

# 支援電話系統之多模式互動網站架構

林俊卿 葉耀明

國立台灣師範大學資訊教育所

juinching@hs.ntnu.edu.tw ymyeh@ice.ntnu.edu.tw

## 摘要

近年來，人們利用全球資訊網來獲取資訊的需求與日俱增，然而傳統瀏覽器利用視覺呈現網頁資訊的方式漸顯其限制，除了無法滿足部份身心障礙者需求外，隨著行動通訊設備的發展，行動上網的需求日增，但大多數的行動通訊設備在螢幕顯示上及輸出入設備上有其限制，間接或直接影響了網頁資訊存取的能力。為此，Universal Accessibility 的概念於是被提出，目的是希望不論一般人或身心障礙者，都能夠隨時、隨地使用任何通訊設備，取得網路上的資訊。本研究擬利用網頁設計技術及語音合成技術，設計一個視覺、語音呈現雙模式，滑鼠、鍵盤操作雙模式的新網站架構，並結合 VoiceXML 技術，建立支援電話系統的多模式互動網站，其目的在於增加全球資訊網的多元存取方式，在此概念下，使用者可利用電腦系統及電話系統存取全球資訊網。本研究並實作多模式互動網站開發系統，以協助使用者快速開發此類多模式互動網站。

**關鍵詞：** University Accessibility、Web Accessibility、Text to Speech、VoiceXML、Mobility。

## 1. 前言

近年來，全球資訊網以 HTML 語言的方式整合各種多媒體素材，其優異的多媒體呈現能力，生動活潑的互動方式，一躍成為網路上資訊傳遞的主流媒體。再加上電腦與網際網路的進步與普及，使得目前各國人民在日常生活中利用全球資訊網來獲取資訊的人數與日俱增。

然而，傳統的網頁以視覺呈現滑鼠操作的方式明顯無法符合部分身心障礙者的需求，為此，網頁可及性(Web Accessibility)的概念於是被提出，其主要概念是透過一些網頁的設計規範和法規條例，來達成身心障礙者可以無障礙的在網際網路上瀏覽網頁資訊。另外，隨著行動通訊設備的發展，行動上網的需求日增，但大多數的行動通訊設備在螢幕顯示上及輸出入設備上有其限制，間接或直接影響了網頁資訊存取的能力，Universal Accessibility 應運而生，強調任何人(代理人程式)可在任何時候，於任何支援網路的平台上，使用任何網頁存取技術以存取網頁上的資源。

本研究參採 Web Accessibility 及 Universal Accessibility 的概念，利用網頁設計技術及語音合成

技術，設計一個視覺語音呈現雙模式、滑鼠鍵盤操作雙模式的新網站架構，並結合 VoiceXML 技術，建立支援電話系統的多模式互動網站，其目的在於增加全球資訊網的多元存取方式，在此概念下，使用者可利用電腦系統及電話系統存取全球資訊網。本研究並實作多模式互動網站開發系統，以協助使用者快速開發此類多模式互動網站。

本文包含下列五個章節，第一章簡介研究動機，第二章介紹相關研究，第三章介紹多模式互動網站，第四章介紹多互動網站產生器，第五章討論本研究的成果及未來的發展。

## 2. 相關研究

### 2.1 Web Accessibility

根據國外一些相關研究報告中指出【3】【4】【5】【6】，傳統的網頁瀏覽器因其操作介面無法符合視障者的需求，而常常導致資訊取得不足或是資訊認知錯誤。Web Accessibility 的概念【8】於是被提出，目的是希望不論一般人或身心障礙者，都能夠非常方便地透過全球資訊網取得完整的資訊服務。Web Accessibility 的主要概念是透過一些網頁的設計規範和法規條例，來達成身心障礙者可以無障礙的在網際網路上瀏覽全球資訊網的網頁資訊。例如，網頁的文字可以使用點字設備顯示或用語音表達等等。

目前主導國際全球資訊網技術的標準機構「全球資訊網協會」(W3C)所形成的「資訊網可及性推動組織」(Web Accessibility Initiative, 簡稱 WAI)【10】已經在 1999 年五月五日訂定出一份「無障礙網頁內容可及性規範」(Web Content Accessibility Guidelines; 簡稱 WCAG)的國際標準。在國內，研考會於九十一年六月「無障礙網頁開發標準暨標章核發作業」委外案，參考 W3C 協會的 WAI 組織的無障礙網頁內容標準相關規範，並參照各國在制訂無障礙網頁相關政策和推廣策略，與國內近年來在身心障礙者保護政策等相關措施，訂定我國的「無障礙網頁開發規範」。

