

94年7月

架構於 MathML 並支援 LaTeX 輸入之數學討論版設計

黃皇男* 顏佩君† 王婕‡

摘要

本文將探討如何在建立以 MathML 標籤語言為基礎，並配合 LaTeX 語法輸入數學式之網路數學討論版。其設計方法，首先建立 LaTeX 語法與 MathML 之間的轉譯機制；進一步將包含 LaTeX 語法編輯的數學式在內的文章，利用討論版設計語言 PHP，將數學式的部分轉譯成 MathML 語言後，存入 MySQL 資料庫，最後透過 PHP+MySQL 產生 XML 檔輸出成網頁。透過此種討論版，人們可利用網際網路的快速與方便，達成討論數學、交換心得的目的。未來系統的設計也將朝向結合圖形輸入介面，使不熟悉 LaTeX 的人也能輕鬆運用此類討論版。

關鍵詞：MathML、LaTeX、XML、PHP、MySQL

一、簡介

設計支援數學符號顯示之討論版的想法，主要源於作者從事數學教學、研究工作當中，經常與學生或研究領域相近學者的交流、討論的需要，但卻無法使用網際網路上的討論版來進行，相當可惜。又由於支援 LaTeX 語法的數學排版軟體在學界廣泛流行，因此，便興起結合現有的網際網路技術，來研製支援 LaTeX 語法輸入、線上呈現數學符號的討論版，促進數學在網路上傳播。

目前網路上所呈現包含數學式的文件，多以圖形的方式呈現數學式，不但無法表達數學式的意義，而且需要使用相當大的磁碟空間。有鑑於此，Worldwide Web Consortium(萬維網聯盟，簡稱 W3C)在 1998 年首次提出網路文件數學式子的表示法 Mathematical Markup Language(數學標籤語言，簡稱 MathML)，最新的版本 MathML2.0 也在 2001 年 2 月發表了[1]。MathML 除了能在如 IE、Mozilla[2,3]等瀏覽器顯示數學表示式，更重要的是它還記錄了數學符號的結構及意義，詳細請參考專書[4]。

但至今 MathML 仍無法廣為流傳，乃是因為他的語法繁複(其原始設計理念是爲了能精確的表達數學式，而非作爲輸入數學式的語法)，爲表達一小行的數學式子，往往就要寫上一大段的

* 東海大學數學系

† 交通大學資訊科學所

‡ 華盛頓中學教師

MathML Code，對於使用者而言耗時費力；相對的，LaTeX 語法就簡潔多了。我們可以從下面的例子中很明顯的感受到：

表一：LaTeX 與 MathML 語法比較

LaTeX	$(x+y)^3$
MathML	<pre> <msup> <mrow> <mf></mf> <mrow> <mi>x</mi> <mo>+</mo> <mn>y</mn> </mrow> </mrow> </msup> </pre>

雖然有這些缺點，如何在網路上利用 MathML 進行網際溝通、含數學式的文件展示、網路計算的工具之研究也不段的進行，目前已經舉辦有兩次國際研討會[5]。

為了解決 MathML 的輸入、顯示、排版的問題，國際上持續在發展相關軟體[6]，而 WC 為了解決 MathML 的規格，另行發展了 Amaya[7]，它是一個瀏覽器也是一個編輯器，可編輯 MathML(還有 XHTML、SVG、CSS)文件；但還無法最到動態網頁，及線上數學討論，以及不支援多國語言輸入與顯示功能。然而由 Design Science 所發展出的 WebEQ[8]，這是一套包含互動數學的建構網頁工具，以及數學符號的圖形輸入介面 WebEQ Editor(可參考交大應數數學交流網[9])；然而它是以外掛程式(Applet)的方式來處理數學式(MathML)的顯示，當開啓網頁的數學式較多時，顯示速度慢，效果稍差。

LaTeX 與 MathML 語法之間的轉換研究，國外學界有加拿大的 ORCCA 之 MathML 計畫[10]在進行，而國內則有中興大學應數系的許昌旺老師。除此之外，尚有一些數學軟體廠商，如 Mathematica、Maple、IBM Techexplorer、WebEQ(Math Type)等，以及少許個人(如 itex2mml[11]、tex4ht[12]、web editeur mathematique[13]、ttM[14]等)在進行，但較顯著的成果都屬於軟體公司的財產，而非估開軟體。如 WebEQ 外，只有 itex2mml 被引進討論版，也是用外掛程式來進行[15]。由於有這些 LaTeX 轉成 MathML 的軟體，以前所整理的大量 LaTeX 文件也可以轉譯成 MathML 再放在網路，進行分享。

