

電子文件註記系統之探討

江柏寬 楊亨利 政治大學資管系

摘 要

目前線上學習的趨勢日漸明顯，學習者隨時隨地都可以從網路上取得電子文件教材，本研究之目的是設計一個便於使用的電子文件註記系統，並將其應用到自我學習與知識分享上。本系統與一般註記系統最大的不同點在於註記檔案和教材文件乃是分開儲存，隨後在呈現時利用錨點演算法將兩者結合。如此一來註記的檔案將會比較小，利於分享與傳遞，甚至可以根據學習者的需求同時呈現數份註記。而且由於使用錨點演算法的關係，就算教材文件經過修改，註記仍能在文章中找到適當的位置。

關鍵詞：註記、後設認知、電子文件、錨點

Annotation System in Electronic Documents

Po-Kuan Chiang, Heng-Li Yang

Dept. of Management Information Systems, National Cheng-Chi University

ABSTRACT

The e-learning environment has become more and more popular. Students can get digital documents from Internet anytime and anywhere. The purpose of this research is to develop a usable annotation system that can be used while students self-learn or share knowledge. The main characteristics of this annotation system are that the annotation files are stored separately from documents, and an anchoring algorithm is applied to integrate them. Therefore, the annotation file will be smaller and be easily transmitted. It can even present several annotations at the same time according to the user's request. If the document is modified, we can also find the right-position of annotations in the modified document owing to the anchoring algorithm.

Keywords: Annotation, Meta-cognition, Electronic document, Anchoring

1.序論

現代科技日新月異，網際網路的使用越來越頻繁，生活上的各種面向都跟網路脫不了關係。而線上學習的趨勢日漸明顯，網路無遠弗屆的覆蓋率和電腦系統隨時待命的特性，使得學習的界限越來越模糊，學習者隨時隨地都可以利用簡單的設備連上網路取得電子文件。電子文件和傳統紙本相比，在取得、傳遞和保存上的確具有顯著的優勢，但是學習者在閱讀電子文件時很難進行註記行為，而註記卻又是傳統上普遍使用的學習策略，這種學習模式的落差造成線上學習的障礙。所以，本研究希望能設計一個便於使用的電子文件註記系統，解決學習者的迫切需求，同時藉由本系統的測試找出適用於電子文件的註記模式。

2.文獻探討

註記是學習中常見的輔助行為，以下文獻，我們將試著由心理學「後設認知」理論，來探討註記對於個人學習與團體學習過程的助益。除此之外，我們將瞭解目前的註記系統與其演算法之可能改進空間。

2.1.註記

學生為何在學習的過程中會有「註記」的行為？Slotte和Lonka等學者認為有兩個主要的理由：其中之一是學生相信在做筆記的過程中能夠幫助自己學習；另一個理由則是學生認為註記行為所產生的「筆記」，在日後複習的時候會非常有用〔17〕。Hidi 和 Anderson 指出將文章中具有代表性資訊的內容做摘要標記，對理解該文章有相當大的幫助〔12〕。Quade認為，在電腦化的教學環境下，利用電腦做筆記的學習效果比在紙筆上做筆記的效果要來的好〔15〕。Ovsianikov等學者把『註記』的種類分為以下幾種類型：標記(Mark up)、寫在頁邊(Write on margins)、寫在頁面上方(Write at the top)、分開寫(Write separately)、寫在行與行之間(Write between lines)。在這當中，以Mark up的功能最常被使用，其次是寫在頁邊、寫在頁面上方、分開寫、寫在行與行之間。他們把『註記』的用途分為以下三方面：記憶(Remember)、思考(Think)、闡明內容(Clarify)〔14〕。