### 2.2 Universal Accessibility & Multimodal interaction

Universal Accessibility 則是 Web Accessibility 的擴充，意指「任何人(代理人程式)可在任何時候，

於任何支援網路的平台上，使用任何網頁存取技術以存取網頁上的資源」。換句話說，網頁資訊的存取不再只限於電腦設備，而可以擴充到個人數位助理(PDA)、WAP phones、Smart phones 或者傳統電話等等。

然而，大部分的行動通訊工具，受限於處理器的能力、螢幕顯示的尺寸、簡化的鍵盤輸入設備或滑鼠設備，直接或間接影響了存取網頁資訊的能力，因此，多模式互動機制(Multimodal interaction)【1】【2】於是被提出，即可藉由多種輸入模式，包括鍵盤、滑鼠、文字、語音及多種無障礙設備與系統展開互動。在支援多種輸出輸入模式的互動下，使用者可針對手中行動設備的特性，選擇最合適的方式存取網頁資訊。

### 2.3 語音合成技術

語音合成 (Speech Synthesizer) 又名文句轉語音(Text-To-Speech,TTS)，是指將輸入的文字或文件，轉換成清晰、流暢、自然的語音輸出技術。語音合成技術大致可分成文句分析、韻律產生器、合成單元產生器、語音合成等四個模組【7】。

目前的語音合成技術發展日趨成熟，許多語音合成器的產品也已經大量地商品化。因此，藉由運用已成熟的語音合成技術，可以大大減少研究時間的消耗，而更能專注於研究的本質。

### 2.4 VoiceXML

VoiceXML(Voice eXtensible Markup Language)植基於 W3C 定義的 XML (eXtensible Markup Language)，是一種人與電腦對話的標示語言。在 VoiceXML 架構下，人們可透過語音播放的提示資訊 (語音合成或事先預錄的音檔)、口述命令或按鍵音輸入，與機器展開互動，就像利用傳統的互動式網頁存取網路資源一般。

換言之，隨著語音技術的成熟，電腦不再是唯一上網的條件；網路資料經過 VoiceXML 的規範，只要利用電話或語音裝置也可隨時連上語音網站 (Voice Site)，而透過語音型態的瀏覽器，使用者可用語音方式取得網站內容，以及其他所需的服務。

本研究擬參考 Universal Accessibility 概念及 Multimodal interaction 機制，結合網頁設計技術、語音合成技術及 VoiceXML 技術，探討支援電話系統之多模式互動網站的可行性。

## 3. 多模式互動網站

### 3.1 多模式互動網站概念

多模式互動網站概念的提出在於增加全球資訊網的多元存取方式，在此概念下，使用者可用以

下兩個管道存取全球資訊網，包括利用電腦系統及電話系統，分述如下：

- (1) 電腦系統：這是傳統瀏覽網際網路的方式，然而本研究提出的多模式互動網頁具有視覺、語音呈現雙模式及滑鼠、鍵盤操作雙模式，所以使用者在瀏覽及操作網頁時將有更大的彈性。
- (2) 電話系統：藉由 VoiceXML 技術，使用者可利用室內電話或手機，透過公用電話系統以按鍵存取全球資訊網。

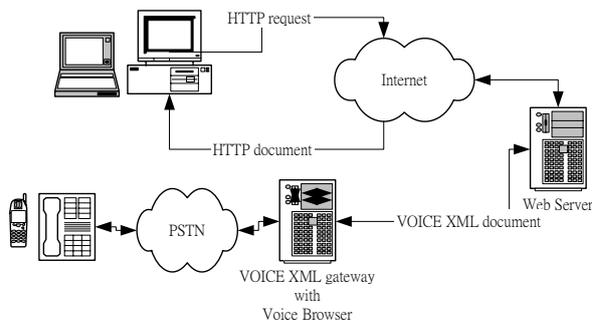


圖 1 語音互動網站概念圖

如同網頁瀏覽器傳遞 HTML 那樣，語音瀏覽器(Voice Browser)在語音開道(VoiceXML gateway)上執行，連結公共交換電話網(PSTN)及網際網路，負責 VoiceXML 文件的處理及音訊輸入(輸出)。其概念如圖 1 所示。