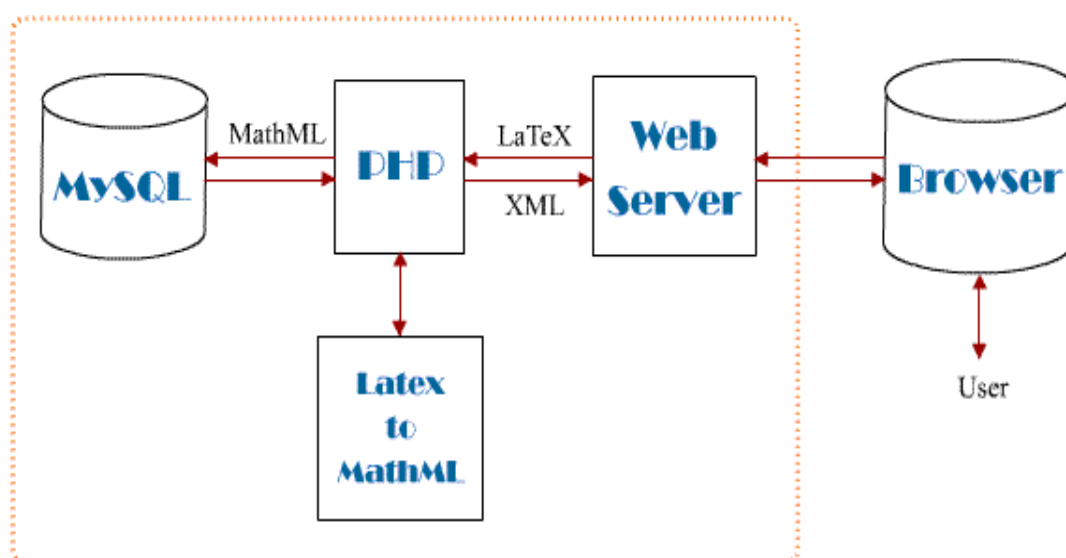
本文所探討的討論版的設計時，不希望採用外掛式的方式，來處理 LaTeX 和 MathML 的轉

譯，因此需先建立 LaTeX2MathML 的 parser 機制(即將 LaTeX 語法建成對應的樹狀結構，再將此樹狀結構轉譯成 MathML)，並將此機制加入討論版的實作。由於目前網路上的討論版製作，以採用 PHP[9]配合 MySQL[10]資料庫為主流，因此本系統也採取這兩者的搭配來實現。

二、系統架構

本系統主要是架構在 Linux 作業系統之上，設計一個適合於 Web Server 上的線上數學討論版。我們使用 PHP 設計一個結合 MySQL 資料庫的討論版，以及 LaTeX to MathML 的轉譯程式。使用者可以經由 PC 連上網際網路，開啓數學討論版網頁；並經由輸入 LaTeX，發表含有數學符號、式子及方程式的文章，因而整個系統的架構如圖依所示，說明如下：

1. 系統設計：存在於 Server 端的 WWW 網站(使用 PHP 語言編纂)。
2. 系統元件：
 - (1) PC 端使用的瀏覽器：IE、Netscape、Mozilla。
 - (2) 網路討論版：使用 PHP 語言及 MySQL 資料庫設計。
 - (3) LaTeX2MathML 轉換程式：使用 PHP 設計。



