

HTML 與 WML 之間的轉換與整合開發環境

A Conversion Environment Between HTML and WML

連耀南 陳宗儀 何彥德 吳曉峰 蔡子傑 張宏慶

國立政治大學資訊科學系

摘要

讓使用者以 WAP 行動電話 (手機) 在行動中透過行動通訊網路連上網際網路已成為近期內資訊通訊界亟欲突破的目標, 然而除了通訊費用以及頻寬有待大幅改進之外, 網際網路的 WML 內容極為貧乏, 而 WAP 手機又不適合直接瀏覽網路上數以百萬千萬計的豐富 HTML 內容, 本文介紹我們設計並實作的一個 HTML2WML 轉換環境, 可將 HTML 文件轉換成 WML 文件, 提供給 WAP 使用者以瀏覽網路上的 HTML 內容。我們針對 WAP 手機的限制及 WML 的能力分析網路上的普遍形式, 並據以設計合適的轉換原則。我們的轉換環境除了一般常見的 HTML2WML 轉換器之外, 還包括了一個 WML2HTML 轉換器, 供內容提供者在進行 HTML2WML 轉換後, 可以利用 HTML 瀏覽器以 HTML 方式檢視整群 WML 網頁。

一、簡介

1.1. WAP 架構與環境

讓使用者以在行動中以行動台透過行動通訊網路連上網際網路是一大技術挑戰, 因為無線設備頻寬既窄且貴, WAP 手機運算資源有限, 螢幕也較小, 不能直接使用網際網路慣用的通訊協定, 而必須制定專門的協定來支援。無線應用協定 WAP(Wireless Application Protocol), 是一種開放、標準的無線應用軟體協定, 類似 Internet 上的 HTTP 協定, 但係針對數位式行動電話網路 (GSM) 所開發的, 供行動電話或 PDA 等無線通訊設備來執行網際網路存取服務的開放標準[4,5]。WAP 文件所使用的語法是 WML 標記式語法, 相當是個人電腦上面瀏覽器中所用的 HTML 程式語法。只要手機能支援 WAP 協定, 就可以經由無線通訊網路存取網際網路資源。

現在一般的 WAP 手機並無 TCP/IP 能力, 且未配有 IP 地址, 無法直接上網, 必須透過行動通訊服務提供者所架設的 WAP Gateway, 手機將 request 透過 WAP 協定送給 Gateway, 而此 Gateway 則連上 Internet 利用 HTTP 協定去執行此 request,

若有任何結果, 則經過適當 encoding 之後 循原機制, 反向送回給手機。

1.2. WAP 的 HTML – WML

WML(Wireless Markup Language)[6], 類似 HTML 語法, 是 WAP 規格中的一部份, 自 XML (eXtensible Markup Language) 衍伸而來, 是用來設定文字或圖形呈現於 WAP 手機螢幕的方式。為了配合 WAP 手機的小螢幕, 以及充分利用網路傳輸的特性, WML 支援 card 與 deck 的觀念:

WML 與使用者的溝通畫面 一個 WML 網頁可有多個 card。

多個 card 組成一個 deck, 每一個 deck 最多可以容納 1500 bytes 資訊。

WML 文件設計者可將一個 WML 文件設計成一個 deck 而將文件內容依照手機螢幕大小切割成較小的區塊, 而將這些區塊設計成 card, 並以 hyperlink 將 deck 內的 card 連結起來, 方便使用者循著 hyperlink 依序點閱。在傳輸時, WAP server 會將整個 deck 全部傳下來以節省通訊時間與費用。

1.3. HTML2WML 轉換

由於網路上有著非常豐富的 HTML 文件, 而絕少 WML 文件, 有必要設計一個轉換工具, 供使用者運用網路上豐富的 HTML 文件。將 HTML2WML 轉換器放在 WAP Gateway 上, 由 WAP Gateway 執行轉換工作, 是比較理想的方式。但是由於理想的忠實轉換工具難以實現, 有必要研究開發離線轉換工具, 讓內容提供者可以在此轉換工具的協助下, 以半自動的方式進行誤差較少的離線轉換。目前我們所設計的轉換器可放在 Kannel WAP Gateway 上[1], 也可作為離線轉換工具。

目前的 WAP 服務都架構在 GSM 系統上, 其 9.6kbps 的頻寬加上以時間計費的特性使得 WAP

的內容服務必須極為小心的控制網頁大小，手機的超小畫面以及瀏覽器的有限能力也對 WML 網頁做了很多的限制。此外，WML 語法與 HTML 語法不盡相容，以致轉換工作受到很大的限制，例如：

HTML 語法容許不嚴謹的格式，例如容許應配對而沒配對的 tag，而 WML 是 XML 之衍伸，有嚴謹的格式定義，未配成對的 tag 不為 WML 文法所接受。因為很多 HTML 文件並未嚴格遵照標準格式撰寫，使得分析 HTML 文件變成相當困難。