### 3.2 網站架構

有別於傳統的網頁架構以 HTML 網頁為基本單元，本研究以目錄為一個基本單元，在目錄下包括有：

- (1) HTML 檔案：具備視覺、語音呈現雙模式，鍵盤、滑鼠操作雙模式的 HTML 網頁，主要供電腦系統管道存取。
- (2) VXML 檔案：VoiceXML 檔案，主要供電話系統管道存取。
- (3) 網頁單元(選項)：表示子網頁單元。其架構如圖 2 所示。

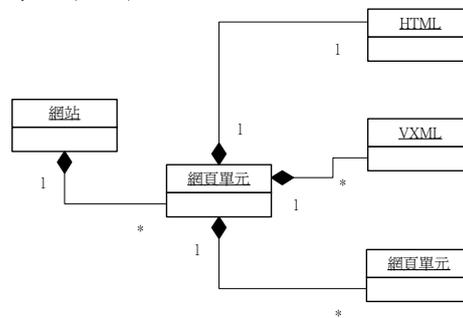


圖 2 網站架構圖

### 3.3 網頁結構

本研究所提出的網頁有以下三點特色：

- (1) 套用網頁樣版：網頁樣版觀念的引進，讓網頁

風格趨於一致，且網站編輯者能輕易的改變網頁的佈景主題，讓網頁呈現更多樣化。

(2) 網頁區塊劃分：網頁區塊的劃分讓網頁更具結構性，有助於瀏覽者對資訊的取得，本研究劃分區塊如下：

- 網頁標幟：放置網頁圖示。
  - 目前位置：顯示目前網頁所在網站中之位置。
  - 子網頁：列出目前網頁有哪些子網頁。
  - 網頁內容：顯示網頁內容。
  - 網頁語音圖示：每個圖示表示一個語音檔案，包含有網頁導覽語音、網頁全文語音及網頁段落語音。
  - 頁尾文字：顯示版權資訊。
- 網頁結構如圖 3 所示。



圖 3 網頁結構圖

(3) 主題語音分段：視覺與聽覺在資訊取得的方式上有著根本的差異，相較於視覺存取方式可以隨機的瀏覽網頁上任何一段資訊，聽覺存取只能以循序方式逐步獲取資訊。為了加速聽覺存取網頁資訊的效率，本研究引進隨機存取的觀念，即除了全文語音閱讀外，編輯者可將全文以主題劃分成若干段落語音，讓使用者可以更有效率的取得需要的資訊。

### 3.4 電腦系統存取模式

有別於傳統視覺呈現滑鼠操作的網頁模式，本研究提出的電腦系統存取模式包含以下三點特徵：

- (1) 呈現雙模式：包括視覺、語音呈現兩種模式。在語音呈現方面，其分類如下：
- 網頁導覽語音：呈現目前所在網頁位置、目前網頁有哪些子網頁及目前網頁有哪些文章段落。
  - 網頁全文語音：提供網頁全文語音。
  - 網頁段落語音：提供網頁段落語音，段落語音的呈現方式可讓使用者以更具彈性、更有效率的方式進行網頁資訊的讀取。
- (2) 操作雙模式：操作雙模式包括鍵盤、滑鼠操作雙模式。在鍵盤操作模式方面，我們將使用者的鍵盤事件分為四種模式，其分類如下：
- 導盲模式：網頁區分為標題區塊、子網頁區塊、內容區塊及頁尾文字區塊，導盲模

式可以讓使用者用快速鍵於各區塊間切換。

- 控制模式：以快速鍵執行系統控制功能，包括回首頁、停止播放、繼續播放、暫停播放、開啟(關閉)語音、顯示輔助訊息及登出等。
- 連結模式：以快速鍵瀏覽頁面上的子網頁。
- 內容模式：以快速鍵瀏覽網頁內容的分段段落。

符合無障礙網頁規範：本網頁的測試基於行政院研考會頒發的「無障礙網頁開發規範」，並採用研考會之無障礙網路空間服務網之網頁檢測功能，其網址為<http://enable.nat.gov.tw/check.jsp>。本研究將以符合機器檢測三星級為目標。