圖一：系統整體架構

三、LaTeX 到 MathML 之轉譯機制

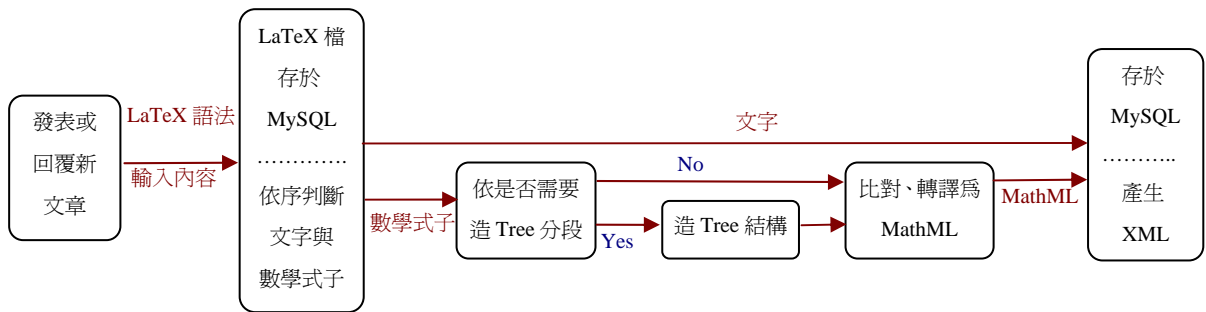
有關 LaTeX 語法轉譯成 MathML 的語言，目前網路上的開放軟體多是採用兩種語法之直接翻譯，並未進行分析。為了提昇程式效率與未來的擴充性，除了盡力 LaTeX 與 MathML 兩者之

間的對照表(轉入資料庫)、與法規那外,尚進行語法分析。此外,轉譯的步驟分成 LaTeX 的語法數建立以及將之轉成 MathML 語法樹等兩大部分,主要程序包含接收輸入、分解與資料庫比對、重組、輸出儲存等,如圖二所示,說明如下。

3.1 LaTeX 語法樹建立

LaTeX 語法樹的建立,依圖二解說如下:

1. 依序判斷文字或數學式區段:



圖二：轉譯處理程序圖

Case 1. 分段封包：可使用串列(linked list)的觀念,依序放入,有兩種處理方式：

- (1) 當陣列的第一個元素非"\$"或是"\$\$"時,則此陣列直到搜尋到"\$"或"\$\$"即結束,最後一個"\$"或"\$\$"不存入陣列。當然其他 LaTeX 表示數學式開頭的符號如" $\begin{equation}$ "," $\begin{eqnarray}$ "等也會一併判斷,以下例同。
- (2) 當陣列的第一個元素為"\$"或是"\$\$"時,則此陣列直到搜尋到"\$"或"\$\$"及結束,最後一個"\$"或"\$\$"存入陣列。

Case 2. 判斷內容：依輸入內容之順序,將依序產生一個個新陣列,每產生一個新陣列,就開始判斷內容並選擇下依程序的處理方法。

- (1) 當判斷陣列之最後一元素非"\$"或"\$\$"時,即為前次分段中的第一類情形,則可判斷此陣列為文字陣列,可直接依序將整個陣列存於 MySQL 中。
- (2) 當判斷陣列之最後一元素為"\$"或"\$\$"時,即為前次分段中的第二類情形,則可判斷此陣列為數學式子陣列,可將此陣列去頭去尾再繼續轉譯為 MathML,即將此陣列去掉第一個及最後一個元素"\$"或"\$\$"再送至下一程序。