2.2 註記系統

目前國內已經有一些關於註記系統的研究，例如陳守賢[4]曾經針對在網路環境上的電子文件設計一套註記系統，能夠閱讀與註記HTML格式的網頁，方便在網路上進行學習。而許政穆與林建丞[5]則是設計了一套適用於線上閱卷的註記系統，將註記用於測驗的評量與互動。邱續瑩[3]則將註記系統結合影像多媒體，讓老師將教材內容與授課時對教材的導引動作完整的記錄下來，老師可以利用註記與影像對教材做說明。另外，谷川[1]則是針對了不同的學習環境設計了一套「無縫式」的註記系統，讓學習者可以利用電腦與手機來閱讀教材，延伸學生線上學習的時間與空間。吳玟玲[2]則針對註記系統在協同學習、同儕互評及回饋機制上的功效進行探討。這些研究所注重的方向各有不同，但

是這些註記工具對於註記和文件分開儲存時會遇到的問題並沒有深入的探討，事實上，文件很有可能被修改過，以致於註記無法在正確的位置呈現。本研究希望能來解決這個問題。

2.3.後設認知

最先使用後設認知 (Meta-cognition) 一詞的學者是Flavell，他認為後設認知是一個人對自己的認知系統的內省知識(Inspective Knowledge)〔11〕。此外，另一位對後設認知具較大貢獻的學者Brown認為後設認知是對自己知道些什麼的瞭解 (Knowing about Knowing) 以及瞭解如何去知道 (Knowing how to Know) 〔8〕。張春興認為後設認知是指面對某種資訊予以處理時，如在認知上超過「知其然」的程度，進而達到「知其所以然」的地步時，即稱為後設認知〔6〕。

「註記」記錄了學習者當時思考模式的歷史軌跡，學習者可以藉由瀏覽註記評估自己對於學習議題的認識程度，同時審視自己當時的學習策略。所以說，註記是一個讓學習者可以「瞭解自己知道什麼」，以及「瞭解自己如何去知道」的工具，即是表示註記具有增進後設認知的功能。

另外，註記不止對個人有所幫助，藉由註記的分享，不但可以達到知識交流的目的，對團體中的其他成員的學習技巧亦有正面效益。Kuhn將個人內心的推理與思考概念化為一種內心的辯論(internal argument) 〔13〕，而Andersen認為同儕之間的互動與交流乃是內心辨證的延伸，這樣的互動模式其實是一種分散式的後設認知(Distributed Meta-Cognition)，藉由分享個人的後設認知知識，將個人的學習技巧傳授給團體中的其他成員，使成員的後設認知也隨之成長〔7〕。

2.4.錨點演算法

Brush 與 Bargeron [9,10] 提出了一種演算法，能夠用於註記的定位。他們的假設是註記和文件是分開儲存的，而且文件本身的內容有可能被更改，所以需要有一種方式來找到新的註記位置。步驟如下：

1. 在註記時同時紀錄以下資訊，用於將來註記與文件的結合
 - (1) 註記在文件中的偏移量(和文件開頭的字元距離)
 - (2) 註記的字元長度
 - (3) 註記的「開頭」、「結尾」與「上下文」的文字內容資訊
 - (4) 註記包含的關鍵字。演算法會尋找在文章中出現頻率最少的字作為關鍵字，關鍵字至少會有三個字，但如果註記本身長度太短，則會記錄所有的被註記文字。
2. 日後將文件與註記結合時必需先確定註記在文件中的位置：
 - (1) 在文件中尋找關鍵字，找到包含關鍵字的區段之後將其列為「候選錨區」，由於候選錨區可能不止一個，必須進一步做篩選。藉由比較各關鍵字之間的相對字元距離，和修改之前的原文差距越小，該候選錨區的「信賴分數」就越高。
 - (2) 演算法會從候選錨區開始，向前向後尋找和原文相符的註記開頭與結尾。同樣的，演算法也會計算新找到的開頭與結尾與關鍵字之間的字元距離，越符合原本距離

者，分數越高。

- (3) 演算法會比較候選錨區的「偏移量」和「長度」，和原文越符合者信賴分數越高。
- (4) 尋找候選錨區的前後文資訊，如果找到符合的前後文，則增加其信賴分數。
- (5) 以得到最高信賴分數的候選錨區為新的註記位置。

