輪廓、相機運動等等，利用這些特徵將影片片段加上標記及索引，最後讓使用者可使用文字、影像甚至影片片段來進行檢索。但影像視訊檢索困難度較高，對使用者來說可能也不是這麼容易進行操作。

聲音檢索方面，一類是利用語音辨識，將語音轉換成文字，使用關鍵字或者經過自然語言處理來進行檢索，現在許多產品使用這個技術。或者利用音訊的波形、音高、旋律等，讓使用者可自行錄製一小段聲音，或者使用一段音訊，來對音樂進行檢索。

最一般的則是使用文字檢索。將影片加上描述影片內容的資料，讓使用者可對影片的標題以及描述內容進行檢索，目前許多影音資料庫都屬於此種類型，可讓使用者以關鍵字找到相關的影片類型，對於相當大量的影音資料，是最合適的方法。如 **youtube**，面對大量的使用者上傳影片，根據標題、關鍵字，透過翻譯、語意，而能對影音資料進行檢索。

而中央研究院數位典藏技術分項所建置的 **VCenter** 則是希望增添影音內容與

使用者的互動，加入了線上影音編輯工具，讓使用者可在影音內容上添加浮水印、加入字幕以及設定影音書籤等；並整合地理資訊系統，讓使用者可標記影片位置，而能夠以地圖呈現或檢索。[7]

但爲了對於影音資料的內容進行更細部的檢索，因此可以透過建立影片的字幕文字並取得字幕時間，讓使用者對字幕也能進行檢索，再利用字幕時間存取影片任一時間點，也就可以達到對影片內容搜索的目的。

Alexander Street Press 所建立的 VAST: Academic Video Online 影音資料庫，就將其所有典藏影片建立了完整可搜尋的字幕，並可根據搜索結果跳轉影片時間。另外可讓使用者自訂影片片段，給予註解，並提供永久連結來讓使用者分享給其他人。[6]

在資訊領域中，對於影片字幕多半著重在於擷取以及辨識的技術，除了靜態字幕之外，也有對於動態字幕擷取的研究，以及對於運動類型的動態影像上的資訊，如分數及球員名字，對影片進行結構化的分析以及辨識。

國內研究在早期，字幕偵測常用於運動視訊上，用來標記影片片段，而對整個影片做摘要或者將影片片段分類，進而能提取運動影片的精彩鏡頭[8]。或者對於移動文字的字幕擷取，先找出可能爲文字邊緣的像素並定位文字區域，再決定臨界值以區分文字與背景，接著消除錯誤辨識的區塊來提升辨識率[9]。近來也有只做影片字幕的偵測並消除，再將移除區域的影像內容修復，也就是結合影片文字偵測及影像修補的研究[10][11]。

而靜態字幕的擷取，通常一開始先找出影片中的文字區域，再針對文字區域做邊緣偵測處理來找出文字的邊界，製作成二值圖像之後，最後進行光學字元辨識以取得文字。

〈基於文字穿越線和筆畫連通性的視頻文字提取方法〉先利用 wavelet transform 與 clustering 得到文字區域，接著定位單行文字之後，利用筆畫的連通性，找出文字的關鍵點，再從關鍵點開始進行區域生長，來得到文字二值化的圖像。[12]

在張曉維的〈基於多影格的精確新聞影片文字偵測與擷取〉中，則先利用時間因素去除大部分的背景來取得文字區域，再利用 Canny edge detector 來判斷文字邊緣，然後進行掃描邊緣圖來分割文字，因而擷取出文字。[13]

〈利用支持向量機的影像文字偵測方法〉則是利用三個不同大小的影像轉換成灰階之後，計算最大差值並應用 k-means 演算法來分群成文字與非文字兩類，把連通區域找出來之後，合併三個梯度的連通區域而找出文字區域，再針對文字區域進行 Sobel 邊緣偵測來決定邊界，接著利用文字的幾何與文理特徵並應用 wavelet transform 進行文字驗證。[14]

也有許多在靜態影像進行即時文字偵測及擷取的，如在 DSP 上可自動對影像做分析，或者可馬上對拍照下來對影像作文字偵測處理與辨識，應用在行動裝備上則可以直接對 camara 拍到的影像進行分析，而可以達到即時翻譯，或者文字搜尋的目的[15]。在行動裝置充斥在生活中的今天，能夠即時將字幕偵測與辨識應用在動態影像上的話，未來應該有很大的展望。

1.2 研究對象及動機

『臺大文學講座』系列影片為臺大出版中心將臺灣近代文學的作家在臺灣大學的演講錄製成影音光碟，主要作家有白先勇、葉維廉、葉嘉瑩、王文興、高行健...等，共 31 部影片（50 片 DVD）（見附錄一）。演講內容包羅萬象，有作家親身的生長求學過程、文學創作的經歷、對於自己作品的剖析、文學美學的見解等等，體裁對象更包括詩、詞、小說、舞台劇、電影等，對於社會大眾與莘莘學子們是很好的閱讀輔助教材。[3]

此系列影片皆儲存在影音光碟上，光碟依舊是目前儲存影像的一個重要的媒介，但在使用上來說，我們無法對於內容做一個全盤的搜索，當使用者只對於影片中某一個部份的內容有興趣時，無法快速的跳到該片段進行觀看。以臺大文學講座系列影片為例，此系列影片皆含演講手冊，演講手冊內容是將整個演講去掉口語化、贅字以及更正一些口誤及文法，也就是把字幕潤飾之後的文字。而使用者在讀演講手冊時，如果看到一段感興趣的文字段落，並且想觀看演講者的現身說法，卻無法快速找到影片的片段；要找到該影片片段，我們只能依靠影片的該段內容，並比對演講手冊判斷先後順序，來不斷的切換影片時間，經過許多錯誤嘗試來找尋。

因此，為了解決這個問題，擷取字幕的文字讓字幕可以檢索；取得字幕的時間，就可以快速的跳躍到影片的該時間點；取得字幕與演講手冊的對應關係，更能讓使用者了解檢索到的字幕是在演講手冊上的何處。而且台大文學講座的影片內容皆是演講，字幕檢索更是適合，對字幕的檢索就等於對影片內容的檢索，使用者就可以對影片內容有全盤的掌握。取得字幕與演講手冊的對應關係，更能讓使用者了解檢索到的字幕是在演講手冊上的何處。

1.3 研究目的

根據研究對象的資料特性，對於「臺大文學講座」系列影片來說，影片內容皆是演講的內容，字幕檢索將是很好的對影片內容檢索的方法。而本論文使用現有軟體來進行字幕的擷取及辨識，再經過人工處理以得到較完整精確的字幕全文以及字幕時間。並利用研究對象特有的演講手冊，與字幕全文比對取得字幕與演講手冊的對應關係，最後建立一個『臺大文學講座』系列影片的字幕檢索系統。

此字幕檢索系統除了可對字幕進行關鍵字檢索之外，更進一步將字幕對應到演講手冊內容上，使得系統也可根據字幕呈現演講手冊中的文字；並且使用後分類的導覽方式，將檢索結果進行簡單分類，讓使用者可對結果篩選，能更快速的找到需要的字幕。而最重要的是在點選字幕之後，即可馬上觀看從該段字幕開始的影片片段。

1.4 研究方法與流程

爲了達成研究目的，我們需要影片字幕的全文、字幕在影片出現時間，還有字幕的發言者，以及字幕與演講手冊的對應關係。因此可建立以下研究流程《圖 1-1》。

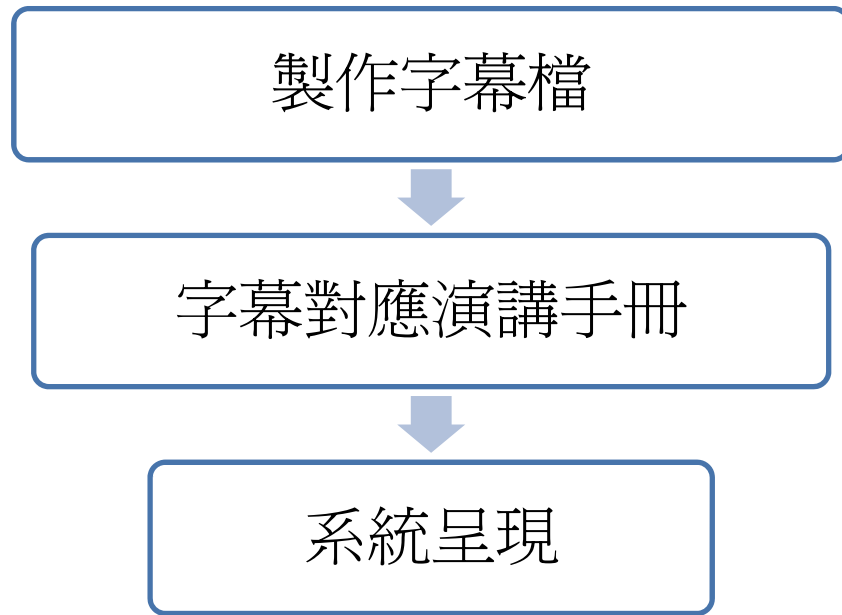


圖 1-1 研究流程圖

首先，爲了能夠讓字幕可以做文字的檢索，我們需要得到字幕的全文；而爲了讓系統能夠切換影片的時間，則需要字幕在影片中出現的時間。因此，製作一個影片的字幕檔，如 subrip、Station Alpha、Vobsub 等字幕檔格式，其中就包含兩者，接著再分析字幕檔的內容，就可以得到字幕全文以及字幕時間。在此步驟我們利用了現有的軟體來分析影片，並辨識文字。我們先使用 esrXP 來對整個影片進行分析，esrXP 利用字幕顏色及字幕外框顏色來過濾影片的每一幀(frame)，再依照畫素(pixel)的變化來決定每段字幕的時間，最後取得過濾出來的字幕圖片，以及字幕圖片對應到的影片時間。接著將字幕圖片使用 Microsoft Office Document Imaging 作光學字元辨識(OCR)，來取得字幕的全文。最後將字幕全文輸入到 esrXP 中，即可把文字對應到字幕圖片上，並輸出成需要的字幕檔。

取得字幕的全文之後，我們就可以拿全文來與演講手冊作比對，首要目的是取得每句字幕的發言者，以便系統呈現；另外，我們將演講手冊分割成數個句子，並找出字幕與演講手冊句子的對應關係。我們利用字幕與演講手冊句子的最長共

同子序列(Longest Common Subsequence)來決定字幕含意是否包含在演講手冊的句子意思當中。設定一個 LCS 分數為 LCS 字串長度除以字幕長度，LCS 分數越高表示共同子序列占字幕的比例越大，因而決定該句字幕是否相關於該演講手冊句子。接著假設每句字幕都要對應到一個句子，而且每個句子都要被對應到，以及字幕與句子的固定順序，建立一個挑選矩陣元素的演算法，而滿足這個演算法的組合中，LCS 分數總和最大的組合，我們就認為是字幕與演講手冊最相似的對應關係。最後以一個遞迴方法來決定 LCS 分數的最大值，並找出此最大值的組合，而得到字幕與演講手冊句子的對應關係。