2. 依序判斷封包是否需要造 Tree 結構後再執行轉譯,判斷情形如下：只有上列程序之第二類陣列會送至此程序判斷：

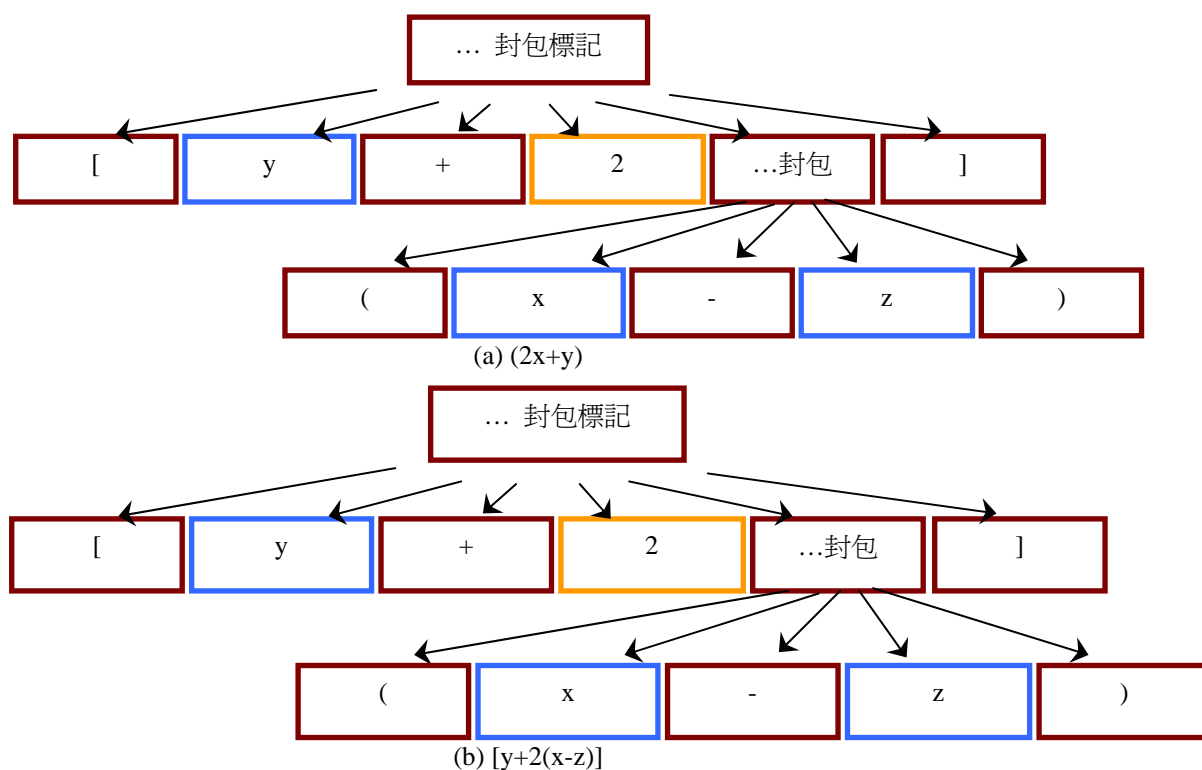
Case 1. 不需形成 Tree 結構,可直接轉譯者；當依序搜尋陣列,無任何符號且開頭不是"\符號之函數表示式時,則可直接將此陣列送去 LaTeX2MathML 轉譯。但要注意的是,若發現"\則需比對符號表,判斷是否可直接找到與 MathML 對應之語法直接轉譯,

若無，則為 Case2。

Case 2. 需先形成樹狀結構者：非 Case1 者，當依序搜尋陣列，發現任何刮號或是開頭為“\”符號之函數表示式時，則依序採分段處理，即遇到需先形成 Tree 結構者才封包，並呼叫程式形成 Tree 結構再進行轉譯。

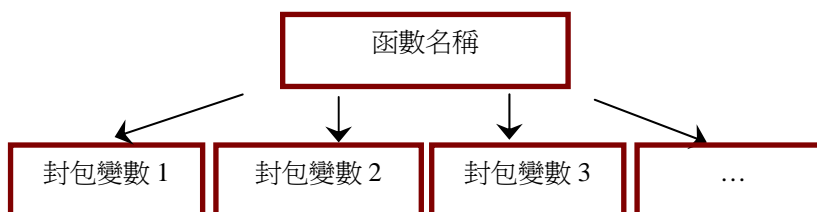
3. 形成 Tree 結構：只有上列程序中屬於 Case2 之部分封包陣列會到此程序：

Case 1. 成對刮號形式：若搜尋得到之封包陣列為成對刮號者，及此段封包陣列之第一與最後一個元素分別為各種形式之左刮極右刮，則屬此類，需先給於一封包標記，以利判斷生成 $\langle mrow \rangle$ 將之涵蓋，例如： $(2x+y)$ 與 $[y+2(x+z)]$ 兩個式子之封包標記如圖三所示。