WML 不支援 frame，而網路上的 HTML 文件常含有複雜的 frame 架構。

WML 不支援 Recursive Table 元件 (table 中又內含有 table)。

HTML 文件中常含有 Javascript 或其他外掛程式，其能力遠超過 WAP 手機中有限資源所能支援的 WMLscript，要將這些 Javascript 程式自動轉換成 WMLscript 程式幾乎是不可能。

、設計構想

我們先根據兩種語言的相容性與轉換之困難度將各種 HTML tag 歸納成三類：

2.1. 中文 Encoding - Unicode

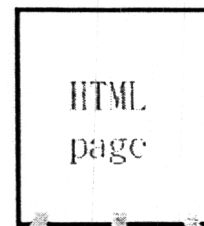
WML 支援 Unicode[3]，且大部分具中文顯示能力的 WAP 手機僅支援 Unicode，但國內網路上的中文 HTML 文件都是以 Big5 編碼的。因此轉換工作之第一件事是 Big5 碼與 Unicode 碼之間的轉換工作。如欲支援多國語言的 HTML 文件轉換，則相對應的編碼轉換是不可避免的。表一是一個簡單的例子。

中文	Big5	Unicode
政		
大		

表一、中文 Big5 與 Unicode 的對照

2.2. 文件分割

網路上的 HTML 文件通常遠大於 WAP 手機螢幕可容納的範圍，必須作適當的分頁，將一個 HTML 文件依適當的大小切割成數量不等的小文件。文件小於 1500bytes 時，每一個小文件可轉換成 WML 的一張 card，而所有同一 HTML 文件的所有 card 組成一個 WML deck，在轉換過程中，我們並依 card 的順序建立對應的 hyperlink，讓使用者可以依序瀏覽。大於 1500 bytes 的文件就不能全放在同一個 deck 中，在此情形下，可以將各個切割出來的小文件轉成一個單獨的 WML 文件。



圖一、分頁

2.3. 資訊減量

無線通訊網路的頻寬既窄且貴，加上 WAP 手機資源不足之限制，有必要去除 HTML 文件內不必要的資訊，以減少轉換後 WML 文件之大小，並節省資訊傳輸量。

某些 tag 或 parameter 在 WML 是無意義的，例如 background, color, font, 等。

HTML 文件內的註解 (comment)，並不一定要保存，可由設計者決定是否保留。

某些 HTML 版面控制用的 tag 或 parameter 在 WML 中也失去版面控制的意義，可以除去。

2.4. Image

HTML 與 WML 都以 <image> 支援圖檔，但 WML 只支援其自定的圖檔格式 (WBMP)，原來 HTML 所支援的 JPG/GIF 等圖檔必須先行轉換成 WBMP 格式。此外，由於圖檔可能遠大於手機螢幕，有時必須考慮降低圖檔之解析度，將圖檔縮小以配合螢幕大小。



圖二、原圖檔(.gif)



圖三 轉換後之 WBMP

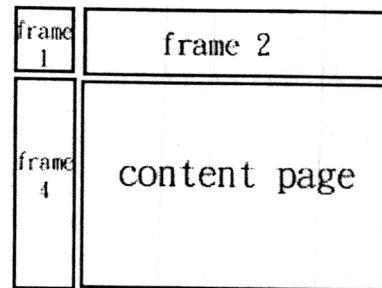
2.5. Frame

網路上流行利用 frame 將一個畫面分割成數個小畫面，而其中一個畫面 (稱為 index frame) 列有資訊索引 (稱為 index hyperlink) 指向另一個 HTML 文件 (稱為次級文件)，使用者可以點選 index hyperlink 以瀏覽次級文件，而點選結果就顯示在另一個 frame 上。不幸的是，WML 並不支援 frame 以致處理此類文件相當棘手。

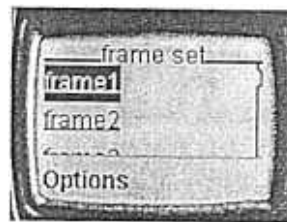
一個理想的轉換器應該有能力辨認 index frame，再將其轉換成一個 frameset 的「首頁」(稱為 index page) 以 hyperlink 方式達到次級文件。但是由於網路上此種型態的網頁沒有固定撰寫規則，辨認 index frame 的難度太高，我們採取另一個較為簡單，但較不經濟的轉換方法。我們的轉換器首先將一個 frameset 所屬的 frame 各自獨立轉換成一個 page (稱為 frame page)，其次，建立一個「首頁」(稱為 super index page) 包含