這一個錨點演算適合於較長的註記，因為較長的註記在選擇關鍵字上比較容易，但是也因此這個演算法對於極短的註記卻比較不適用。而且該演算法所使用的「偏移量」乃是註記在整篇文章中的字元位置，在這種情況下，如果文章被修改過的話，有可能所有註記的偏移量都會改變，似乎不是很理想的作法。

以上文獻顯示從「後設認知」的來說，註記對於個人學習過程的記憶與思考有很大的助益，而且分享註記能夠增進團體中同儕的學習技巧。在系統建置上，我們認為將文件與註記分開儲存有利於分享，但是在分享註記的過程中，被註記的文件有被修改的可能，團體中各別的學習者所持有的文件有可能略有不同，所以需要有一個有效的方式來將註記與文件作適當的結合，而文獻中的演算法來還有改良空間。

3. 研究架構

3.1. 電子化學習中與註記相關之思考

3.1.1 自我學習：

註記系統最基本的功能就是對於電子化的教材或文件作「筆記」，在閱讀中隨時記錄下自己的感想與心得。這個紀錄的過程對於學習有很大的幫助，可以讓使用者能夠在日後快速複習該篇文件，迅速回復當時的學習記憶。除此之外，對文件進行註記還有增進學習技巧的效果。根據「後設認知」理論，我們認為當使用者藉由「做筆記」的行為來紀錄當時的學習心得時，稍後學習者可以藉由省視自己的學習歷程，針對其學習過程作內在的省思，了解到自己獲得了什麼知識，以及這些知識是如何獲得的，達到控制、引導及改進學習心智歷程的目的，讓學習者自我改善其學習技巧。

3.1.2 協助測驗：

註記系統不只可在自我學習的時候才被使用，我們認為在線上測驗和電子化測驗的過程中也需要使用到註記功能。首先，學生可以使用註記來對考題作答，或是在考試中記錄自己運算和思考的過程。而考試結束之後，教師也可以利用註記的方式來「批改」考卷，藉此改善傳統線上考試教師很難對考試的結果進行「回饋」的缺點。教師可以藉由註記功能來對學生的試卷進行批改與建議，而學生同樣也能利用註記系統和老師請教尚未澄清的疑問，這樣的互動模式才能達到考試真正的效果。

3.1.3 分享與討論：

註記之後的文件不只對於原註記者有所助益，同時也能幫助他人減少原本閱讀時會遇到的障礙，快速掌握該篇文章的意涵。其次，文件的註記可以作為和同儕討論互動的媒介，利用這項功能來為彼此解惑或是交換學習的心得。由於人際互動所帶來的參與感

和成就感，不單單可以提升學習效果，還可以激發學生的學習意願、促使其自動學習。另外，根據後設認知的文獻內容，我們認為藉由分享註記的方式來傳播學習心得可以增進團體的後設認知，也就是說，經由分享個人學習的技巧，能讓團體中的其他成員也獲得這項技巧，將個人的後設認知推廣到團體中，改善整個團體的學習能力。

3.1.4 文件與註記分離：

在儲存註記的時候，我們應該讓註記檔案能和文件分開，這麼做的用意是為了不讓原本的文件被修改，保有原文件的再用性。此外，分開儲存還有另一個好處，那就是方便分享。在註記分離的模式下，學習者可以只傳遞註記給其他的夥伴，不必連文件也一起傳送，縮小傳輸時的資料量。另外，在閱讀的時候可能會需要同時檢視多份註記，這時候學習者只需要開啟一份文件檔，再加入所需要的註記即可，不必同時開啟多份文件。