### 3.5 電話系統存取模式

使用者利用市內電話或手機透過公共交換電話網路經過 VoiceXML gateway 連結到多模式互動網站後即可開始存取網頁內容，其步驟包含下列動作：

- (1) 網站傳回導覽語音。
- (2) 使用者依照導覽語音進行瀏覽動作。
- (3) 重複步驟(1)(2)，使用者可對整個網站內容進行完整的存取。

其中使用者對網站的瀏覽動作包含以下兩種模式：

- (1) 導覽模式：用在網頁連結的切換。導覽模式語音包括位置訊息、控制訊息、連結訊息及段落訊息等。語音訊息及快速鍵詳見表 1。
- (2) 閱讀模式：用在網頁內容(全文或段落)的閱讀。閱讀模式語音包括位置訊息、控制訊息及內容訊息等。語音訊息及快速鍵詳見表 2。

表 1 導覽模式語音訊息及快速鍵對應表

訊息類別	語音訊息
位置訊息	您目前的位置是...
控制訊息	回上一頁請按 0 閱讀全文請按 1 重新播放請按 2
連結訊息	共有 n 個子網頁 前往子網頁 1 請按 3 前往子網頁 2 請按 4 ... 前往子網頁 n 請按 n+2
段落訊息	共有 m 個段落 閱讀段落 1 請按 1+(n+2) 閱讀段落 2 請按 2+(n+2) ... 閱讀段落 m 請按 m+(n+2)

導覽模式語音範例如下：

您目前的位置是網頁  
 回上一頁請按 0  
 閱讀全文請按 1  
 重新播放請按 2  
 共有 2 個子網頁  
 前往子網頁 1 請按 3  
 前往子網頁 2 請按 4  
 共有 2 個段落  
 閱讀段落 1 請按 4  
 閱讀段落 2 請按 6

表 2 閱讀模式語音訊息及快速鍵對應表

訊息類別	語音訊息
位置訊息	您目前的位置是...
控制訊息	回上一頁請按 0 重新播放請按 1
內容訊息	網頁內容如下：...

閱讀模式語音範例如下：

您目前的位置是「靜夜思」  
 回上一頁請按 0  
 重新播放請按 1  
 網頁內容如下：  
 床前明月光,疑似地上霜。舉頭望明月,低頭思故鄉。...

## 4. 多模式互動網站產生器

### 4.1 系統架構

本系統是以 Apache Tomcat 為網頁伺服器，並採用 JSP(Java Server Pages)伺服器端網頁技術開發而成的網頁應用程式，使用者區分為編輯者及一般使用者，使用案例圖如圖 4 所示。網站編輯及匯出僅限編輯者使用，網站建構共分為三個步驟，分述如下：

- (1) 登入：進行身份確認後方能使用系統。
- (2) 網站編輯：網站管理以檔案總管方式呈現，網站編輯以所見即所得方式呈現，如圖 5 所示，編輯者應可輕易上手。
- (3) 網站匯出：將編輯完成之網站匯出後成為 ZIP 壓縮檔，解壓縮後置於任何網頁伺服器上即可順利執行，不需進行額外設定。

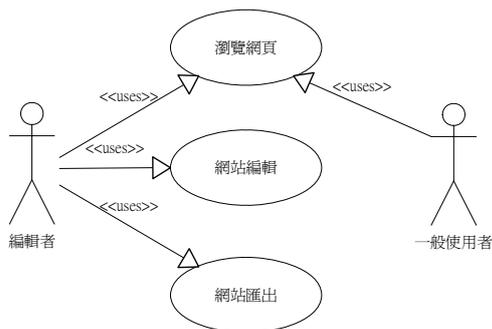


圖 4 使用者案例圖



圖 5 網站編輯介面圖

而系統的組成包括以下五個模組：

- (1) 網站編輯模組：產生網頁單元的 XML 記錄檔案及 HTML 檔案。
- (2) 網頁語音處理模組：分析 XML 記錄檔案並呼叫語音製作模組產生語音檔案，並嵌入 HTML 及 VXML 檔案中。
- (3) 語音製作模組：接收來自網頁語音處理模組的文字檔案並製作語音檔案。
- (4) 快速鍵處理模組：分析 XML 記錄檔案產生快速鍵處理程式碼並嵌入 HTML 檔案中。
- (5) VXML 處理模組：產生電話系統所需的 VXML 檔案。

