

行政院國家科學委員會專題研究計畫 成果報告

運用視訊處理技術於網路英語學習(2/2)

計畫類別：個別型計畫

計畫編號：NSC91-2213-E-032-034-

執行期間：91年08月01日至92年07月31日

執行單位：淡江大學資訊工程研究所

計畫主持人：郭經華

計畫參與人員：宋立群、吳炳煌、藍永孝、何啟弘、陳祐凡、郝維華、陳世彰

報告類型：完整報告

報告附件：出席國際會議研究心得報告及發表論文

處理方式：本計畫可公開查詢

中 華 民 國 92 年 10 月 29 日

行政院國家科學委員會補助專題研究計畫成果報告

運用視訊處理技術於網路英語學習(2/2)

計畫類別：個別型計畫

計畫編號：NSC 91-2213-E-032-014

執行期間：91年08月01日至92年07月31日

計畫主持人：郭經華

共同主持人：

計畫參與人員：宋立群、吳炳煌、藍永孝、何啟弘、陳祐凡、郝維華、
陳世彰

本成果報告包括以下應繳交之附件：

赴國外出差或研習心得報告一份

赴大陸地區出差或研習心得報告一份

出席國際學術會議心得報告及發表之論文各一份

國際合作研究計畫國外研究報告書一份

執行單位：私立淡江大學

中華民國 92 年 10 月 29 日

行政院國家科學委員會專題研究計畫成果報告

計畫編號：NSC 91-2213-E-032-014

執行期限：91年08月01日至92年07月31日

主持人：郭經華 淡江大學資訊工程研究所

計畫參與人員：宋立群、吳炳煌、藍永孝、何啟弘、陳祐凡、
郝維華、陳世彰 淡江大學資訊工程研究所

一、中文摘要

近幾年來我們致力運用多媒體技術與網際網路進行互動式的英語學習，在智慧型互動式英語學習整合系統中，我們企圖有效提昇學生學習英文基本的聽、說、讀、寫能力。在我們所發展的系統中包括了(1)互動式線上英語寫作系統、(2)多媒體創意即席寫作系統、(3)情境式影片英語學習系統、(4)智慧型閱讀系統及(5)同步式寫作環境。透過這五個子系統的整合，我們期望為英語學習環境開創一個新的里程碑。在充分運用多媒體、網際網路及語言文字分析的技術於英語學習環境的同時，我們希望學生能在愉悅的學習過程中紮實地成長。智慧型互動式英語學習系統發展至今輪廓已漸成形。在此計畫中，我們以『情境式影片英語學習系統』為主要設計目標。在既有的基礎上，整合多媒體技術、學習科技以及計算語言學的專業領域知識來設計網路型互動式英語學習環境，於此計畫中我們完成了以語法為基礎的視訊資料搜尋技術，藉此，使用者更方便的搜尋出所欲取得之視訊資料。

關鍵詞： 隨選視訊、電影自動分段、線上英語學習、語法搜尋

Abstract

Learning in an efficient manner is a very important issue. The key of the accomplishment of this objective depends on the integration of the expertise from information technology, learning technology, and domain knowledge. In this proposal, we will design and implement a web-based

authentic English learning environment to illustrate the effectiveness of our approach. In our previous work, we have successfully built a system to support interactive English writing on the Internet. Many schools and students currently use the system. Here, we focus on the design of the authentic learning environment. As it is well known, the lack of authentic environment is the main obstacle for non-native speakers. The developed environment intends to improve learners' basic language skills such as listening, reading, speaking and writing by using the creation of networked learning environment. We make use of movies to create a virtual environment for learning. In this project, we have completed a video retrieval technique based on syntax research. As a result, users demand can be satisfied even better.

Keywords: VoD, Movie segmentation, On-line English learning, video syntax search

二、緣由與目的

在情境式影片英語學習系統中，我們提供一個可以讓學生在欣賞英語電影的同時，提昇自己的英語聽力及熟悉正確的情境對話及習慣用語的環境，我們提出了一個新的看電影學英文的學習方式。有別為於市面上販賣之看電影學英文的CD-Title，只單純提供電影播放及字幕查詢的功能；我們設計一套可以讓老師及教材編輯者可以編輯包含電影片段的多媒體教材的系統，老師及教材編輯者（以下簡述為老師）可以依電影主題或英語教學主題來編輯教材，這樣的多媒體教材可以包括