由於文件與註記分開儲存，原本的文件內容還是有可能被修改，但是我們還是必須將註記顯現於適當的位置，所以我們使用一種演算法來決定註記的錨點，讓註記和文件能完美契合。

3.2. 註記系統功能需求分析

考量到學習者對註記的需求和系統規模的限制，我們列出了應該具備的功能如下：

1. 使用者記錄：由於註記不只供個人使用，更可能被用在分享、討論甚至是測驗上，所以必須具備紀錄使用者相關資料的功能，如使用者姓名、進入與離開系統的時間。
2. 劃記功能：在閱讀的過程中時常需要將重點和關鍵句做明顯的標記，本系統提供了一些改變字型呈現模式的功能，如底色方塊、劃線、變色、粗體、斜體等，讓電子文件能像實體文件一樣被劃記。
3. 插入文字：提供使用者能將自己的心得、意見和疑問加註在文件之內的功能。註記有很多不同的呈現方式，例如「頁邊註記」是讓使用者在文章呈現的同時，能在旁邊撰寫相關的意見。「記號點」是在文字之間插入一個明顯的小標記，其包含的文字內容平常是隱藏的，在滑鼠移到標記上方時，才會浮現註記的文字框。「文字方塊」則是將文字放置於方塊之中，再將方塊自由的置放於文件的任何位置。
4. 搜尋功能：可以用輸入關鍵字的方式來搜尋字詞，也可以將註記的類型列為搜尋條件。
5. 同步顯示功能：考慮到使用者可能會需要同時觀看多人的註記，或是一邊瀏覽別人的心得，一邊撰寫自己的心得。所以系統必須要能夠將不同作者的註記同時呈現，方便使用者參考、比較與評論。

3.3. 系統架構

根據功能分析的結果，本系統將採用以下架構：

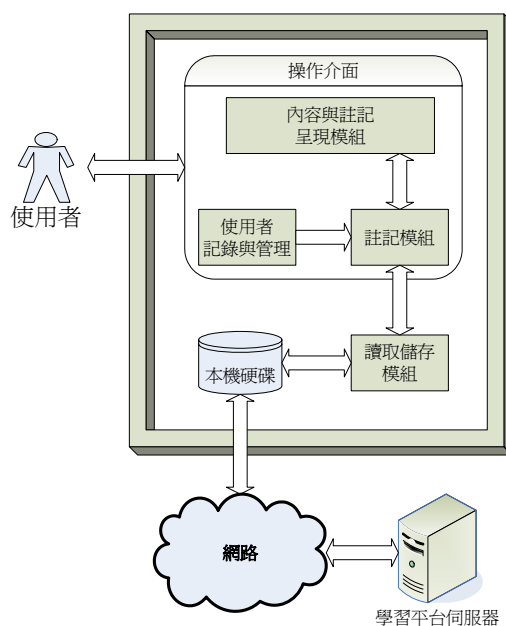


圖 1.系統模組圖

1. 在系統架構中的伺服器主要是提供教師分派教材、考卷，以及同學彼此分享註記的平台，所以至少必須隨時有一個學習平台伺服器運作中，負責管理檔案、提供資料交換。
2. 內容與註記呈現模組負責呈現文件與註記結合的畫面。
3. 使用者記錄模組會記錄使用者的姓名資料及當次進入時間和離開時間，隨後傳遞給註記模組，將使用者資料寫入註記檔案中。
4. 註記模組所負責的工作主要有三項，其中之一是滿足使用者對文章進行註記的需求，提供介面讓使用者選擇各種劃記模式和插入文字註解。註記模組的第二項功能則是依循錨點演算法，在產生註記時記錄其相關資訊，以便在下次呼叫的時候能找到正確的註記位置。註記模組最後一項工作則是在讀取註記檔案的時候，藉由錨點演算法來決定註記適當的位置，將文件和註記結合。

3.4.文件與註記錨點

由於文件與註記是分開儲存的，所以在呈現的時候必須重新將其結合為一，這個時候就需要有一套規則來決定其正確的位置，尤其是文件的內容有可能會被修改，在這種情況之下更需要一個演算法來找到原本被註記的內容。本演算法是參考 Brush and Bargerion 的演算法修改而成，原本文獻中的演算法較適用於文字較多、長度較長的註記，因為其註記錨點的定位十分依賴關鍵字，所以註記越長，關鍵字的選擇就越有彈性，也因為如此，該演算法對於短的劃記比較不適合。另外，該演算法並不重視註記與其所所在的段落之間的關係，反而比較重視在文章中的位置。這樣的作法似乎有改進的空間，以下的演算法將會對這些部份進行修正，和文獻中所提的演算法最大的不同點在於本演算法認為正確的註記必定會位於正確的段落中，所以在尋找錨點時，段落的可信度為評估錨點可信度的重要依據。