圖三：數學式所對應之封包標記

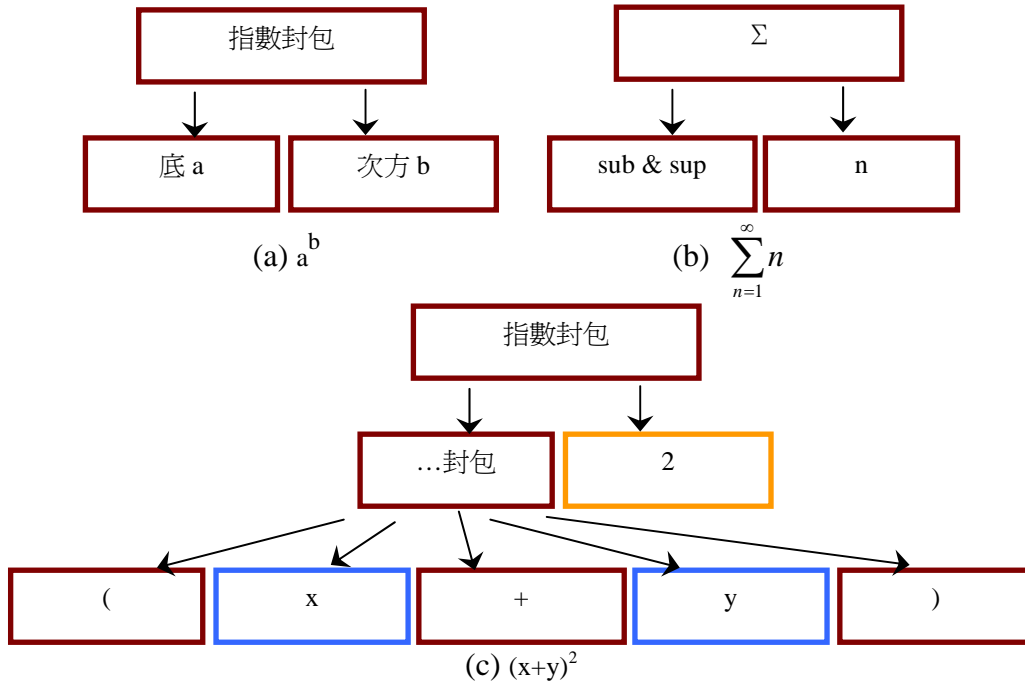
Case 2. 特定函數語法形式：若搜尋得到之封包陣列開頭為“\”，且經比對資料庫後得知緊接著“\”的是特定的 LaTeX 函數語法，則需依照不同函數語法，產生不盡相同之樹狀結構，如圖四所示：



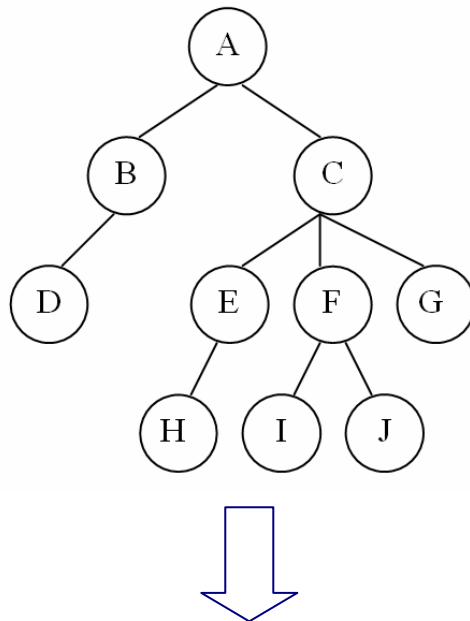
圖四：函數的封包標記

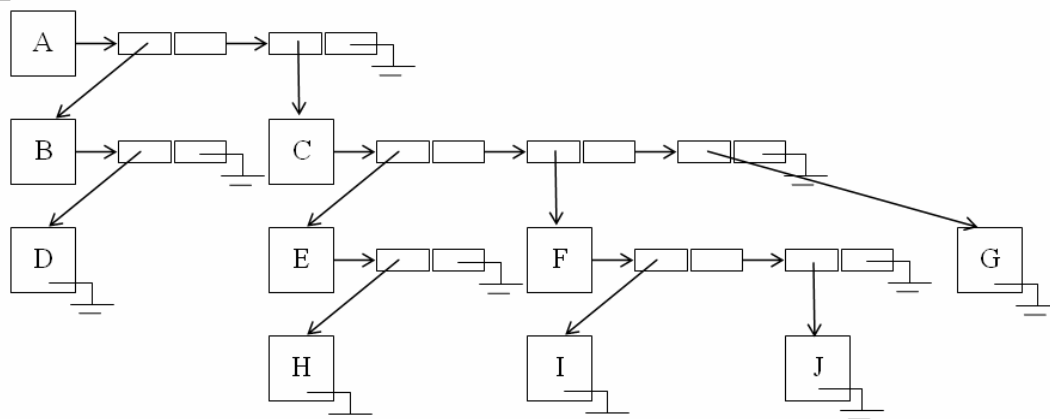
例如：指數函數 a^b 、 $\sum_{n=1}^{\infty} n$ 與 $(x+y)^2$ 之對應封包標記如圖五所示。