最後，將得到的字幕全文以及字幕時間，還有字幕的發言者以及對應到的演講手冊句子，一起匯入資料庫中，並以系統呈現。系統呈現上，我們對於檢索結果加上多階層後分類的導覽方式，讓使用者可以對檢索結果進行篩選，可選擇以字幕出處、發言者及影片區間等分類組合來檢視檢索結果，方便使用者找到想要檢視的影片片段；影片播放則使用 HTML5 的 video 標籤，讓使用者只要使用支援 HTML5 的瀏覽器即可播放影片而不用另外嵌入元件；透過控制 video 標籤的參數就可以存取影片的任意時間點，而使得影片可以從字幕時間開始播放；另外，在檢索項目中及影片下方，都有字幕對應到的演講手冊句子，讓使用者更能知道搜尋到的字幕在演講手冊上是屬於哪個段落，其中影片下方的演講手冊句子會隨著影片時間而變換，也可讓使用者知道該影片時間對應到的演講手冊段落。

1.5 論文架構

本論文共分為五章。第一章介紹目前數為人文及資訊領域中，對字幕檢索與字幕偵測與辨識的背景資料，並介紹研究對象「臺大文學講座」系列影片的資料特性，以及研究的動機，而最後目的則是建立一個「臺大文學講座」系列影片的字幕檢索系統。

第二章「擷取影片字幕」講解方法流程中的建立字幕檔步驟，詳細說明如何根據字幕類型的不同而使用不同的方法來擷取字幕文字，主要使用 esrXP 來得到字幕圖片，接著 MODI 的 OCR 功能來辨認字幕字幕圖片取得文字，最後建立影片的字幕檔。

第三章「字幕與演講手冊的對應」則講解流程中的字幕對應演講手冊，先定義字幕、句子與 LCS 分數，再闡述整個對應的方法，包括字幕與句子的對應假設，以及說明找出相似度最高組合的遞迴方法，最後解釋為何要使用此方法的原因。

第四章「系統實作與功能」則是系統呈現部分，介紹本論文提出的檢索系統模型、實作過程，並以簡單範例簡介系統頁面與系統功能。

第五章「結論與未來工作」總結整篇論文，並思考未來方向。

第二章 擷取影片字幕

2.1 使用工具介紹

2.1.1 esrXP

esrXP 為 cphk (網路代號) 所建立的程式，用來擷取嵌在影片畫面上的字幕。esrXP 並不會自動分析影片然後直接將字幕抽出來，而是要通過設定顏色過濾器，利用字幕顏色固定、或者有外框的特性，對影片的每一格畫面作顏色過濾，再依照過濾出的畫素變化來決定一段字幕的時間，而抽取出影片中的文字部分及對應的影片時間。[16]

首先使用 esrXP 載入一部影片，這邊使用的編碼為 x264，封裝方式為 MPEG-4，解析度為 720*480。解析度越高，除了能增進 esrXP 過濾字幕的效果，對於之後 OCR 軟體辨認文字更為重要。

接著開啓過濾畫面《圖 2-1》，點擊在預覽畫面的下方長條處，可看到不同時間的影片預覽，接著選擇影片畫面中字幕的範圍，再根據影片設定字幕顏色及邊線顏色，然後啓動過濾器並調整過濾器設定，可設定色度、光度、色澤差距《圖 2-2》，直到字幕出現。



圖 2-1 esrXP 過濾畫面

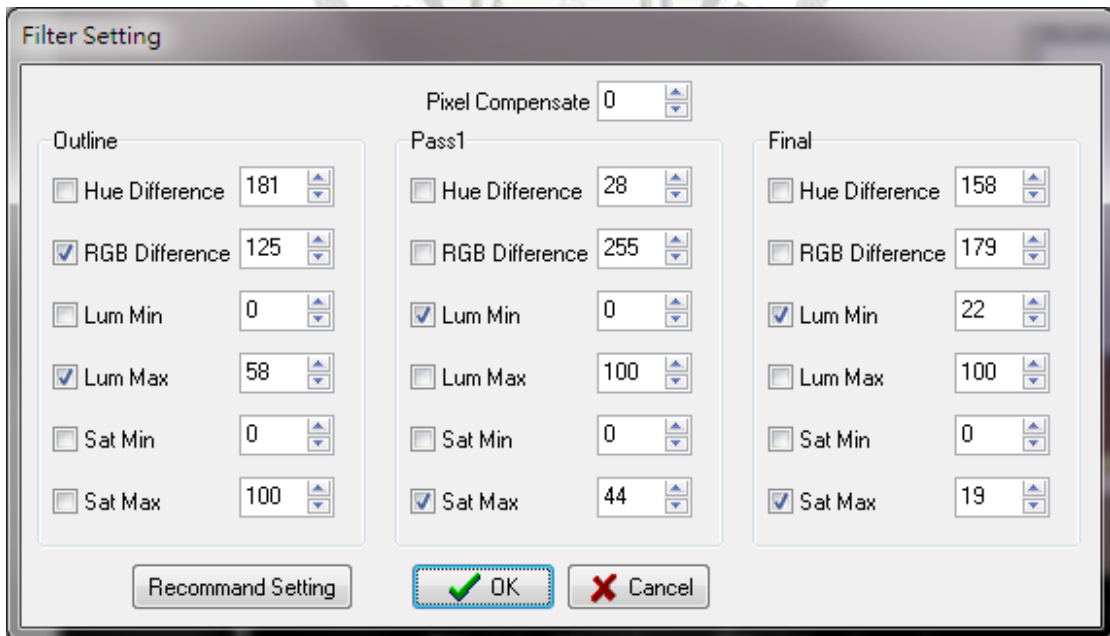


圖 2-2 esrXP 過濾器設定

跟著設定後處理及抓取設定，即可讓 esrXP 開始對影片進行分析。分析完整部影片之後，由於使用簡單的顏色過濾，會得到許多錯誤的字幕圖片，因此再開啓字幕管理畫面，刪除及合併多餘的字幕。此時可選擇輸出字幕圖片《圖 2-3》，將圖片交由 OCR 軟體來辨認並取得文字。

今天的這場演講之前
這有一件事情還是要先說明宣布
並且表示感謝
我們在上個學期開始創辦了
臺文所開始創辦了
臺灣大學臺大文學講座
在當時
在當時所得到的
教務長的協助跟支持
我們這個講座的主要的目的
就希望能夠替臺灣文學
留下一系列的重要的見證
所以我們都邀請
臺灣文學的重要的作家
那麼請他們來
做一個類似文學自傳的演講
然後希望把這個演講同時錄影
那麼將來還會製作成這個光碟
那麼提供給以後
不論是從事臺灣文學研究
或教學的人來使用
那麼我們上個學期總共舉行了三場
那麼這個第一場
在邀請的就是我們這個
我們文學院外文系的這個名譽教授

圖 2-3 字幕圖片

由於已分析完整部影片，esrXP 已取得每句字幕的起迄時間，所以接著在主視窗《圖 2-4》的文字區輸入對應到右方每句字幕的文字，可使用 OCR 辨認出來的文字或者自行打字輸入，最後就可以儲存字幕檔，字幕檔輸出格式可選擇 Sub Station Alpha (.ssa)、Subrip (.srt) 或者 Vobsub (.idx .sub)。

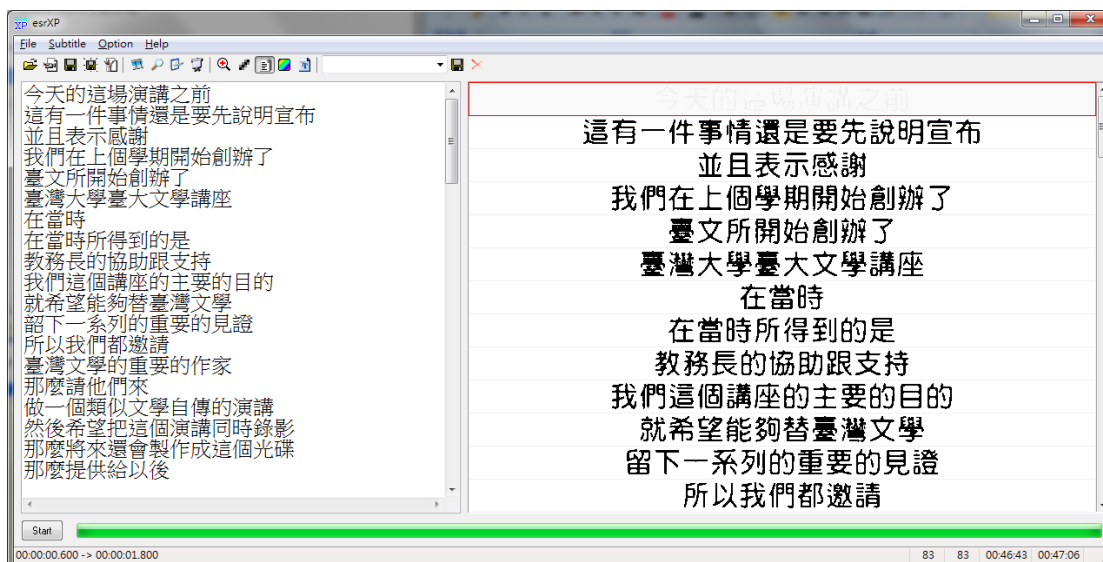


圖 2-4 esrXP 主視窗

2.1.2 Microsoft Office Document Imaging

Microsoft Office Document Imaging (以下簡稱 MODI)為 Microsoft Office 2007 中，用來掃描文件的軟體，它在掃描文件時會自動執行光學字元辨識 (OCR)。而它也可以另外載入 TIFF 圖檔，並執行光學字元辨識。