在儲存註記檔案之前，我們會先使用雜湊演算法對教材文件進行運算，紀錄文件的內容特徵，我們所採用的演算法是 MD5 演算法，利用 MD5 可以對任何長度的檔案資料進行運算，得到一組固定長度為 128 位元的 “fingerprint” (check sum)。將來我們可以利用這組資料來辨認文件是否有被修改過，如果對文件進行雜湊的結果改變的話，即代表該文件已經被修改過〔16〕。使用 MD5 所產生的 fingerprint 其重複率低於百萬分之一，所以可信度相當高。

註記檔案會記錄註記的型態，並且儲存與定位相關的資訊如下：

1. 從註記的兩端向外選擇連續且獨一無二的文字作為註記的開頭與結尾。
2. 註記包含的關鍵字。演算法會尋找在文章中出現頻率最少的字作為關鍵字，關鍵字至少會有三個字，但如果註記本身長度太短，則會記錄所有的被註記文字。
3. 註記的長度。
4. 各段落的關鍵字以及關鍵字之間的字元距離，關鍵字以整篇文章中出現次數最少次的文字為優先。
5. 註記位於哪一個段落
6. 註記與段落開頭的字元距離

本演算法決定註記錨點的規則如下：

1. 使用 MD5 演算法來確認文件是否有經過修改，如果文件沒有改變的話，直接使用原註記的位置與長度來呈現註記後結束。如果發現文件被修改過，才進入第二步驟。
2. 搜尋註記的開頭、結尾與其關鍵字，如果發現找到唯一且完全符合的區段，則決定其為註記所在的位置。
3. 如果找不到完全符合的區段，將所有部份符合的區段都列為「候選註記區」，隨後評估其關鍵字數量、開頭結尾與註記長度，和原註記資料越接近的候選註記區將得到越高的「信賴分數」。
4. 藉由關鍵字搜尋，找到各段落所在位置。如果發現文章中出現兩個段落都具有相同關鍵字的話，兩者都將被列為「候選段落」，再比較關鍵字之間的相對字元距離，和原文差距最小者，信賴分數越高。此為「候選段落」的信賴分數。
5. 最後，將「候選註記區」與「候選段落」的信賴分數加總。其規則如下：100%的信賴分數之中，有 60%為「候選註記區」分數，另外 40%則是由「候選段落」來決定。「候選註記區」的分數加上其所屬「候選段落」的分數，加總之後取得最高分的「候選註記區」將成為新的正試註記區。

4.系統雛形

根據以上的文獻以及架構探討，本研究的目標是發展一註記系統，可以使用瀏覽器開啟且不需要安裝任何其餘的程式便可以使用，可以離線使用，在需要傳遞資訊的時候也能配合線上工作。現今的作業系統一般皆附有瀏覽器，不需要使用者另行安裝，如此可以減少使用上的不方便。目前需要註記之電子文件需先轉成 HTML，整個註記系統環

境包含 HTML、VML、Java Script、在 .NET Framework 架構下的 ASP.NET、以及 IIS 伺服器及伺服器端的 SQL Server 資料庫而開發的，以下將介紹本研究所開發的工具及系統所提供的註記功能：

在開始註記之前，使用者必須先使用系統的 CGI 功能從伺服器下載所需要的教材與註記。隨著使用情境的不同，下載的資訊也不同。例如，當學習者單純要瀏覽教材的時候，下載教材之後即可開始閱讀，而在閱讀的過程中利用本系統提供的功能來進行註記。如果學習者是要利用本系觀摩看他人的註記或是觀看教師對作業、測驗的評分批改的話，那他除了教材之外還必須下載他人的註記，在系統中結合兩者以觀看其內容資訊。當使用者想要進行自己的註記活動的話，首先，使用者必須要先輸入自己的個人資料，系統也會自動記錄當時的操作時間。然後學習者即可開始進行註記。註記系統本身可以同時顯示數個不同作者的註記，也可以將他人註記加入自己的註記(但是不能修改別人的註記)。所以說，即使正在觀看他人註記，也可以在其上加入自己的註記，然後儲存成同一份檔案，這樣的疊加儲存功能可以幫助與同儕的討論溝通。當完成註記之後，可以再度使用系統的 CGI 功能將資料上傳。