一旦 LaTeX 的語法樹分析完後，逐一建立以串列資料型態儲存的樹狀結構，此種樹狀結構如圖六所示。



圖五：常見函數的封包標記





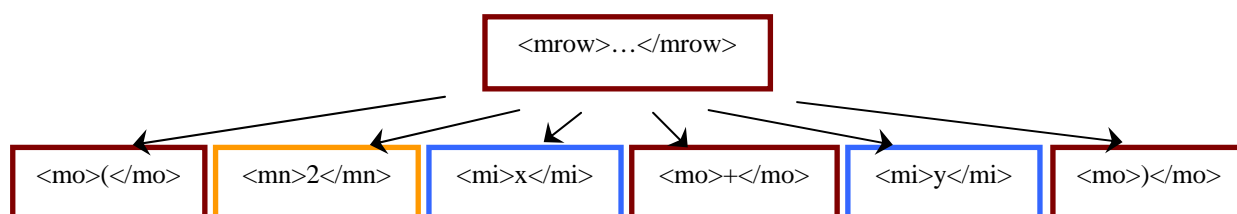
圖六：LaTeX 語法樹之儲存結構

3.2 翻譯為 MathML

圖二所示的轉譯過程中的最後一個程序是轉成 MathML，這階段程序最重要的就是 LaTeX 語法與 MathML 語言之比對，系統會建立一個完整的資料庫來反應兩者之關係，以便能夠很快速地為 LaTeX 語法找到與其對映的 MathML 語言，依語法性質的不同，這類比對表可分為三種，分別是(1)刮號封包，(2)函數符號，(3)一般符號。說明如下：

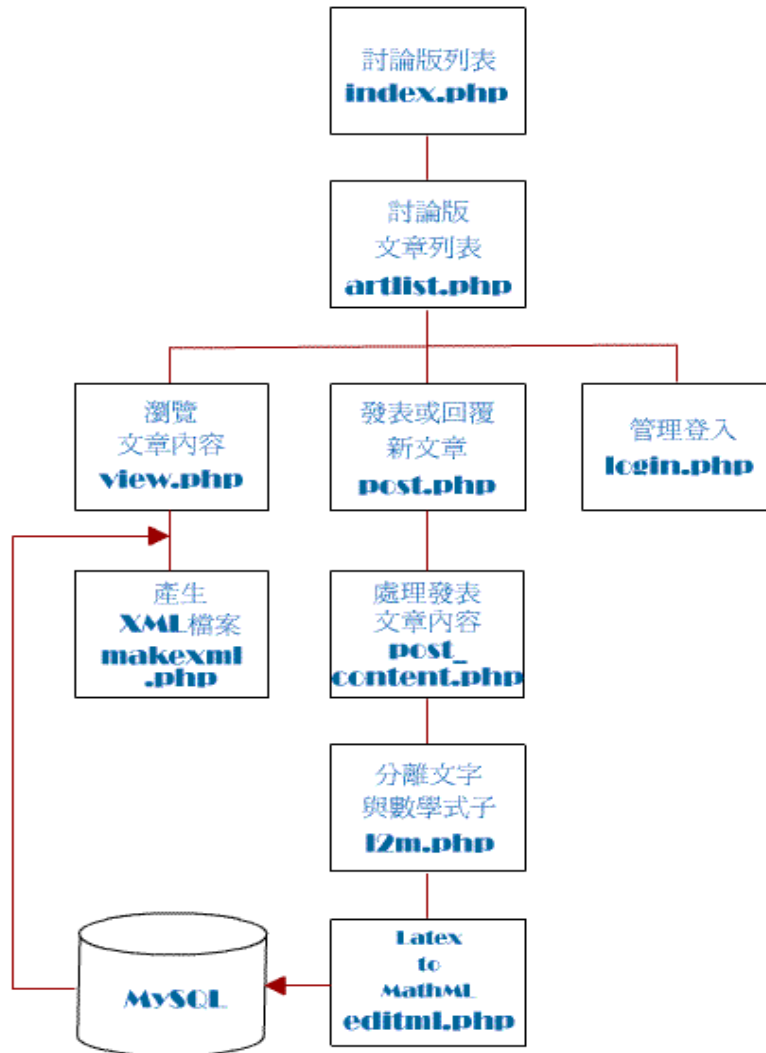
- (1) 刮號封包：刮號封包比對表，放置成對刮號的 LaTeX 程式碼，當比對到此部分的資料，刮號中的內容包含其他刮號符號在內，需先以 `<mrow>` 分段封包，進而再處理其封包內容。
- (2) 函數符號：函數符號比對表，放置各函數的基本語法對映，其各種基本語法均有不同要件。
- (3) 一般符號：至於一般符號比對表，則以 MathML 之分類法來區隔，為 `<mo>`、`<mi>`、`<mn>`，分別整理各 LaTeX 語法中可直接對映之內容。