使用方式很簡單，開啓一個 TIFF 圖檔如《圖 2-5》，並選擇傳送文字到 word 即可，此時會產生一個副檔名為 html 的檔案，內含光學字元辨識的結果。

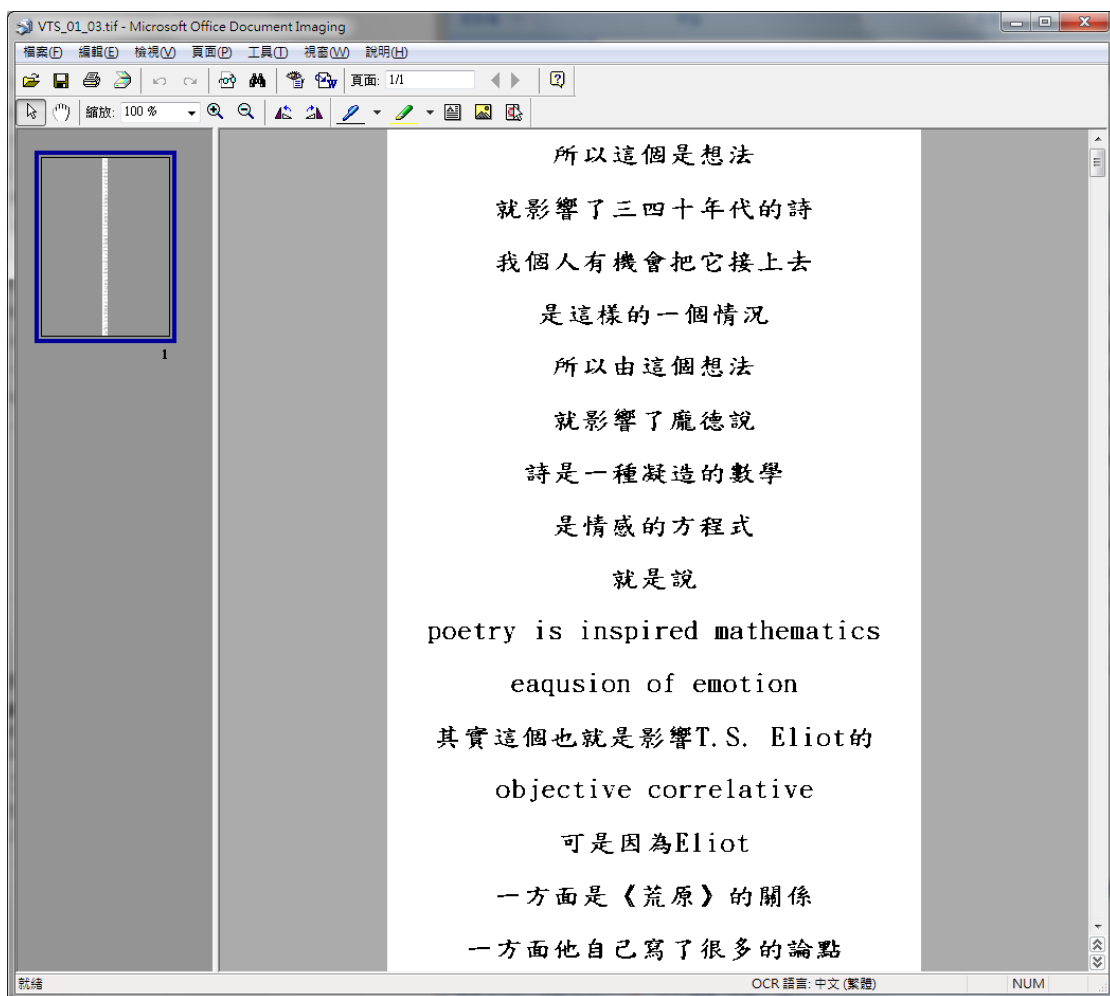


圖 2-5 Microsoft Office Document Imaging 畫面

2.2 材料分析

在 31 部影片（50 discs）當中，其中 3 部（3 discs）為 DVD 外掛字幕，剩餘的 28 部影片（47 discs）皆為字幕直接嵌在影像畫面上。在這 28 部影片裡面，其中 15 部（34 discs）為字幕與影像畫面分別獨立《圖 2-6》，13 部(13 discs)字幕與影像畫面重疊《圖 2-7》。



圖 2-6 字幕與影像畫面獨立



圖 2-7 字幕與影像畫面重疊

2-3 擷取流程

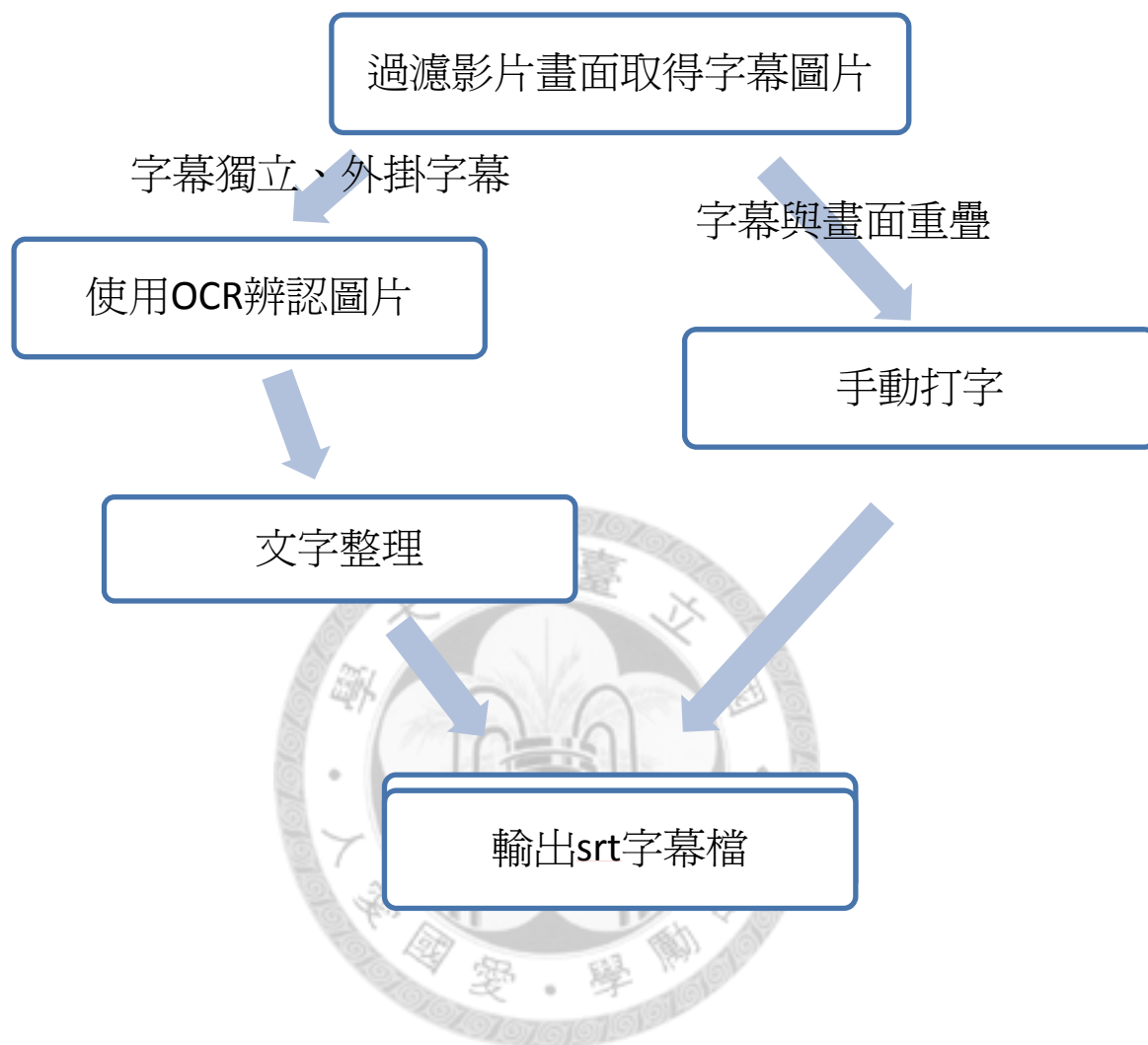


圖 2-8 擷取字幕流程圖

如流程圖《圖 2-8》所示，首先將 DVD 影片全部轉檔成以 x264 編碼的 MPEG-4 檔案，再使用 esrXP 載入影片，過濾影片畫面取得字幕的圖檔（BMP）之後，依照字幕鑲嵌在影像畫面上的情況分成不同的做法。

從字幕獨立的影片取得的字幕圖片，經過轉檔將 BMP 轉成 TIFF，交由 MODI 做光學字元辨識，取得辨識結果之後，再將辨識出來的文字經過整理，接著將辨識結果輸入到 esrXP 的文字區中，即可輸出字幕檔，而取得每句字幕的起迄時間。

文字整理部分，首先對幾份 OCR 辨認出來的文字進行人工更正錯字，再比對人工更正後的文字與 OCR 辨認結果，取得 OCR 較常辨認錯誤的文字組合，製作一個錯字取代列表（見附錄二），用來對剩下的 OCR 辨認結果做更正，加速工作的進行。而從字幕與畫面重疊的影片取得的字幕圖片，由於背景容易有雜訊，OCR 效果很差，因此直接進行人工打字在 esrXP 的文字區內，最後輸出字幕檔。

最後用程式分析字幕檔，取得每段字幕的開始時間及結束時間，然後匯入資料庫中，供系統網頁存取使用。本論文使用的是 Subrip 格式的字幕檔（srt），其格式如《圖 2-9》所示。



圖 2-9 srt 字幕檔格式

第三章 字幕與演講手冊的對應

3.1 演講手冊說明

演講手冊為每部影片所附的一本冊子，為影片中的演講內容的逐字稿，但經過潤飾，去掉許多口語化的語助詞，並修改文字及文法使得語句上更為通順有條理，因此與影片上的字幕有些許差異。演講手冊內容範例如下所示：

取自最初的讀者(林文月主講)演講手冊

柯慶明教授：

今天，開始這場演講之前，有一件事情要先說明並且表示感謝。我們（臺文所）上個學期開始創辦了「臺灣大學臺大文學講座」，當時得到了教務長的協助跟支持。以下略。

執行人：

大家好，我是洪建全基金會的張耀德。很高興有機會能夠跟臺大支持這樣一個計畫，因為我們基金會非常支持文學跟藝術的創作。也希望說能夠扮演所有的創作者跟欣賞者—也就是在座的各位—的一個溝通平台。很高興今天能夠來參與這樣的活動，希望以後臺大所有的文學講座會越辦越精彩。謝謝大家。

柯慶明教授：

林老師是我們臺大中文系的名譽教授，所以你們還是可以直接稱老師。但是假如她已經她不是名譽教授，而是完全退休的話，我們就要強迫各位叫太老師。因為林老師除了是我們的老師，她也是我的文學啟蒙人。以下略。

林文月教授：

謝謝柯所長。各位先生、各位女士、各位老師、各位同學，我今天這個題目呢，其實是有一點在那裡耍酷。我要講的內容非常簡單，我想談談我的課外活動。也就是教書研究以外的一些文字工作，我的經驗、我的想法。所以這個是非常非常個人的，很不學術的東西。我所要講的是關於自己的兩部分的寫作。一個是創作，以散文為主。另外一個呢是翻譯。但是我的翻譯當然是有限的，只能夠翻日文到中文。我也曾經翻過一點點英文的東西，可是留下的不多。以下略。


由以上範例可知，演講手冊內容是按照整個演講順序，給予發言者的名字，以及其所說的語句。因此，如果將每個字幕對應到演講手冊的句子，除了知道字幕與演講手冊句子的對應關係，還可得到該段字幕是何人所說的。