了文字、圖片、聲音、電影片段，並且可以透過寬頻網路（如 ADSL 或校園網路）播放。

三、討論與結果

老師們在這個系統中設計教材時，可以選擇電影中的片段做為教材中的素材。（如圖一）然而由於電影資料庫中有非常多的電影，而且每一部電影的播放時間也都很長，若是沒有設計良好的篩選工具，老師們將難以從龐大的電影資料庫中選取適當的對白或片段做為教材的素材。



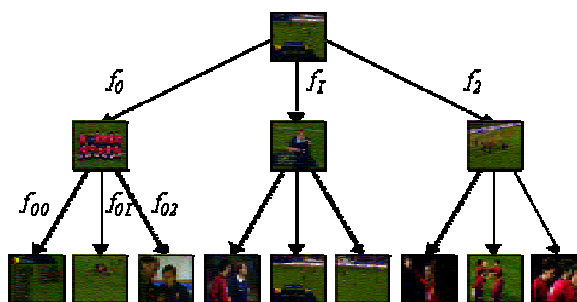
圖一 使用線上教材編輯環境編輯線上教材

在本系統中，本系統提供了數個方法讓老師們可以在電影資料庫中選擇適當的電影片段，以下為前期本系統中已實作的方法

1. 利用關鍵字查詢電影對白（圖二）。
2. 利用影片時間及對白擷取長片段。
3. 關鍵字原型查詢，如查 understand，同時查詢 understanding, understands 和 understood。
4. 利用 WordNet 擷取關鍵字的同義字、廣義字、狹義字、反義字等相關字，連同關鍵字一同查詢，已查出代表關鍵字概念的影片對白。
5. 利用一大型語料庫如 BNC Corpus，擷取 WordNet 無法定義的相關字，如使用者如要查詢 hospital，則同時查詢 doctor 和 nurse 會使影片出現醫院的機率增高。
6. 利用語意關聯的段落將影片分解成一階層式結構，以利查詢語意片段（圖三）。



圖二 利用關鍵字查詢電影對白



圖三 階層式結構

以上的方法皆集中在使用關鍵字查詢，然而使用者，尤其是教師，在選擇影片對白或其他語料時，希望有其他的查詢方式以取得較適合教學的影片。在台灣的英語教學上，語法是很重要的一部份，然而在一般的語料搜尋環境裡，並無語法搜尋的機制。所以本系統試圖在所有的語料上，加上一語法搜尋介面，當然也可以利用在影片搜尋上，以下就提出如何實作此機制的步驟及實作後的擷取畫面。

本系統實作的語法搜尋機制目前能運用在以下的語法搜尋：

1. 關鍵字組合：任意數量的關鍵字，且不必相鄰，例如 no matter、as ... as。
2. 關鍵字和詞類的組合：任意數量的關鍵字及詞類，且不必相鄰，例如 as Adj as、How to Verb。
3. 詞類組合：任意數量的詞類，且不必相鄰，例如 WH Adj。

其中又以第一及第二種較為實用。

基於以上的語法搜尋原則，本系統使用 regular expression indexing 的方法來實

作，以提供快速且有彈性的查詢機制。

Regular expression indexing 已被多位學者證實其實用性，茲將我們實作的步驟列舉如下：

1. **前處理**：在本系統中由於要提供詞類查詢，所以需將原始語料加上詞類的資訊，我們運用在自然語言處理上常使用的方法來實作 Part of Speech tagger：以一已加註詞類的語料庫，如 BNC Corpus，及馬可夫模型來訓練，再將訓練成的 tagger 將原始語料加註。又為了利於以後查詢結果的呈現，將句子儲存成 XML 的格式，如下例：

原始：I will do my best.

詞類加註：I/PNP will/VM0 do/VDB my/DPS best/NN0 ./PUN

XML 格式：<s><w POS="PNP">I </w><w POS="VM0">will </w><w POS="VDB">do </w><w POS="DPS">my </w><w POS="NN0">best </w> <w POS="PUN">.</w></s>

2. **索引製作**：直接運用現成的 regular expression engine 查詢所有系統中的語料，當語料庫非常龐大時，反應時間會非常久，大大降低其實用性。所以本系統利用分析 regular expression 以擷取其中的關鍵字串，在本系統中即為關鍵字及詞類。再利用關鍵字串搜尋出候選物件，以加速搜尋速度。而利用關鍵字串搜尋就需要使用索引以增快速度，由於本系統選定 regular expression 作為搜尋語法，故使用 Suffix array 作為製作索引的方法，茲利用下列說明 Suffix array 所產生的索引形式：

原始：

1. This is an example.
2. This example is for suffix array.

索引 (Length of suffix >=2)：

Th	(1, 1), (2, 1)
Thi	(1, 1), (2, 1)
This	(1, 1), (2, 1)
is	(1, 2), (2, 3)
an	(1, 3)
ex	(1, 4), (2, 2)
exa	(1, 4), (2, 2)