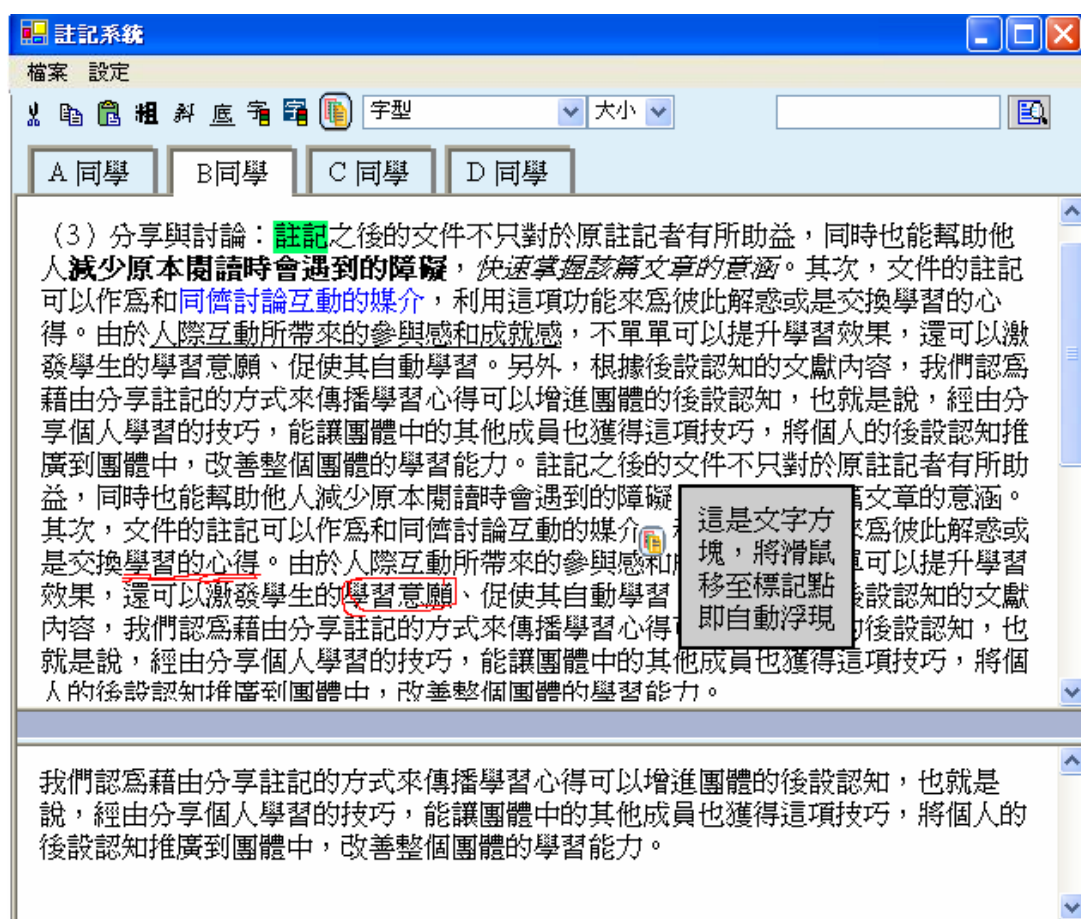


圖 2.系統雛形介面

在圖 2 中顯示了系統的介面，其中包括了許多註記相關的功能，如將文字的背底加上色彩方塊，有如使用螢光筆畫記。改變文字的呈現樣貌，使其呈現粗體、斜體或是在

下方附加底線。也可以選擇直接改變文字的顏色。另外，使用者可以在文章中任何位置加入文字方塊，藉由在文章中插入一個標示點，如果將滑鼠游標移至標記點上時，隱藏的文字方塊就會浮現。如果覺得附加的文字方塊版面太過狹小，不夠容納想紀錄的資訊的話，也可以使用頁面下方的註記區，可以容納大量的文字。除了改變文字型態的註記法之外，藉由 VML 的幫助，使用者也可以利用滑鼠及鍵盤在畫面上留下向量圖形，讓學習者能自由的劃記文章。此外，除了註記功能之外，版面上也有附加搜尋文章內容的功能，可以搜尋輸入的字句，也可以搜尋某種型態的註記，例如搜尋被「色塊」標記的「註記」二字。在畫面上方有下拉式選單，藉由選擇「檔案」選項，使用者可以讀取文件和註記，以及儲存自己所編輯的註記。選擇「設定」選項則是進行相關的功能設定，如調整「色塊」及「變色」功能的顏色。

5. 結論

本論文主要是希望能提供一個適合學習者使用的註記系統，提升線上學習者的學習效率，利用註記功能提供教師與學生進行測驗與作業的批改與回饋，同時藉由分享註記的方式來促進團體的互動，經由人際互動所帶來的成就感以及分享後設認知，增進學習者的學習意願及學習技巧。目前本系統尚未發展完備，在與離型使用者訪談調查之後，期望能設計出更適合現代線上學習的註記模式。目前本系統缺少多媒體的呈現功能，對於教材文件的檔案類型也有所限制，如何增加系統能夠辨認的檔案類型，以及增加影像及聲音的處理功能，將是未來的研究目標。

6. 參考文獻

1. 谷圳，「朝向無縫式的註記與發問學習輔助系統」，國立中央大學資訊工程研究所碩士論文，民國93年6月。
2. 吳玢玲，「線上多人批閱及分析系統—以C語言教學為例」，中華大學資訊管理系碩士論文，民國93年6月。
3. 邱續瑩，「開發SCORM教材之多媒體編輯工具實作」，成功大學工程科學所碩士論文，民國92年6月。
4. 陳守賢，「建構易用線上註記平台」，中原大學資訊管理系碩士論文，民國90年6月。
5. 許政穆、林建丞，「應用於網路教學的Web-based線上閱卷工具—A Web-based Scoring Tool for E-Learning」，TANET 2001，2001年10月。
6. 張春興，張氏心理學辭典，1991。
7. Andersen,C.”Collaboration as argument”,Presented at the meeting of the American Educational Research Association,New Orleans,LA,April,2000
8. Brown,A.L.”Metacognitive development and reading”,Theoretical issues in reading comprehension, 453-481,1980.