3.2 對應方法

3.2.1 定義

a. 字幕：同一時間在影片畫面上出現的一段文字。

例如最初的讀者字幕檔的前三句：



1
00:00:34, 533 --> 00:00:35, 800
今天的這場演講之前

2
00:00:35, 800 --> 00:00:40, 066
這有一件事情還是要先說明宣布

3
00:00:40, 066 --> 00:00:43, 333
並且表示感謝

「今天的這場演講之前」、「這有一件事情還是要先說明宣布」和「並且表示感謝」，這三句為三段字幕。

b. 句子：演講手冊中，以句號或分號斷開的一段文字。

例如最初的讀者(林文月主講)演講手冊中：

柯慶明教授：

今天，開始這場演講之前，有一件事情要先說明並且表示感謝。我們（臺文所）上個學期開始創辦了「臺灣大學臺大文學講座」，當時得到了教務長的協助跟支持。以下略。

其中「今天，開始這場演講之前，有一件事情要先說明並且表示感謝。」以及「我們（臺文所）上個學期開始創辦了「臺灣大學臺大文學講座」，當時得到了教務長的協助跟支持。」為兩個演講手冊的句子。

c. LCS 分數： $(\text{LCS 字串長度} / \text{字幕字串長度})$

此處 LCS 為字幕與句子的最長共同子序列(Longest Common Subsequence)，計算 LCS 時不包括字幕與句子字串中的標點符號，接著將 LCS 字串長度除以字幕的字串長度而得到 LCS 分數，用來表示該字幕與句子的相似程度。

3.2.2 目的

為了將整個字幕與演講手冊對應起來，字幕以每段字幕最為一個單位，演講手冊則以句子為一個單位，把每段字幕對應到手冊中包含該字幕意思的句子。

因為句子通常是比較長的，包含完整意思的，而字幕則可能是短短的一兩個字，語意上要與前後段字幕連貫的。因此，每個演講手冊上的句子可能會對應到多段字幕。以上述例子為例：

我們要將「今天的這場演講之前」、「這有一件事情還是要先說明宣布」和「並且表示感謝」三段字幕，對應到「今天，開始這場演講之前，有一件事情要先說明並且表示感謝。」這個句子上。

而「我們在上個學期開始創辦了」、「臺文所開始創辦了」、「臺灣大學臺大文學講座」、「在當時」、「在當時所得到的」、「教務長的協助跟支持」是接下去的六段字幕，則對應到「我們（臺文所）上個學期開始創辦了「臺灣大學臺大文學講座」，當時得到了教務長的協助跟支持。」這個句子上。

3.2.3 方法

首先將每段字幕對每個句子比對，取出最長共同子序列(Longest Common Subsequence)，並計算出 LCS 分數。

以字幕 1「今天的這場演講之前」與句子 A「今天，開始這場演講之前，有一件事情要先說明並且表示感謝。」為例，兩者的 LCS 為「今天這場演講之前」。並且算出 LCS 分數 $8/9 = 0.89$ 。

字幕 2「這有一件事情還是要先說明宣布」與句子 A「今天，開始這場演講之前，有一件事情要先說明並且表示感謝。」的 LCS 為「這有一件事情要先說明」，LCS 分數為 $10/14 = 0.71$ 。

字幕 3「並且表示感謝」與句子 A「今天，開始這場演講之前，有一件事情要先說明並且表示感謝。」的 LCS 為「並且表示感謝」，LCS 分數為 $6/6 = 1$ 。

字幕 4「我們在上個學期開始創辦了」與句子 A「今天，開始這場演講之前，有一件事情要先說明並且表示感謝。」的 LCS 為「開始」，LCS 分數為 $2/12 = 0.17$ 。

依此類推，將所有字幕與每個句子都算出 LCS 分數，以範例《表 3-1》來說，為字幕 1-29 與句子 A-E，最後存成一個矩陣 X。其中(1, A)、(2, A)、(3, A)、(4, A)就是上述四段字幕對應到句子 A「今天，開始這場演講之前，有一件事情要先說明並且表示感謝。」的 LCS 分數。

	A	B	C	D	E
1	0.89	0.11	0.22	0.33	0.22
2	0.71	0.00	0.21	0.14	0.14
3	1.00	0.00	0.00	0.17	0.00
4	0.17	0.92	0.17	0.33	0.42
5	0.25	1.00	0.25	0.25	0.25
6	0.10	1.00	0.40	0.60	0.40
7	0.00	0.67	0.00	0.33	0.00
8	0.00	0.63	0.25	0.25	0.25
9	0.00	1.00	0.11	0.22	0.22
10	0.25	0.50	0.83	0.50	0.50
11	0.00	0.40	0.90	0.40	0.20
12	0.18	0.09	1.00	0.36	0.36
13	0.00	0.29	0.00	0.86	0.43
14	0.10	0.50	0.80	1.00	0.60
15	0.00	0.17	0.00	0.67	0.50
16	0.17	0.33	0.33	0.92	0.42
17	0.23	0.23	0.23	0.54	0.15
18	0.08	0.08	0.15	0.54	0.15
19	0.00	0.00	0.00	0.57	0.14
20	0.09	0.36	0.45	0.73	0.27
21	0.00	0.25	0.25	0.88	0.38
22	0.07	0.47	0.13	0.27	0.73
23	0.29	0.14	0.43	0.14	0.57
24	0.10	0.30	0.20	0.50	0.60
25	0.07	0.40	0.27	0.47	0.87
26	0.14	0.29	0.14	0.21	1.00
27	0.08	0.08	0.15	0.31	1.00
28	0.13	0.13	0.25	0.25	1.00
29	0.00	0.00	0.00	0.13	1.00

表 3-1 LCS 分數矩陣 X

矩陣 X 中每一列(row)代表同一段字幕對 ABCDE 五個句子的 LCS 分數，反過來說，每一行(column)則是同一句子對 1-29 字幕的 LCS 分數。

矩陣中標記部分的 LCS 分數很高，代表著該段字幕與該句子重複的文字很多，於是我們認為該段字幕對應到該句子，而目的就是要找出這所有標記的部分，也就是字幕與句子的對應關係。

因此，根據我們的目的，再加上字幕與句子都有固定的排序，滿足這些條件的設定為：

1. 每段字幕對應到一個句子。
2. 每個句子都要被對應到。
3. 字幕與句子的對應之後保持原來的排序。

以一個 $m \times n$ 矩陣來看，即是：

1. 每一列(row)取一個 $m(i,j)$
2. 每一行(column)至少取一個 $m(i,j)$ 。
3. 取 $m(i+1,k)$ 時， $k = j$ or $j+1$ 。

根據這個取法，會有 H_m^n 個組合。而將所有 LCS 分數加起來最大的組合，就代表著字幕與句子最為相似的對應關係，也就是我們要的組合，以《圖 3-1》來看就是所有標記的部分。

而這個問題的解決辦法可以利用動態規劃（Dynamic Programming）來解決。首先我們利用遞迴求滿足上述條件的 LCS 分數總和最大值的組合。

對一個 $i \times j$ 矩陣，其 LCS 分數最大值總和 $M(i, j)$ ，即為取 $M(i-1, j-1)$ 與 $M(i-1, j)$ 之中較大者，再加上矩陣座標 (i, j) 的 LCS 分數 $m(i, j)$ 。於是可得下列遞迴的式子：

$$M(0, 0) = 0$$

$$M(i, 0) = -\text{infinity}$$

$$M(0, j) = -\text{infinity}$$

$$M(i, j) = m(i, j) + \max(M(i-1, j-1), M(i-1, j))$$

上述式子表示，對於每一個矩陣，我們只要去比較兩個較小的矩陣的 LCS 分數最大值，即可決定自己本身的 LCS 分數最大值。

這些 $i \times j$ 矩陣中較小的分割矩陣的 LCS 分數最大值 M ，我們另外把它記錄下來成為另外一個最大值矩陣 X' 《表 3-2》，除了避免一直重複計算相同的矩陣的最大值，此矩陣 X' 也用來追蹤擁有最大 LCS 分數總和的組合。

我們只要從矩陣 X' 的座標 $(1, A)$ 開始，一列一列的往下找每一列的最大值，並把座標記錄下來，就可以得到最大值矩陣 X' 中標記的部分，也跟 LCS 分數矩陣 X 的標記部分是一樣的，就是我們想知道的字幕與句子的對應關係。

	A	B	C	D	E
1	0.89	-999.89	-999.78	-999.67	-999.78
2	1.60	0.89	-999.56	-999.52	-999.52
3	2.60	1.60	0.89	-999.36	-999.52
4	2.77	3.52	1.77	1.22	-998.94
5	3.02	4.52	3.77	2.02	1.47
6	3.12	5.52	4.92	4.37	2.42
7	3.12	6.19	5.52	5.25	4.37
8	3.12	6.81	6.44	5.77	5.50
9	3.12	7.81	6.92	6.66	5.99
10	3.37	8.31	8.64	7.42	7.16
11	3.37	8.71	9.54	9.04	7.62
12	3.55	8.80	10.54	9.91	9.41
13	3.55	9.09	10.54	11.40	10.34
14	3.65	9.59	11.34	12.40	12.00
15	3.65	9.75	11.34	13.07	12.90
16	3.82	10.09	11.68	13.99	13.49
17	4.05	10.32	11.91	14.52	14.14
18	4.13	10.40	12.06	15.06	14.68
19	4.13	10.40	12.06	15.63	15.21
20	4.22	10.76	12.52	16.36	15.91
21	4.22	11.01	12.77	17.24	16.74
22	4.28	11.48	12.90	17.50	17.97
23	4.57	11.62	13.33	17.65	18.54
24	4.67	11.92	13.53	18.15	19.14
25	4.74	12.32	13.80	18.61	20.01
26		12.60	13.94	18.83	21.01
27			14.09	19.13	22.01
28				19.38	23.01
29					24.01

表 3-2 紀錄 LCS 分數最大值的矩陣 X'

根據 LCS 分數最大值矩陣 X'，我們找出滿足該最大值的組合之後，我們就可以建立一個字幕、句子以及發言者的表格，如《表 3-3》，只列出前九個字幕與句子的關係。

字幕	句子	發言者
今天的這場演講之前	今天，開始這場演講之前，有一件事情要先說明並且表示感謝。	柯慶明
這有一件事情還是要先說明宣布	今天，開始這場演講之前，有一件事情要先說明並且表示感謝。	柯慶明
並且表示感謝	今天，開始這場演講之前，有一件事情要先說明並且表示感謝。	柯慶明
我們在上個學期開始創辦了	我們（臺文所）上個學期開始創辦了「臺灣大學臺大文學講座」，當時得到了教務長的協助跟支持。	柯慶明
臺文所開始創辦了	我們（臺文所）上個學期開始創辦了「臺灣大學臺大文學講座」，當時得到了教務長的協助跟支持。	柯慶明
臺灣大學臺大文學講座	我們（臺文所）上個學期開始創辦了「臺灣大學臺大文學講座」，當時得到了教務長的協助跟支持。	柯慶明
在當時	我們（臺文所）上個學期開始創辦了「臺灣大學臺大文學講座」，當時得到了教務長的協助跟支持。	柯慶明
在當時所得到的	我們（臺文所）上個學期開始創辦了「臺灣大學臺大文學講座」，當時得到了教務長的協助跟支持。	柯慶明
教務長的協助跟支持	我們（臺文所）上個學期開始創辦了「臺灣大學臺大文學講座」，當時得到了教務長的協助跟支持。	柯慶明

表 3-3 字幕與句子對應表格(例)