3. **搜尋**：在以上步驟都完成後，就可利用現成的 regular expression engine 搜尋候選物件了，取出結果後，利用 XML DOM Parser 顯示單句及標示出 regular expression 的關鍵字串，特別需要處理的部分只有當關鍵字串是詞類時，也需要標示出詞類所代表的字。搜尋介面則與關鍵字版本相同，如（圖四）。



圖四 語法搜尋結果

四、計畫成果自評

計畫之執行，有非常豐碩的成果，除感謝國科會之贊助外，亦是淡江電腦與網路實驗室師生努力的成果。本計畫中所研發之所得成果亦以發表於國際研討會[7] [13] 以及投稿至國際期刊[1] [6]。我們將以我們已經完成的系統功能為基礎，繼續研究與開發以電影內容為素材設計網路多媒體英語教學教材所須要的相關技術。我們預計在 2004 年 1 月份時，將我們的整個系統設計理念及技術寫成論文投稿至 International Conference on Applications of Natural Language to Information Systems 以及相關領域的期刊上。希望國科會可繼續贊助我們在這方面的研究。

五、參考文獻

- [1] Chin-Hwa Kuo, David Wible, Meng-Chang Chen, Li-Chun Sung, Nai-Lung Tsao, and Chia-Lin Chio, "The Design of an Intelligent Web-based Interactive Language Learning System," *Journal of Educational Computing Research*, Vol. 26, 2002.
- [2] Chin-Hwa Kuo, David Wible, and Chia-Lin Chio, "A Synchronous EFL Writing Environment for the Internet," *Journal of Universal Computer Science*, Vol. 7 No. 3, pp. 240-253, March 2001.
- [3] Chin-Hwa Kuo, David Wible, and Nai-Lung Tsao, "On the System Design of Web-based English Writing Environment and Learner Corpus," *IEICE Transactions on Information and Systems*, Vol. E84-D, No. 8, August 2001.
- [4] David Wible, Chin-Hwa Kuo, Feng-yi Chien, Anne Liu, Nai-lung Tsao, and Hong-Sheng Chui, "A Web-based EFL Writing Environment: Exploiting Information for Learners, Teachers, and Researchers," *Computers & Education*, vol. 37, pp. 297-315, 2001.
- [5] David Wible, Chin-Hwa Kuo, Anne Liu, and Nai-Lung Tsao, "An online Writing Platform for Language Teachers," *Journal of Universal Computer Science*, Vol. 7 No. 3, pp. 278-289, March 2001.
- [6] 郭經華、衛友賢、陳孟彰, "運用多媒體於英語學習" 資訊與教育, 2001年4月。
- [7] Chin-Hwa Kuo, David Wible, Nai-Lung Tsao, Tzu-Chuan Chou, "Active Writing Through Images: The Designing of a Web-based English Writing Environment," *AACE/WebNet 2001*, Florida USA, October 2001.
- [8] Shu-Ling Liao, Chin-Hwa Kuo, Tzu-Chuan Chou, and Chen-Chiou Chiu, "The Design of Authoring Tools for Instruction and Content Management on the Web-Based Learning," *AACE/WebNet 2001*, Florida USA, October 2001.
- [9] Chin-Hwa Kuo, David Wible, Meng-Chang Chen, Li-Chun Sung, Nai-Lung Tsao, and Chia-Lin Chio, "Design and Implementation of an Intelligent Web-based Interactive Language Learning System," *IEEE/ICME2001*, Tokyo Japan, August 2001.
- [10] David Wible, Chin-Hwa Kuo, Feng Yi Chien, Nai-Lung Tsao, "Automating Repeated Exposure to Target Vocabulary for Second Language Learners," *IEEE/ICALT*, Madison Wisconsin, August 2001.
- [11] Chin-Hwa Kuo, David Wible, Chih-Chiang Wang, "The Design of a lexical Difficulty Filter for Language Learning on the Internet," *IEEE/ICALT*, Wisconsin USA, August 2001.
- [12] Chin-Hwa Kuo, David Wible, Tzu-Chuan Chou, "CanAM- A Web-based Multimedia Authoring and Management Tools for CALL," *IEEE/ICALT 2002*, April, 2002.
- [13] Chin-Hwa Kuo, Tzu-Chuan Chou, Nai-Lung Tsao, and Yung-Shiao Lan, "CanFind - A Semantic Image Indexing and Retrieval System," *ISCAS2003*.