經由資料庫比對之後，即可形成完整的 MathML 語言。以前文引用的數學式 $(2x+y)$ 為例，其封包標記如圖三(a)所示，建成對映的 LaTeX 語法樹，轉成樹狀結構，在進行比對後，即可組合成 MathML 語言的封包，如圖七所示。樹狀結構的處理，採由下層往上層，且由左而右的搜尋掃描方式，進行重組 MathML 封包。

圖七：數學式 $(2x+y)$ 所對應之 MathML 封包

四、系統實作

以 LaTeX 語法轉譯成 MathML 語言的機制為基礎，利用 PHP 程式語言配合 MySQL 資料庫討論版的進行實作，針對使用者輸入的文章，只要是數學式的部分，都會陸續依序送至此轉譯機制處理，處理完畢後，再與原輸入的純文字部分結合，儲存至資料庫。為了方便未來系統發展的需要，原輸入之 LaTeX 數學式，也以註解的方式併入 MathML 資料而同時儲存之。在使用者瀏覽討論區時，這些儲存在資料庫中的資料，會被 PHP 程式取出，及時產生 xml，而送到使用者端，由瀏覽器(IE, Netscape, Mozilla 等)顯示之。整個系統所需的程式如圖八所示，實作的討論版稱為 Math Messenger，圖九則顯示實際瀏覽時，系統會出現的畫面、對話視窗與數學符號的呈現型態。



圖八：網路討論版程式概圖

Math Messenger		
最新發表日期	討論區	文章數
2003 - 12-05 13:55:32	微積分	7
2003 - 12-05 15:17:18	三角函數	14
2003 - 12-05 19:19:03	幾何	1

©Copyright 2003, MM group

(a) 討論版首頁顯示討論區列表

Math Messenger		三角函數討論版	新增文章 管理登入
日期	作者	主題	回覆
2003 - 12-05 13:55:32	nhuang	New Topics	0
2003 - 11-15 12:31:15	Sandy	特殊角	1
2003 - 11-15 12:26:16	Iris	正弦定理	4
2003 - 11-04 19:56:31	John	積化和差	1
2003 - 11-04 19:15:37	Emma	和角公式	3

(b) 三角函數討論版文章列表

三角函數討論版	和角公式	新增 回覆 管理
2003 - 11-04 19:15:37	Emma (218.162.195.144)	
$\sin(x + y) = \sin x * \cos y + \cos x * \sin y$ $\cos(x + y) = \cos x * \cos y - \sin x * \sin y$		

(c) 三角函數討論版文章內容

三角函數討論版：回覆文章 回覆編號：1

*姓名：

電子信箱：

*主題：

*文章：

$$\sin\frac{\theta}{2} = \sqrt{\frac{1 - \cos\theta}{2}}$$

$$\cos\frac{\theta}{2} = \sqrt{\frac{1 + \cos\theta}{2}}$$

(*表必填欄位)

(d) 輸入 LaTeX 語法的數學式，回覆文章

三角函數討論版	和角公式	新增 回覆 管理
2003 - 11 - 04 19:15:37	Emma (218.162.195.144)	
	$\sin(x+y) = \sin x * \cos y + \cos x * \sin y$ $\cos(x+y) = \cos x * \cos y - \sin x * \sin y$	
2003 - 11 - 04 19:32:56	Debby (218.162.195.144)	
RE: 和角公式	$\sin \frac{\theta}{2} = \sqrt{\frac{1 - \cos \theta}{2}}$ $\cos \frac{\theta}{2} = \sqrt{\frac{1 + \cos \theta}{2}}$	