3.3 使用此對應方法的理由

使用這個找出字幕與句子相似程度最大值的方法，原因在於，假設我們給定一個簡單的演算法如下。

1. 字幕 $\text{Sub}(i)$ 對句子 $\text{Sent}(j)$ 取 LCS 分數 $\text{LCS}(i,j)$
2. if $\text{LCS}(i,j) > \text{常數 } c$ ， $\text{Sub}(i)$ 對應到 $\text{Sent}(j)$ 且 $i = i+1$ 。
else $\text{Sub}(i)$ 對應到 $\text{Sent}(j+1)$ 且 $i = i+1, j = j+1$ 。
3. 回到步驟 1。

當出現以下的例子《圖 3-1》時，字幕將會對應到錯誤的句子。

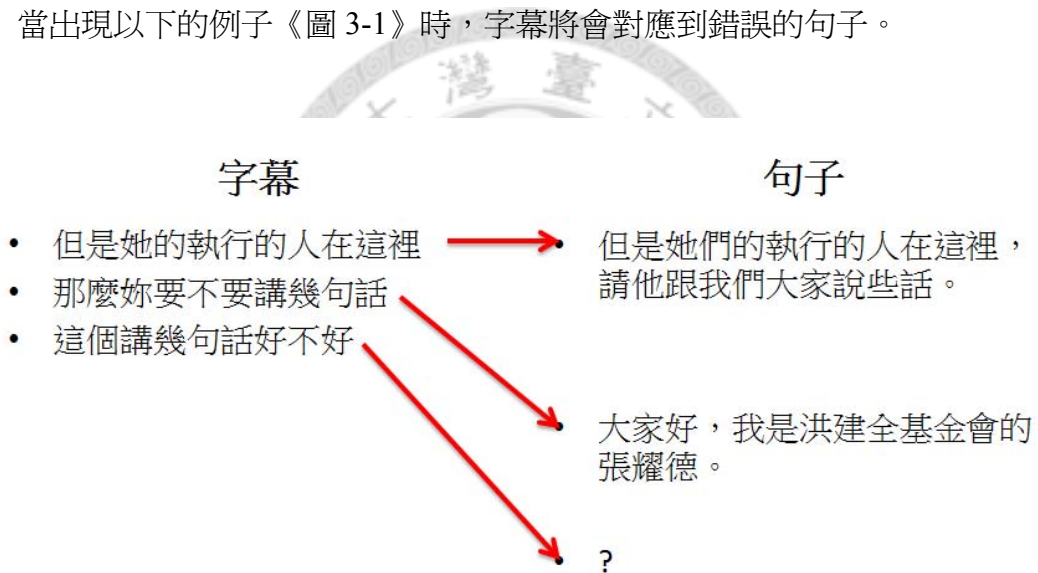


圖 3-1 對應錯誤範例

其中字幕「那麼妳要不要講幾句話」對句子「但是他們的執行的人在這裡，請他跟我們大家說些話。」的 LCS 分數只有 $1/10 = 0.1$ ，假設常數 c 為 0.5 ，那麼此字幕就會對應到下一個句子「大家好，我是洪建全基金會的張耀德。」，即產生了錯誤的對應。下一句字幕「這個講幾句話好不好」則不知道會對應到哪個句子。

當然這是一個過於簡單的演算法，但對於此類迭代式的演算法，常數 c 的值將會很難拿捏，無論是用來決定相似，或者決定不相似，都需要不斷的嘗試調整數值。於是我們採用了本論文提出的對整個字幕與句子計算相似程度最大值的方法。



第四章 字幕檢索影音系統

4.1 系統概述

從第二章和第三章當中，我們得到了每段字幕的起迄時間，以及字幕與演講手冊的對應關係，還有每段字幕的發言者。再加上各個影片的標題，我們就可以實做一個網頁模式的字幕檢索影音系統。

4.1.1 系統架構

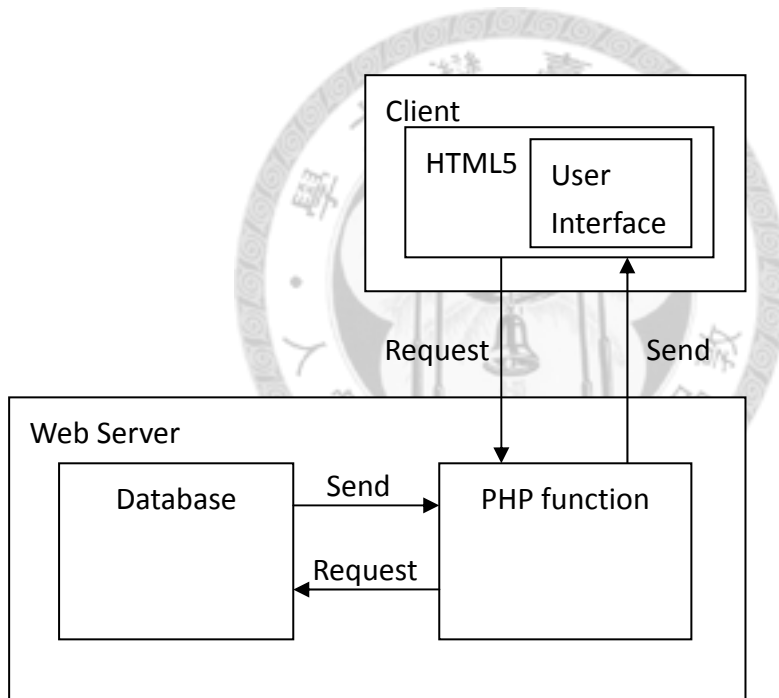


圖 4-1 系統架構圖

在系統架構上，簡單分成客戶端與伺服器端，首先由客戶端對伺服器端發出要求，伺服器端收到要求之後利用 PHP 函式向資料庫取得需要的資料，並轉換成 HTML 網頁送回給客戶端，此處利用到 HTML5 中的 video 標籤作為影音播放器。

4.1.2 系統環境

系統建置與運作的環境

- Apache Web Server Version 2.2.8
- PHP Script Language Version 5.2.6
- MySQL Database Version 5.0.51
- phpMyAdmin Database Manager Version 2.10.3

測試瀏覽器

- Google Chrome Version 20.0.1132.57
- Mozilla Firefox Version 14.0.1

影音格式

- MP4: x264 編碼，解析度 480*320，位元率 192 bps，幀率 25 fps。
- webm: VP8 編碼，解析度 480*320，位元率 192 bps，幀率 25 fps。
- 音訊: aac 編碼，位元率 128bps，採樣頻率 44100Hz。

伺服器基本頻寬

- 上傳: 320bps 以上 (192bps+128bps)。

4.1.3 HTML5 video 標籤介紹

video 標籤為 HTML5 新增的一個標籤，可以簡單的將影片嵌入網頁中，而不用另外嵌入 Flash 播放器或其它外掛元件來播放影片。[17]

最早 video 標籤為 Opera Software 在 2007 年二月所提出，目的是希望每種不同的瀏覽器都能支援一種影片格式而能直接在網頁上播放。最早所建議的影片格式為 Ogg Theora，後來 H.264/MPEG-4 AVC 也加入戰局，Google 也跟著提出以 VP8 編碼（開源）當基礎的 WebM 影片格式。

主流瀏覽器所支援的格式如《表 4-1》所示，因此最少只要準備兩種影片格式即可讓所有主流瀏覽器播放影片。

影片格式 瀏覽器	Ogg Theora	H.264	VP8(WebM)
Google Chrome	O	O(removal planned)	O
Mozilla Firefox	O	X	O
Internet Explorer	X	O	X
Apple Safari	X	O	X
Opera	O	X	O

表 4-1 各主流瀏覽器對 video 格式支援表

在使用 video 元件上，首先在網頁上嵌入一個 video 元件，再根據點選的字幕，來更換 video 元件的來源影片檔案，接著只要在 video 元件讀取完該影片檔案的 metadata 之後，即可透過控制 currentTime 這個參數，來隨意存取影片檔案的任何一個時間點。

4.2 系統功能及操作實例

4.2.1 關鍵字檢索

使用者可以對所有影片的字幕以關鍵字檢索，可得檢索結果畫面，此為以「白先勇」為關鍵字的檢索結果《圖 4-2》。



圖 4-2 檢索結果畫面

檢索結果條目中《圖 4-3》，左側為影片截圖，灰底藍字是檢索到的字幕，灰底右方為發言者。下方則有該字幕在影片出現的時間及影片名稱，以及該句字幕對應的演講手冊句子，可讓使用者判斷是否是演講手冊中感興趣的段落。

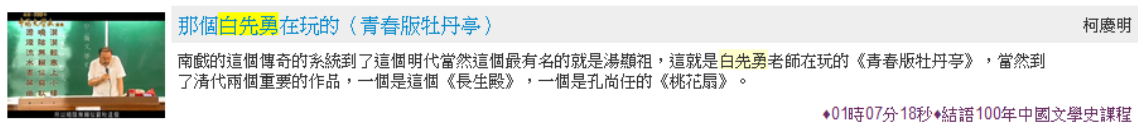


圖 4-3 檢索結果條目

4.2.2 後分類索引

出處	發言者
+ 白先勇的藝文世界-2-閱讀白先勇	51
+ 白先勇的藝文世界-1-開幕式	34
+ 白先勇的藝文世界-8-驀然回首	27
+ 白先勇的藝文世界-5-白先勇的舞臺劇	8
+ 白先勇的藝文世界-6-尹雪艷與秦可卿	7
+ 白先勇的藝文世界-7-青春版牡丹亭	6
+ 白先勇的藝文世界-4-玉卿嫂的電影創作	4
+ 王文興與現代主義	4
+ 白先勇的藝文世界-3-影像與文學的對話	3
+ 人文批評的反思	3
+ 結語100年中國文學史課程	2
+ 神思的機遇	2
+ 家變例講	1
+ 電影與文學間的曖昧關係	1
+ 生活與寫作	1
+ 在舞臺上尋找女性角色	1

圖 4-4 後分類索引

得到檢索結果之後，系統根據檢索結果的詮釋資料(metadata)，如出處、發言者及字幕時間，來進行分類處理，有兩種多階層的分類方式，分別為出處→發言者→發言時間，以及發言者→出處→發言時間，發言時間是以每十分鐘做為一個區段，讓使用者可依照不同的分類組合來篩選檢索結果。

最後在系統畫面左側的一個區塊呈現，如《圖 4-4》，上方有「出處」與「發言者」的分頁標籤。「出處」代表的是字幕出自於哪部影片，「發言者」則是字幕出自於何人之口。點擊文字前的十字，可展開下個階層的分類；分類「出處」底下是「字幕者」，再往下還有「發言時間」，為字幕在影片出現的時間區段，如《圖 4-5》；相對的，分類「發言者」底下則是「出處」，再往下也有「發言時間」。