模組相關圖如圖 6 所示，各模組將於以下各節分別介紹之。

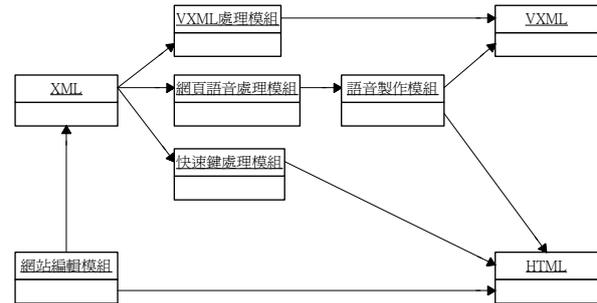


圖 6 模組相關圖

### 4.2 網站編輯模組

網站編輯模組包括以下三個主要功能：

- (1) 網站編輯：提供檔案總管的操作方式，進行網站結構的維護。
- (2) 網頁編輯：提供所見即所得的操作方式，進行網頁內容的設計。
- (3) 網站匯出：將設計好的網站打包成 ZIP 壓縮檔案。

在網頁編輯方面，編輯者所編輯的資料網頁，將會被分類記錄到多個 XML 檔案，而後套用網頁樣版，產生以視覺呈現方式的 HTML 網頁，其概念圖如圖 7 所示。

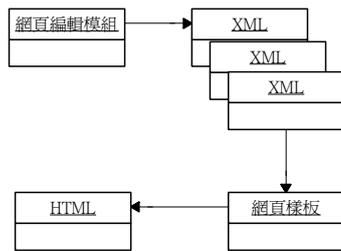


圖 7 網頁編輯模組循序圖

### 4.3 語音製作模組

語音製作模組可以將文字檔案轉換成 mp3 語音檔案，其動作包含以下兩個步驟：

- (1) 語音轉換：利用 Text to Speech 文字轉語音引擎將文字檔案(txt)轉換成語音檔案(wav)。
  - (2) 語音壓縮：利用 LAME encoder 壓縮程式將轉換後的語音檔案(wav)壓縮成 mp3 語音檔案。
- 其概念如圖 8 所示。

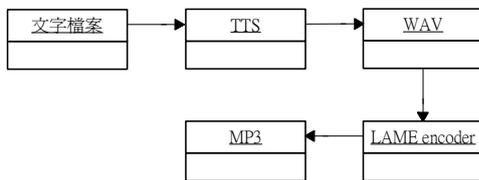


圖 8 語音製作模組循序圖

### 4.4 網頁語音處理模組

網頁語音處理模組目的在產生網頁瀏覽所需要的各種語音檔案，並將語音檔案嵌入網頁中，以達到呈現雙模式的目標。其步驟如下：

- (1) 分析記錄檔案：分析網頁編輯模組產生的各項 XML 記錄檔案。
- (2) 產生文字檔案：產生網頁瀏覽各種模式所需要的文字檔案，包括導覽模式、全文模式及段落模式文字檔案。
- (3) 呼叫語音製作模組：呼叫語音製作模組將各種文字檔案轉換成相對應的語音檔案，包括導覽語音、全文語音及段落語音檔案。因 VXML 檔案不支援 MP3 語音格式，故語音檔案包含 MP3 及 WAV 兩種格式。
- (4) 嵌入語音檔案：將 MP3 語音檔案嵌入由網頁編輯模組所產生的 HTML 檔案，嵌入語音後的網頁將具備視覺、語音呈現雙模式。另外，WAV 語音檔案將嵌入由 VXML 處理模組產生的 VXML 檔案。

其概念如圖 9 所示。

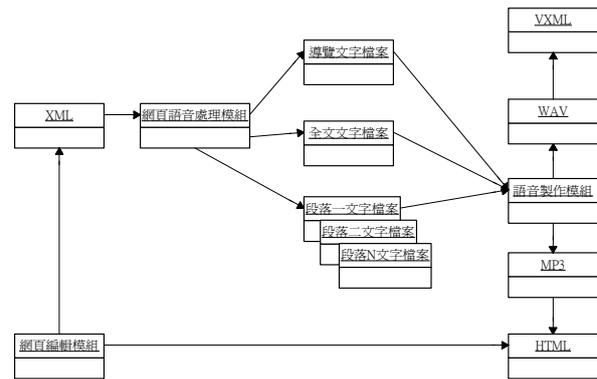


圖 9 網頁語音處理模組循序圖

### 4.5 快速鍵處理模組

快速鍵處理模組目的在產生網頁所需的快速鍵處理程式碼，並將程式碼片段嵌入網頁中，以達到操作雙模式。其步驟如下：

- (1) 分析記錄檔案：分析網頁編輯模組產生的各項 XML 記錄檔案。
- (2) 產生各種鍵值處理程式碼：產生各種鍵盤操作模式的處理程式碼，包括導盲模式、控制模式、連結模式及內容模式程式碼。
- (3) 嵌入程式碼：將四種鍵值程式碼嵌入 HTML 檔案中，嵌入程式碼後的網頁將具備滑鼠、鍵盤操作雙模式。