一組 hyperlink 指到各個 frame page。使用者可以循 super index page 中的 hyperlink 去點閱各 frame page。Index frame 中原含有之 hyperlink 只需稍加處理，即可順利轉換成適當的 WML index page。

這樣的轉換方法可以順利轉換此類網頁。但免不了有一些缺點：多了一層指到各 frame page 的 super index page，使用者不但必須多往下走一層，而且必須自行從數個 frame page 中尋找 index page。



圖四、一般頁面的分割



圖五、轉換分割的結果

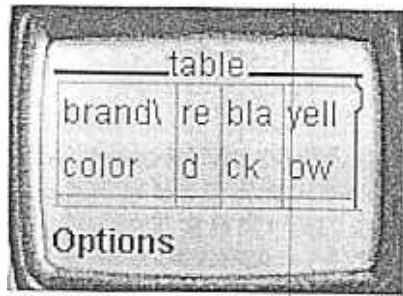
2.6. Table

HTML 文件中常含有 Table 元件，無論在資訊量或複雜度均遠大於 WAP 手機所能承受的範圍。因此，必須進行切割，或以其他方式展現。此外，HTML 的 Table 元件容許 recursive table，而 WML 卻不支援。

目前已存在的 HTML2WML 技術中，有一種作法是將 Table 元件展開成為 recursive list。我們則採用 table 切割的方法，將一個 table 切割成數個小 table。

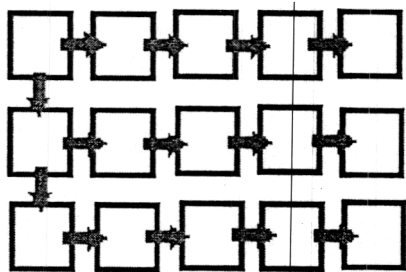
brand/color	red	black	yellow
nokia	23	45	77

圖六、簡單的表格



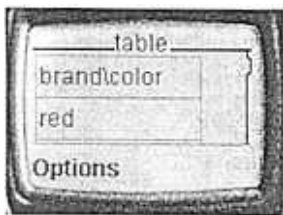
圖七、未經切割直接轉換表格的結果

首先，我們將一個 table 內的所有 cell 以 row major linked list 串起來，如圖八所示。

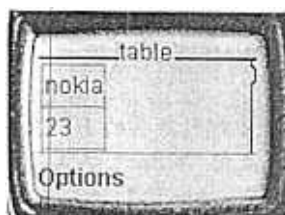


圖八、表格的切割與組織

接者依據 WAP 手機螢幕大小為度選擇適當數量的 cell 依次以直行顯現。每一組 cell 放在一張 card 中，而同一 table 中的所有元件則組成一個 deck，card 與 card 之間，則以 hyperlink 相連，如圖九、十所示。

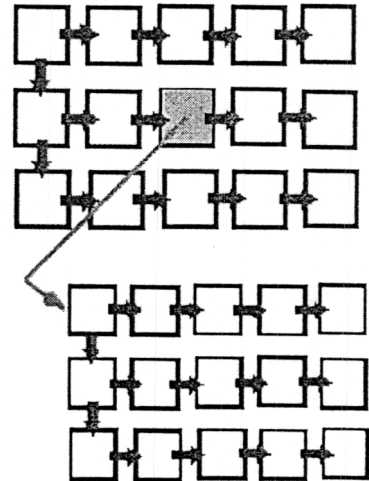


圖九、新表格第一列



圖十、新表格第二列

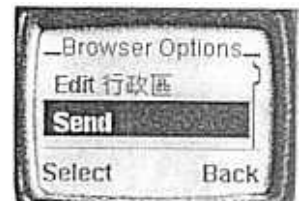
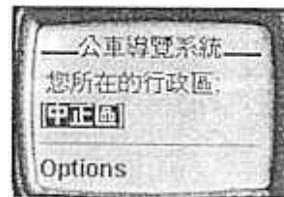
在處理 recursive table 時，很自然的也使用 recursive conversion 技術將 subtable 視為一個新的 table 處理，而以 hyperlink 連結，如圖十一所示。



圖十一、Recursive Table 之處理

2.7. User Interaction and CGI

在 WWW 上，server 多藉由 Form 與 CGI 的方式跟使用者溝通，這對於轉換工作是一大挑戰。網路上的網頁，在使用 Form 時，常常為了畫面之整齊，將各項元件以 table 排列（即 <form> 與 </form> 之間含有 table），但 WML 並不支援這個方式，必須除掉 table。



圖十二、經轉換後之表單

對於 CGI 的處理，我們可以採取靜態或動態兩種方式處理。靜態處理，是將原來的 CGI 全都以人工方式改寫為 WML 的格式，這種方式之工作量相當大，而維護兩種版本的 CGI 程式也是麻煩的事。比較理想的作法是採用動態轉換的方式，在 WAP Gateway 中，加入轉換程式，將 CGI 傳回來的 HTML message 自動轉換成 WML message。此法可以沿用原有的 CGI 程式，但必須仰賴完美的全自動轉換器，所幸，一般 CGI 傳回來的 message 之形式多半很簡單，轉換誤差之機會並不高。