出處	發言者	
+ 白先勇的藝文世界-2-閱讀白先勇		51
+ 白先勇的藝文世界-1-開幕式		34
- 白先勇的藝文世界-8-驀然回首		27
- 陳若曦		9
40-50分		4
50-60分		5
+ 鄭恆雄		6
+ 陳泰然		3
+ 李歐梵		2
+ 提問者		2
+ 葉國良		2
+ 葉維廉		1
+ 司儀		1
+ 顧敏		1

圖 4-5 後分類階層展開

當檢索結果資料量過多時，爲了避免讓使用者一則一則字幕下去找尋需要的字幕，可利用此兩項分類以及底下的階層分類來選擇需要的分類組合；出處底下會顯示檢索結果該出處的發言者分類，底下的發言時間亦然，因此使用者就可以依照想看的影片出處，以及演講者來篩選檢索結果。點擊文字即會在右邊檢索結果頁面中，顯示滿足該條件的搜尋結果，並反白文字提醒《圖 4-6》。若不是想要的篩選結果，點選其他分類字樣就可以對原本關鍵字檢索結果進行重新篩選。

- 白先勇的藝文世界-8-驀然回首	27
- 陳若曦	9
40-50分	4
50-60分	5

圖 4-6 點擊分類反白提醒

4.2.3 影片播放

當使用者點擊字幕之後，即跳出一個小視窗，這裡用的是 JQUERY-UI 的 dialog 屬性，透過將 div 標籤設定為 dialog 屬性，以及 CSS 與 javascript 控制，可把網頁中的元件先隱藏起來，點選字幕時才跳出小視窗顯示，視窗標題為點選字幕的影片出處。

播放器使用 HTML5 的 video 元件，我們準備了以 x264 (H.264 based) 編碼的 MPEG-4 檔案及以 VP8 編碼的 webm 檔案，使用 Google Chrome 或 Mozilla Firefox 瀏覽器，皆不需另外嵌入其他物件，可直接播放影片。

點選字幕之後，首先將 video 元件的影片來源更換成字幕的影片出處，包括 MP4 檔案及 webm 檔案，網頁讀取時，video 元件會根據瀏覽器決定使用的來源檔案，或者顯示不支援的訊息。在 video 元件讀取完影片的 metadata 之後，就可以透過控制 currentTime 這個 video 元件的參數來決定影片開始播放的時間。因此，就算不關閉此影片視窗，也可以直接點選其他字幕，播放器一樣會直接跳轉時間。

而下方顯示演講手冊的部分，則是在呈現檢索結果時，先將此次頁面呈現的字幕出處影片的所有字幕先從資料庫取出並儲存起來，由於同個頁面出現的不同影片出處通常不會太多，因此不太會影響系統效能。接著則是讀取同樣 currentTime 這個參數，再根據影片時間來判斷該顯示的演講手冊句子。下一頁是點擊「那個白先勇在玩的〈青春版牡丹亭〉」出現的畫面。



南戲的這個傳奇的系統到了這個明代當然這個最有名的就是湯顯祖，這就是白先勇老師在玩的《青春版牡丹亭》，當然到了清代兩個重要的作品，一個是這個《長生殿》，一個是孔尚任的《桃花扇》。

圖 4-7 影片播放視窗

第五章 結論與未來工作

5.1 結論

現有的影音光碟播放模式中，由於硬體及軟體上先天性的缺點，而讓使用者不能對整個影音資料做全面的搜索。利用臺大文學講座系列影片皆為演講的特性，擷取影片字幕之後，透過字幕檢索，即可對整個影片內容有全盤的掌握。

建立字幕與演講手冊的對應關係之後，可讓使用者清楚了解到字幕出自於演講手冊的哪個段落，幫助使用者確認目標字幕。但由於演講手冊對演講內容的潤飾，造成手冊文字與字幕文字的些許差異，使得字幕無法簡單的對應到手冊的句子。因此，我們以最長共同子序列來決定字幕與演講手冊句子的相似程度，且為了避免字幕對應到錯的句子，選擇了找出字幕與句子相似值總和最大值的方法。

取得字幕檔中的字幕和字幕時間，以及字幕與演講手冊的對應關係等資訊，即可建立一個隨選字幕的影音系統。再加入後分類索引之後，系統提供了一個更有效率的檢索介面，對檢索結果有初步的分類，讓使用者更能掌握整個檢索結果，而可根據點選後分類來篩選檢索結果。

另外使用 HTML5 的 video 元件，讓使用者只要擁有支援 HTML5 的瀏覽器如 Google Chrome、Mozilla Firefox 等，就可以播放影片而不需其他附加元件。並透過控制 video 元件，根據字幕的時間跳躍到影片的各時間點，達到我們對影片內容檢索的目的。

5.2 未來工作及展望

在影片字幕擷取上，本論文並沒有提供一個好的解決方法。也許可將 esrXP 輸出的字幕圖片作去污處理，來增進 OCR 的文字辨認效果。或者直接對於動態影像，利用字幕背景會移動的特性，來擷取靜態的字幕。如此一來，即可對於所有類型的影像檔案皆可快速的擷取影片字幕，並製作字幕檔，而讓影音系統能夠利用。

在系統效能上，本系統使用傳統的 HTML 網頁設計，因為前後兩個頁面中的大部分 HTML 碼是相同的，因此浪費了許多客戶端與伺服器端的回應時間，造成網頁反應時間過長。將介面改為使用模組化的架構，以及版面配置的設計加上 AJAX 網頁設計，應可使系統效能更進一步。

在數位人文領域中，除了可建立對於許多資料都可使用的一般性工具，也可以根據數位資料特性來建立相關的系統或工具，讓使用者能處於一個為特定資料設計的研究環境底下，來進行資料的使用、分析或處理。本系統原立意於讓影片可讓大眾快速存取及使用，但也利用了一些資料特性，如演講手冊。若要將此影音資料看做研究材料的話，則需要進一步的思考，來製作一些分析工具。

參考文獻

- [1]陳詩沛、杜協昌、項潔，〈史料整體分析工具之幕後-介紹臺灣歷史數位圖書館的資料前置處理程序〉，數位典藏與數位人文國際會議，臺北市，2009
- [2]蕭屹灵，〈日治法院檔案系統及其後分類呈現〉，碩士論文，資訊網路與多媒體研究所，國立台灣大學，2008
- [3]臺大文學講座系列，臺大出版中心<<http://www.press.ntu.edu.tw/>>
- [4]王駿發，〈多媒體影音系統〉，《科學發展》2007年3月，411期，6-13
- [5]臺灣歷史數位圖書館(Taiwan History Digital Library)，<<http://thdl.ntu.edu.tw/>>
- [6]VAST: Academic Video Online，
<<http://alexanderstreet.com/products/vast-academic-video-online>>
- [7] Hsiang-An Wang, Chih-Yi Chiu, Yu-Zheng Wang, "VCenter: A Digital Video Broadcast System of NDAP Taiwan," 11th European Conference on Research and Advanced Technology for Digital Libraries, 2007.
- [8]夏勝雄，〈結合字幕與視覺特徵以進行棒球視訊中語意式事件之偵測與分類〉，碩士論文，電機工程研究所，國立中正大學，2004
- [9]陳永健，〈適用於數位視訊中移動字幕之偵測、定位以及擷取方法〉，碩士論文，電機工程研究所，國立中央大學，2005
- [10]方志倫，〈電腦視覺特徵值萃取於字幕視訊處理及視訊防手震系統設計之研究〉，博士論文，電機工程研究所，國立中央大學，2011
- [11]陳萬鴻，〈視訊字幕區域偵測與修復〉，碩士論文，電機工程系研究所，國立臺北科技大學，2011
- [12]田破荒、彭天強、李弼程，〈基於文字穿越線和筆畫連通性的視頻文字提取方法〉，電子學報 2009 年 1 期，72-78
- [13]張曉維，〈基於多影格的精確新聞影片文字偵測與擷取〉，博士論文，資訊工程學系，淡江大學，2011
- [14]魏貽誠，〈利用支持向量機的影像文字偵測方法〉，碩士論文，電子工程系，國立臺灣科技大學，2011
- [15]呂信德，〈一個應用於攝影機擷取文字影像之光學文字辨識前處理系統〉，博士論文，資訊工程研究所，國立中央大學，2010
- [16]cphktool，<<https://sites.google.com/site/cphktool/esrxp>>
- [17]W3C，<<http://www.w3.org/TR/2011/WD-html5-20110113/video.html>>

附錄一 臺大文學講座系列影音光碟列表

- 1.我對台灣文學與台灣文學研究的看法(齊邦媛主講)
- 2.原鄉與我的創作(席慕蓉主講)
- 3.我與文學(司馬中原主講)
- 4.家變例講(王文興主講)
- 5.背海的人例講(王文興主講)
- 6.結語 100 年中國文學史課程(柯慶明主講)
- 7.最初的讀者(林文月主講)
- 8.我是怎麼寫起詩來(痲弦主講)
- 9.我的文學自傳(葉維廉主講)
- 10.我如何寫小說(王文興主講)
- 11.一個 文藝青年能做些什麼一個文學出版社能做些什麼(隱地主講)
- 12.小說與我(鄭清文主講)
- 13.電影與文學間的曖昧關係(小野主講)
- 14.生活與寫作(陳若曦主講)
- 15.從臺北人到青春版牡丹亭(白先勇主講)
- 16.觸機(張曉風主講)
- 17.顛躓在詩路上的扁平足(商禽主講)
- 18.我的詩路歷程(杜國清主講)
- 19.神思的機遇(葉維廉主講)
- 20.管管腦袋開花(管管主講)
- 21.長篇小說自剖(李喬主講)
- 22.在舞台上尋找女性角色(汪其楣主講)
- 23.神龍見首不見尾(葉嘉瑩主講)
- 24.汪精衛詩詞中的精衛情節(葉嘉瑩主講)
- 25.陳曾壽詞中的遺民心態(葉嘉瑩主講)
- 26.鏡中人影(葉嘉瑩主講)
- 27.白先勇的藝文世界
- 28.王文興與現代主義(葉維廉主講)
- 29.詩體驗(葉維廉主講)
- 30.人文批評的反思(李歐梵主講)
- 31.文學與美學(高行健主講)