其概念如圖 10 所示。

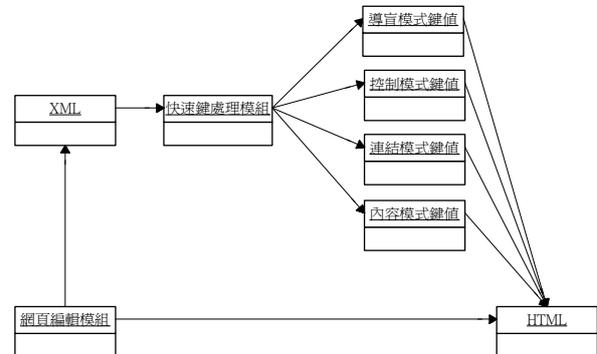


圖 10 快速鍵處理模組循序圖

快速鍵處理模組利用 JavaScript 的鍵盤處理機制捕捉使用者的鍵盤活動，其演算法如下：

```

//四種鍵盤模式之處理
function handleKeys(evt) {
  evt = (evt) ? evt : ((window.event) ? event : null);
  if (evt) {
    //導盲模式鍵值處理
    //if (keyCode=...) ...
    //控制模式鍵值處理
    //if (keyCode=...) ...
    //連結模式鍵值處理
    //if (keyCode=...) ...
    //內容模式鍵值處理
    //if (keyCode=...) ...
  }
}

```

```

}
//設定由 handleKeys 處理鍵盤事件
document.onkeyup = handleKeys;

```

#### 4.6 VXML 處理模組

VXML 處理模組的目的在產生電話系統存取模式所需的 VXML 檔案,包括有導覽模式的 VXML 文件及閱讀模式的 VXML 文件:

(1) 導覽模式 VXML 文件: 導覽模式 VXML 文件包含目前位置、現在網頁下的子網頁、網頁內容區分為幾個段落及快速鍵對應等資訊,其格式如下。

```

<?xml version="1.0"?>
<vxml version="1.0">
<menu>
<prompt><audio src=" 導覽.wav"/></prompt>
<choice dtmf="0" next=" ./ 導覽.vxml"/>
<choice dtmf="1" next="全文.vxml"/>
<choice dtmf="2" next=" 導覽.vxml"/>
<choice dtmf="3" next="子網頁 1/ 導覽.vxml"/>
<choice dtmf="4" next="子網頁 2/ 導覽.vxml"/>
...
<choice dtmf="n+2" next="子網頁 n/ 導覽.vxml"/>
<choice dtmf="1+(n+2)" next="段落 1.vxml"/>
<choice dtmf="2+(n+2)" next="段落 2.vxml"/>
...
<choice dtmf="m+(n+2)" next="段落 m.vxml"/>
</menu>
</vxml>

```

(2) 閱讀模式 VXML 文件: 閱讀模式包括全文及段落兩種 VXML 文件,文件內容包括全文(段落)語音檔案,並提供回上一層及重聽等選項,其格式如下所示。

```

<?xml version="1.0"?>
<vxml version="1.0">
<menu>
<prompt><audio src="全文(段落).wav"/></prompt>
<choice dtmf="0" next=" 導覽.vxml"/>
<choice dtmf="1" next="全文(段落).vxml"/>
</menu>
</vxml>