9. Brush.A.J.B.”Annotating Digital Documents for Asynchronous Collaboration”,

Technical Report 02-09-02,September,2002.

10. Brush,A.J.B.and Bargeron D.,“Robustly Anchoring Annotations Using Keywords”,
Technical Report MSR-TR-2001-107,November 16,2001.
11. Flavell,J.H.,”Metacognition and cognitive monitoring: A new area of cognitive
developmental inquiry”,American Psychologist ,34,906-911,1979.
12. Hidi,S., and Anderson,V., “Producing WritingSummaries: Task Demands, Cognitive
Operations, and Implications for Instruction,” Review ofEducational Research, 1996.
13. Kuhn,D.,”The skills of argument”,Cambridge University Press,New York,1991
14. Ovsianikov,I.A.,Arbib,M.A.,Mcneill,T.H.,”Annotation Technology”.International
Journal of Human-Computer Studies,50,329-362,1999.
15. Quade,A.M.,“An assessment of retention and depth of processing associated with
notetaking using traditional paper and pencil and on-line notepad during
computer-delivered instruction”,Eric Document Reproduction Service,No.ED383
330,1996.
16. Ronald L.R. “The MD5 Message-Digest Algorithm”. RFC 1321, April 1992.
17. Slotte,V.and Lonka,K.,et al.,“Note-taking review – Practical value for learners”,
Arob@se,www.arobase.to,volume 1-2, pp. 79-86, 2003.