附錄二 MODI 常見錯字取代列表

正確	錯字	正確	錯字	正確	錯字	正確	錯字	正確	錯字
啓	啟	卻	谷 p	假	！跋	錦	金帛	敘述	放述
叫	p 川	弭	弓耳	倒	僅 l	衝	種 i	捕捉	插捉
叫	p 圳	惜	，IH	倒	侄 l	禧	示喜	或者	感者
叫	p 糾	證	言登	剩	乘 {	傳統	傳絲	或是	感是
域	1 或	詳	言羊	須	丰頁	候	{ 吳	札	丰 L
批	杜【	讀	言賣	跳	劉【	談	言炎	改造	改浩
剛	岡 1	讀	i 賣	瞬	目舜	說	言兌	唾棄	睡棄
剛	岡 {	別	另叮	瞬	口舜	演	' 寅	語言	飴言
剛	岡 l	別	另 U	咬	口交	頂	J 頁	語言	謚言
剛	岡 [誦	言甬	挑	打 [情	l 青	細緻	細繳
可能	司能	餘	食余	謀	言樂	柯慶明	祠慶明	堵塞	堵塞
可是	司是	關鍵	關鏟	顧	雇頁	幾乎	競乎	自我	自扎
可以	司以	悟	' 一吾	鳴	q 鳥	順	川頁	我們	扎們
可惜	司惜	淵博	淵搏	護	i 藍	問題	問題	博物	搏物
講	言莆	觸	角蜀	河	；可	問題	閒題	這些	這竺
講	言黃	憶	' l 意	恥	耳崧	批	妣	舞臺	舞憂
講	言驚	曦	日羲	慚	！斷	困難	困雞	舞臺	舞轟
翻譯	畫羽譯	亂	商 [昵呢	呢呢	什麼	什座	舞臺	舞畫
翻	畫 ! l	貼	貝占	洋	i 羊	什麼	什曆	舞臺	舞贏
翻	畫弱	幾句	繼句	哎哨	哎暗	怎麼	怎座	說書	說畫
翻	鄱】	臺	臺	間諜	間謀	怎麼	怎曆	已經	已細
翻	劃 l	們	1} !	劇	\$ U	怎麼	怎魔	重新	面新
性	' 一生	們	！門	嘎	嘎	怎麼	怎庭	努力	另力
性	{ 生	們	一門	破	鉸	什麼	什魔	家	豕
性	' 一生	們	{ 門	《	((這麼	這魔	觀眾	靦眾
性	l 陸	們	1 門	》))	那麼	那魔	當眾	當羅
迴轉	迴轉	輔	車甫	》))	那麼	那庭	觀眾	觀羅
話	一言古	馬	· 鳥	欸	歟	麼	魔	形象	形魚
臘	月饑	類	套頁	祥	才羊	這麼	這魔	關係	關保
願	原頁	讚	；贊	裏面	裏面	這麼	這曆	男女	另女
感動	威動	讚	i 贊	〈	<	這麼	這庭	藝術	藝南
情感	情威	順	」一頂	〉	>	這麼	這反	藝術	藝備
感情	威情	於	方全	慢慢	幔幔	那麼	那曆	藝術	藝萄

正確	錯字	正確	錯字	正確	錯字	正確	錯字	正確	錯字
感受	威受	盜蹊	盜鑣	慢慢	慢慢	怎麼	怎曆	藝術	藝婉
感覺	威覺	聚黨	聚室	慢慢	幔幔	恰恰	怡怡	藝術	藝衛
感通	威通	傳	！專	慢	{曼	恰恰	怡怡	藝術	藝婚
感到	威到	傳	l專	慢	{雙	恰恰	怡怡	情緒	蜻緒
感受	成受	傳	{專	壠	土龍	完全	完金	藝術	醒術
感覺	成覺	衡	模i	例	仔l	龐大	廳大	藝術	蘿術
感通	成通	讚美	讀美	個	l固	企圖	企圓	藝術	垂術
感到	成到	祿	宇泵	個	1固	暢銷	暢鎖	藝術	騷術
感	威	驥尾	驗尾	個	{國	衝動	衡動	藝術	惡術
感覺	威覺	悱	付}	個	l國	邏輯	還輯	藝術	藝砲
感受	威受	「小	rl」，	個	1國	主觀	主靦	藝術	騷婚
感通	威通	「小	rl」，	個	l國	瑣事	琅事	藝術	驪術
感到	威到	小	ll，	個	l同	我們	找們	藝術	噩術
情感	情威	小	‘，’	個	1同	象徵	象徽	藝術	惡術
議	言義	小	’，’	真實	佩實	象徵	象徽	藝術	藝當
膽	月晉	辭鏡	掛申言竟	真實	員實	意象	意濛	藝術	蘇術
統	糸充	顏	彥頁	哦	p我	裡頭	理頭	利用	利月
於是	方令是	排	相}	哦	峨	彼岸	很岸	重複	孟複
掙扎	主爭才 L	粘衰草	牯衰草	哇	p圭	宣洩	宣浪	繪畫	繪蠱
十	+	憤	rl*	哇	哇	候	9矣	繪畫	續畫
抖	料	憤	rl\$	釧	811	是	杲	老師	毛師
的	白勺	憤	l質	繡	繙		，.	印象	印魚
到	至l」	怪	111呈	枝幹	技幹			抽象	抽魚
到	至l』	謾	言曼	樹枝	樹技		，.	抽象	抽廬
到	至lj	衽	鮐	慷慨	慷慨		，.	抽象	抽盞
到	至11	股	月貧	恆	11互		.	抽象	抽濛
倒	fj	滕	月樂	恆	j叵		→	抽象	抽激
因為	黑集	憐	.ln	恆	t叵	」	』	形象	形激
專科	再科	憐	l轟	匈奴	匈奴	瀘	瀘	形象	形盞
手稿	手不高	憐	.l轟	恨	j限	嚴肅	嚴祖	連繫	連繫
邵	召B	喚	p奐	盜跖	盜路	眾生	羅生	換句話說	換句話說
狀	糾犬	鄭	奠B	早夭	早天	改造	政造	翻譯	翻諸

正確	錯字	正確	錯字	正確	錯字	正確	錯字	正確	錯字
多年	多無	請	言青	怨俳	怨俳	虛構	處構	觀眾	觀試
老年	老無	顛	才顛	俳	付l	清	i青	這麼	這廬
一年	一無	藏闖	藏闕	衝擊	衡擊	擺脫	攝脫	籠統	龍統
二年	二無	金縷	金樓	諫	言東	擺脫	握脫	輝煌	認煌
兩年	兩無	踣	脚	撼動	據動	虛幻	廬幻	形象	形濛
三年	三無	則	貝lj	孤戀花	抓戀花	認識	認議	膚淺	廬淺
四年	四無	遶門	澆門	編	褊	知識	知議	判斷	判勵
五年	五無	蜂	虫筆	觀	蓬見	意識	意議	溝通	麗通
六年	六無	邛	亡仔	觀	瞿見	敘述	教述	賦予	賦予
七年	七無	野	里予	蟬	虫單	敘述	斂述	邏輯	遜輯
八年	八無	蜂	虫華	濟	'齊	敘述	數述	法	婁
九年	九無	蜂	虫輩	讓	言襄	敘述	故述	我們	栽們
十年	十無	元積	元積	噴呐	噴呐	聯繫	聯繫	我的	栽的
百年	百無	吟	口今	暗	日音	聯繫	聯繫	這個	馴團
千年	千無	吟	吟	斷	昧斥	浪漫	浪湯	忌	·已
今年	今無	始	力台	轟	轟	滑稽	滑稽	往往	柱柱
年齡	無齡	蔡琰	蔡談	題材	題村	屈原	屈原	慢	饅
年輕	無輕	剔	易l	公司	公司	世紀	世記	以後	以復
年紀	無紀	同時	異皚	卿	鯁)	年紀	年記	語言	語旨
年代	無代	心	'乙	卿	卯)	托爾斯泰	托爾斯泰	企圖	全圖
年代	坏代	心	'已	因緣	因緣	繼續	繼緬	白鴿	白鴿
年級	坏級	心>'已		機緣	機緣	最後	最復	了	T
當年	當無	研	石升	歡	薑欠	顛覆	頭覆		二
幾年	幾無	痙弦	痘弦	臺	薑	顛覆	願覆	臺灣	毫灣
每年	每無	痙弦	痙弦	臺	薑	顛覆	頭覆	推	雅
多年	多無	話	言舌	像	〕象	可是	司是	言	旨
幼年	幼無	關於	喝昆方合	像	!象	些	訾	感	威
青年	青無	眼	目艮	細膩	糸叭貳	瞬間	瞬間	領	頷
結	糸吉	流	i奈	快	'l央	括	才舌	犧牲	犧牡
錯	金昌	傻	一谿	劇	'j	何	1司	欸	欵
仁	舛	往	拄						

附錄三 資料處理流程

1.DVD 轉檔

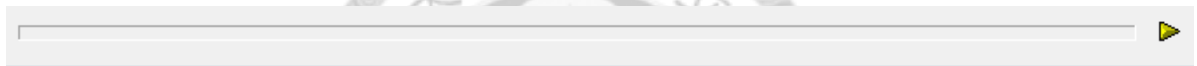
使用軟體將 DVD 轉換成 mp4 檔案(編碼 x264(H.264)，解析度 720*480，位元率 768bps，幀率 30fps)。此處為 DVD 影片原解析度及幀率，給 esrXP 分析影片使用。並另轉出 mp4 檔案(編碼 x264(H.264)，解析度 480*320，位元率 192 bps，幀率 25fps)，以及 webm 檔案(編碼 VP8，解析度 480*320，位元率 192 bps，幀率 25fps)。後兩者給系統使用，放在 media 這個資料夾中。