2.8. Hyperlink

WML 支援 hyperlink，因此可以順利轉換一般的 HTML hyperlink。麻煩的是，很多網路上的 HTML 文件中的 hyperlink 是以 image icon 或 image map 作為螢幕上的提示文字，例如：

```
<A href="targetfile.htm"></A>
```

而 WML 無法支援此種作法。如果原文中的 hyperlink 元件含有 ALT 參數，我們可以摘出參數值，作為 hyperlink 的提示文字。如果沒有，我們就以圖檔的檔名作為提示文字，期望圖檔的檔名與文件內容原就具有高度相關性。此外，我們還須設法將 URL 轉成絕對地址，因為轉換過的 WML 檔案不一定會被放在與原文相同的目錄。

三、Development Environment

在進行文件轉換的過程中，設計者必須要不斷地去檢視轉換後的新文件，以檢視諸如正確性、外觀、及親和性等品質。現有唯一可用的工具是 WAP 模擬器，可在 PC 上模擬 WAP 手機，使用者可用 WAP 模擬器逐一檢視每一個 WML deck。由於一個 HTML 文件可能被轉換成數十個 deck，這種檢視方式非常的耗時費力，尤其是轉換後的 WML 文件常常是支離破碎，使用者必須在很多 card 與 deck 之間跳來跳去，無法觀照全局，親和力極低，對文件設計者是一項沈重的負擔。為了方便文件設計者檢視轉換過的 WML 文件（或未轉換之原始 WML 文件），我們開發了一個 WML2HTML 轉換器，可將轉換過的 WML 文件逆向轉成 HTML 文件，以方便檢視轉換成果。（逆向轉換只是將轉換過的 WML 文件以 tag by tag 的方式轉成 HTML 文件，並未改變原來的 WML 文件架構，所以此逆向轉換的功能並非將轉換過的 WML 文件「還原」成原始 HTML 文件。）我們設計的 WML2HTML 逆向轉換器將許多 deck 全部放在一個 HTML 文件中，並按原有的 WML hyperlink 結構，將這些 card/deck 以 HTML hyperlink 連起來，使用者可在一個 HTML 文件中模擬從一張 deck/card 跳到另一張 deck/card 的過程，視線可同時觀照所有的 deck/card，比 WAP 模擬器方便多了。

當然，這個 WML2HTML 轉換器也可將未經轉換之原始 WML 文件轉換成 HTML 文件，但是如同 HTML2WML 一樣，會有許多轉換誤差。

四、未來工作

要提高 HTML2WML 自動化程度、減少轉換誤差，轉換程式必須加入更多的 intelligence，例如：

辨認 hyperlink icon 內的藝術文字 做為 hyperlink 的提示文字。

辨認 frameset 內的 index frame 提升 frameset 的轉換品質。

此外，由於行動通訊頻寬的限制 有必要進一步進行資訊減量，例如：

文字摘要，關鍵字自動抽取

類似電子報的方式，將文字內容以精要的一句話或幾行列出，以使用者可辨識為原則。

圖檔摘要

辨識並抽取圖檔重點，甚至將原圖予以適當切割，以使用者可辨識為度，若有需要，可讓使用者自由點取原圖。

五、結論

在電話和網路結合之後，WAP 是提供無線上網能力的第一步，由於其種種限制與不便，使得 WAP 手機上網仍未能獲得普遍使用，在未來 GPRS[2] 甚至第三代行動電話成熟之後，無線頻寬將會提高、通訊費用將會下降，行動上網仍將是一個很有潛力的技術。除了技術因素之外，一個網路能否提供使用者所需的「服務內容」將會是一個更重要的關鍵[7]，HTML2WML 將是一個重要的工具，讓 WAP 手機能 access WWW 網路上豐富的 HTML 資源。

參考資料

1. Kannel, <http://www.kannel.org/>.
2. MobileGPRS.com, <http://www.mobilegprs.com/>.
3. Unicode Web site, <http://www.Unicode.org/>.
4. Tsu-Chieh Tsai, Hung-Chin Jang, Yao-Nan Lien, Shi-Chi Huang and Chi-Yin Liu, "Introduction to Wireless Application Protocol(WAP)", *Proc. of the 6th Workshop on Mobile Computing*, March 2000, pp. 145-149.
WAP Forum, <http://www.wapforum.org/>.
XML Reference, <http://www.w3.org/XML/>.
Yesmobile, <http://www.yesmobile.com.tw/>.