```

其概念圖如圖 11 所示。

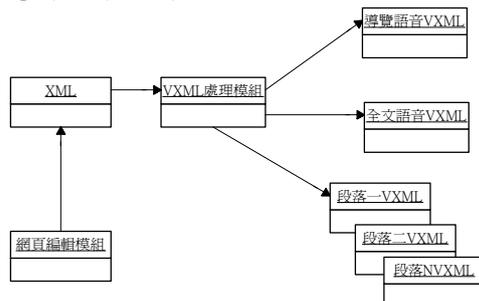


圖 11 快速鍵處理模組循序圖

#### 4.7 電話系統實例模擬

本節將使用 IBM 公司的 Websphere Voice Server SDK 這套軟體來模擬電話系統存取多模式互動網站,假設網站架構圖如圖 12 所示。

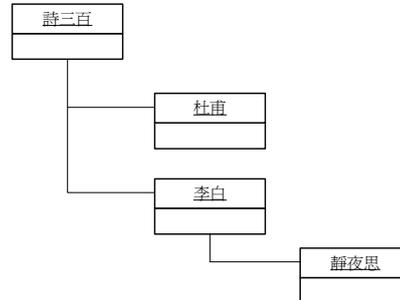


圖 12 範例網站圖示

若要閱讀「靜夜思」則對話流程如下:

您目前的位置是「詩三百」  
 回上一頁請按 0  
 閱讀全文請按 1  
 重新播放請按 2  
 共有 2 個子網頁  
 前往「杜甫」請按 3  
 前往「李白」請按 4  
 按 4 選擇「李白」。

您目前的位置是「詩三百/李白」  
 回上一頁請按 0  
 閱讀全文請按 1  
 重新播放請按 2  
 共有 1 個子網頁  
 前往「靜夜思」請按 3  
 按 3 選擇「靜夜思」。

您目前的位置是「詩三百/李白/靜夜思」  
 回上一頁請按 0  
 重新播放請按 1  
 網頁內容如下:  
 床前明月光,疑是地上霜,舉頭望明月,低頭思故鄉。

#### 5. 結論

本研究參考 Universal Accessibility 概念及 Multimodal interaction 機制,結合網頁設計技術、語音合成技術及 VoiceXML 技術,提出視覺、聽覺呈現雙模式,滑鼠、鍵盤操作雙模式,且支援電話系統存取的多模式互動網站架構,以協助身心障礙者及行動通訊族對網路資訊的取得。雖然多模式互動網站有如此的優點,但以傳統方式製作此類網站有其難度,網站編輯者須具備程式語言基礎、語音檔案製作及 VoiceXML 的背景知識。本研究提出一個簡單的編輯器,以檔案總管的方式管理網站架構,以所見即所得的方式編輯網頁,大大簡化多模

式互動網站的製作難度。

展望未來，除了多模式互動網站的推廣外，本研究更將朝以下方向繼續進行：

- (1) 支援多國語言：目前本研究僅支援中文，依照本系統模組化的設計，其他語言的支援僅需更換文字轉語音引擎即可。
- (2) 結合自然語言辨識系統：目前本系統以類似電話語音選單系統方式進行網頁的操作，未來希望可結合自然語言辨識系統，讓網頁的操作更具人性化。

## 參考文獻

- [1] Consortium, W.W.W. 2000a, Multimodal Interaction Use Cases. Available at <http://www.w3.org/TR/mmi-use-cases/>.
- [2] Consortium, W.W.W. 2000a. Multimodal Interaction Framework: Available at <http://www.w3.org/TR/2003/NOTE-mmi-framework20030506/>.
- [3] James, F. "AHA: Audio HTML Access", The Six International World Wide Web Conference. Ed, by Michael R. Genesereth and Anna Patterson, Santa Clara, CA, 7-11 April 1997. IW3C2, pp. 129-139.
- [4] James, F. "Presenting HTML Structure in Audio: User Satisfaction with Audio Hypertext", ICAD 96 Proceedings, Xerox PARC, 4-6 Nov. 1996, pp. 97-103
- [5] James, F. "Lessons from Developing Audio HTML Interfaces", ASSETS 98, April 1998, pp. 15-17.
- [6] Rollins, S. and N. Sundaresan, "AVoN calling: AXL for voice-enabled Web navigation", Elsevier Science, Computer Networks, Vol: 33, Issue: 1-6, pp.533-551, June 2000.
- [7] Speech synthesize note, <http://www2.cyut.edu.tw/~s9154610/speech.html>
- [8] Waters, C. "Universal Web Design", New Riders Co., 1997.
- [9] White, B. "Web accessibility, mobility and findability", Web Congress 2003, pp. 239 – 240.
- [10] WWW Consortium, "Web Accessibility Initiative". <http://www.w3.org/WAI/>