2.字幕圖片擷取

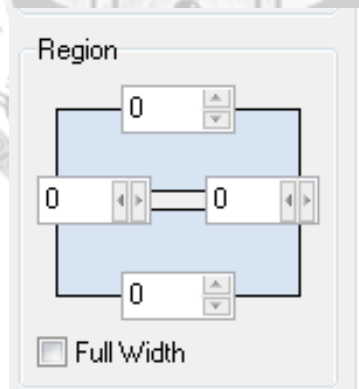
2.1 字幕嵌在影片畫面上的影片

a.開啓 esrXP，點選 File -> Open Video...，選擇轉檔好的影片檔案(.mp4)。

b.點選 Subtitle->Filter，打開過濾畫面，先點擊預覽畫面下面的 bar，找到影片中有字幕出現畫面。



c.選取字幕區域，直接點擊預覽畫面可選取字幕區域，或可開啓 Full width 來進行選取。



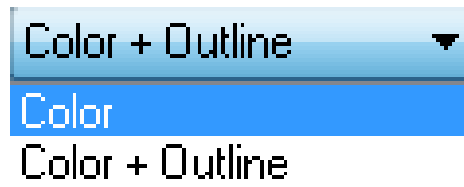
d.設定字幕顏色及字幕外框顏色。



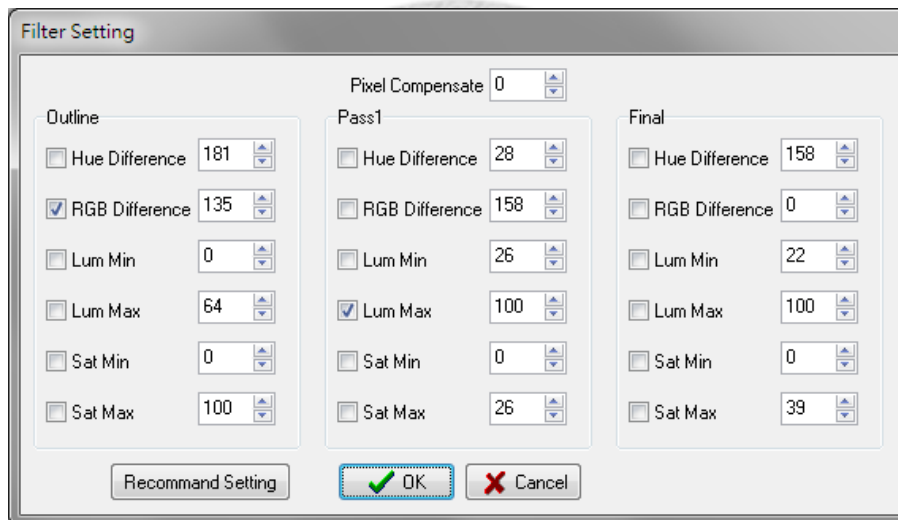
可用滑鼠指著預覽畫面中的顏色，RGB 值會在左下方顯示。



e.視情況選擇過濾方式，可選字幕顏色或者字幕顏色加外框

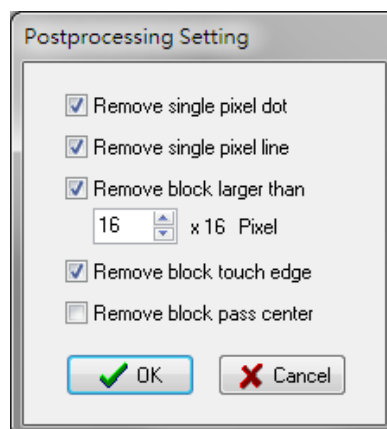


f.勾選 Enable filter，並開啓 Advance 來調整過濾器，直到字幕出現，並選擇其他部分的影片預覽畫面來確定其他部分的字幕也有過濾到。

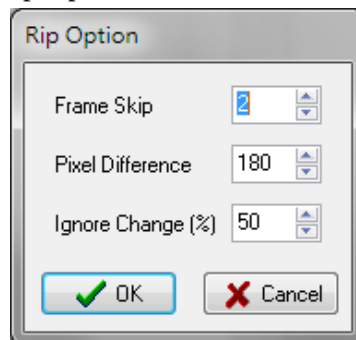


可點選 Recommend Setting 來使用軟體建議的設定。

g.開啓 Postprocessing 來選擇後處理選項。

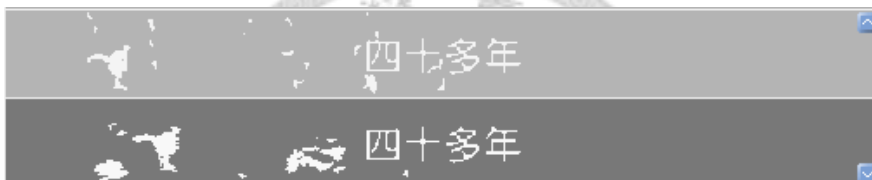


h.回到 esrXP 主視窗，設定 Rip Option。



i.按下左下角的 start，開始分析影片。

j.等到進度條跑完，開啓 Subtitle->Manager，刪除及合併多餘的字幕。左鍵點選字幕是刪除該字幕；右鍵點選第一個你要合併的字幕，再點選最後一個要合併的字幕，就會將這段時間的字幕都合併起來，合併的字幕會出現上下箭頭在字幕上，接著可用左鍵選取這段合併字幕中要顯示的字幕。



k 選擇 Option->White Background 將字幕反白，最後點選 File->Save OCR Image 存成給 OCR 辨識的圖檔(.bmp)。設定每個圖片包含 500 subtitles(太多圖檔會太大)，scale 爲 1。

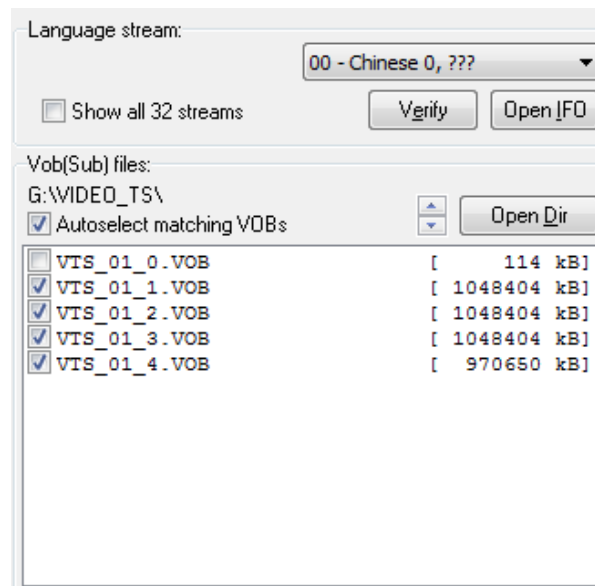
各位老師 同學
這個非常歡迎今天一起來
這個聆聽王老師
他對自己著作的詮釋
我想這個機會是非常難得的
我覺得從臺灣這個小說史來看
王老師的這個作品
我覺得真的他被詮釋的以及被評價的
我覺得真的是遠遠不夠

2.2 外掛字幕的影片

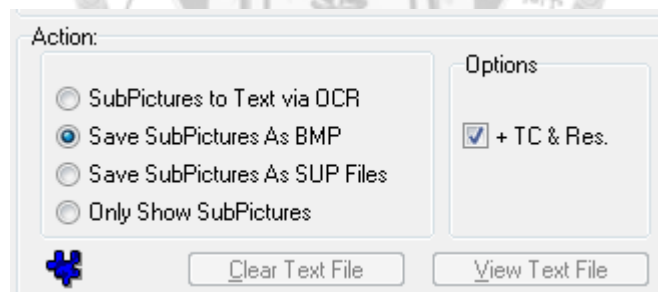
a. 安裝 subrip 並開啓。

b. 點選 File->Open VOBs，在開啓的視窗中再點選 Open IFO，選擇 DVD 中的 IFO 檔案，通常為較大那個 IFO 檔案。

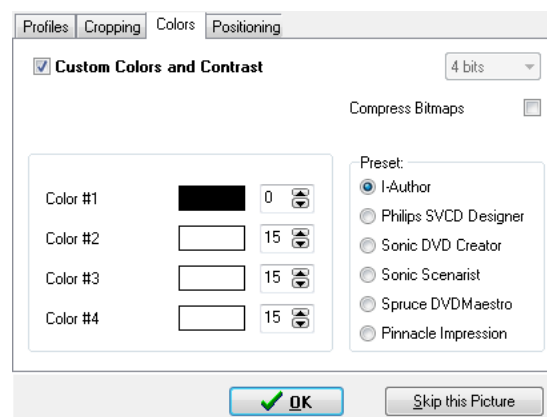
或者也可以點選 Open Dir，再選擇 DVD 中的 vob 檔案。



c. 右邊選擇 Save Subpictures as BMP 以及勾選 + TC & Res.，然後點選 start。



d. 選擇儲存資料夾及輸入檔名之後，在開啓的視窗中選擇 I-Author，以及左方設定字幕的顏色，設定成字幕看起來清楚，筆畫分明的顏色，點選 OK。



e.在下方的子視窗中，點選 Output Format->Set Output Format，在 text Formats 中找到 SubRip(*.srt)，並勾選 Extended Format，再按 Convert to this Format。

f.等程式將字幕讀取完之後，點選子視窗的 File->Save as，將 srt 檔案與圖檔存在一起，這種 srt 檔案是沒有文字用圖檔代替。

g.開啓 esrXP，點選 File->Open...，開啓剛剛存好的 srt 檔案。

h.一樣另存 OCR Image，讓 OCR 軟體辨識。

3.OCR 辨識圖片

a.將 BMP 檔轉成 TIFF 檔。(直接小畫家另存即可)

b.用 MODI 開啓 tiff 檔。

c.點選工具->傳送文字到 word...，選取儲存資料夾，等 OCR 辨識完，會自動開啓 word 來顯示辨認結果，不過這個檔案其實是 html 檔，所以之後要開啓可用其他程式。

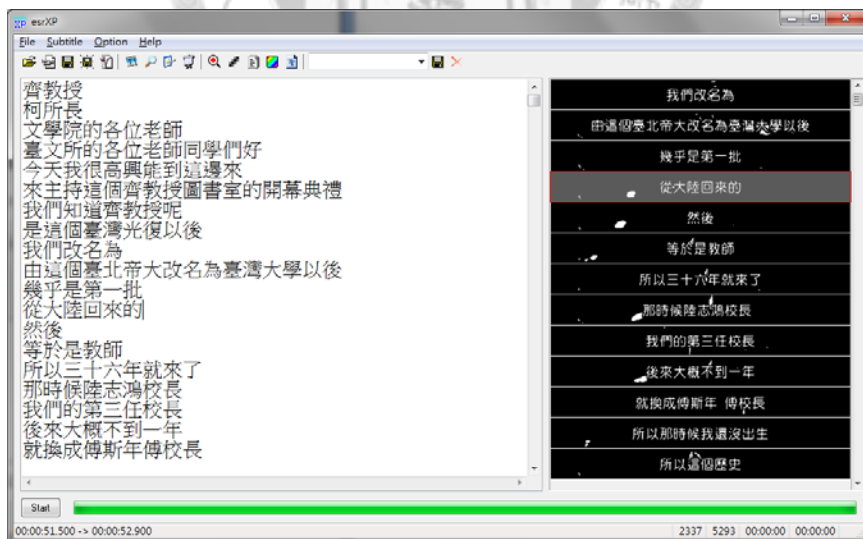
4.處理辨認結果及輸出字幕檔

a.將辨認結果都存在一個 txt 檔案中。

b.修改 correct.php 中第 19 行讀取的檔案爲此 txt 檔案。

c.執行 correct.php，此處使用瀏覽器執行，畫面上會顯示處理後的文字。注意要將 txt 檔以及 replacement.txt 跟 correct.php 檔案放在一起。

d.將處理後的文字貼回 esrXP 主視窗中的文字區域。



e.由於使用 MODI 的關係，辨認結果的斷句並不完美，因此需要手動斷句。如果使用其他 OCR 軟體，應該可以直接貼上結果。

f.點選 esrXP 中 File->save as，另存成 srt 檔案(非 srt with bmp)。

5.匯入系統

a.將演講手冊處理成以下格式並存成 txt 檔。

人名:

演講內容

人名:

演講內容

b.將 srt 檔案放到 rdy 資料夾中，txt 檔案放到 txts 資料夾中，取名相對應的檔名(example.srt 跟 example.txt 表示同一個影片的字幕檔以及演講手冊)。如果 srt 檔案沒有相對應的 txt 檔，則在 importdata.php 的第 83 行及第 176 行，在無逐字稿檔案中加入 srt 檔案判斷。

c.在 DVD_titles.txt 中加入新增的字幕檔以及 DVD 標題和主講者姓名(重要)。

d.執行 importdata.php，需要跟 rdy 資料夾及 txts 資料夾放在一起，此程式因為跑 LCS 會花很多時間以及記憶體，得使用 cmd line 來執行，php 的執行時間也得設定。

註:

設定:

1.若要使 Firefox 可以讀取 webm，apache 設定檔中要加上 AddType video/webm webm。

2.php.ini 設定 memory_limit upload_max_filesize post_max_size 皆 800M

max_input_time = 3600

max_execution_time = 36000