(e) 回到討論版文章內容，觀看 XML 輸出結果。

圖九：Math Messenger 討論版之圖示

五、 結論與外來研究方向

本文說明線上呈現數學表示法之討論版的設計方法，以及採用 PHP 語言實作。目前完成的系統其功能以能夠呈現一般數學式，足供在網路進行線上討論使用。然而對於研究學者而言，在使用此類討論版時，可能會有其他功能的需求，例如數學式之自動編號、一式多行、複雜矩陣等，也將陸續加入，以使本系統能夠較完整的支援與數學相關之教學、研究、討論等。此外，對於線上展示數學討論版的發展，未來的研究方向為：

- 1 加掛圖形的顯示，甚至加入線上畫圖功能(如支援 SVG 標籤語言)，讓數學與圖形解說能相輔相成。
- 2 如同時下流行的 MSN Messenger 或 Yahoo Messenger 一般，設計數學即時通，以搭配一般的即時通，使即時通不在只是單純的聊天工具，也能支援數學的即時討論。
- 3 設計 MathML2LaTeX parser，將線上討論資料轉成一般文件，方便離線編輯、修改。

誌謝

本研究之部分經費承蒙行政院國科會 NSC 93-2115-M-029-006 計畫補助，謹此誌謝。

參考文獻

- [1] W3C Math Home – <http://www.w3.org/Math/>
- [2] MathML in Mozilla project – <http://www.mozilla.org/projects/mathml>
- [3] Mozilla – <http://www.mozilla.org>
- [4] Pavi Sandhu, The MathML Handbook, Charles River Media, Hingham, 2003.
- [5] MathML International Conference 2000, 2002 – <http://www.mathmlconference.org>
- [6] MathML Software – <http://www.w3.org/Math/implementations.html>
- [7] Amaya – <http://www.w3.org/Amaya/>
- [8] WebEQ – <http://www.dessci.com/en/products/webeq/>
- [9] 交大應數 數學交流網 – <http://forum.math.nctu.edu.tw/>
- [10] mathml@ORCCA – <http://www.orcca.on.ca/MathML/>
- [11] itex2mml – <http://pear.math.pitt.edu/mathzilla/itex2mml.html>
- [12] LaTeX and TeX for Hypertext – <http://www.tug.org/applications/tex4ht/mn.html>
- [13] Web editeur Mathematique – <http://www.mathosphere.net/editeurml/WeM.html>
- [14] A TeX to MathML translator – <http://hutchinson.belmont.ma.us/tth/mml/>
- [15] iTeX to MathML – <http://golem.ph.utexas.edu/~distler/blog/archives/000101.html>
- [16] PHP – <http://www.php.net>
- [17] MySQL – <http://www.mysql.com>
- [18] 專業 PHP 程式設計, Jesus Castagnetto, Harish Rawat, Sacha Schumann, Chris Scollo, Deepak Veliath 著(2000), 許鳴程譯, 碁峰。
- [19] 精通 PHP4&MySQL 動態網頁入門實務 HTML·SQL 與物件導向, 位元文化編著(2000), 文魁。
- [20] Teach Yourself PHP4, Matt Zandstra 著(2001), 李正源編譯, 第三波。
- [21] 自學 MySQL, Mark Maslakowski 著(2001), 陳清豪、袁瑋譯, 博碩文化。

Design of MathML-based Bulletin Board System with LaTeX Syntax Input Support to Mathematical Expressions

Huang-Nan Huang^{*} Pei-Jun Yan[†] Chieh Wang[‡]

Abstract

In this paper, a discussion bulletin board system (using PHP) will be presented to support the teaching and communication of mathematical language. The input interface of our board can accept a LaTeX syntax (surely MathML Syntax) to input mathematical expressions. Once the information with mathematical expressions is received, it will be transformed into MathML (for data storage) according to our own LaTeX-to-MathML mechanism. In order to support the users with different browsers, our PHP program can pack the output into XML format with proper style sheets to be displayed via Internet Explorer (with MathPlayer) or Netscape (or Mozilla). In order to help the users of our system who are not familiar with LaTeX, a graphic user interface for input mathematics will be integrated into the system in the near future.

^{*} Department of Mathematics, Tunghai University, Taichung 407, Taiwan.

[†] Department of Computer Science, National Chiao Tung University, Hsinchu 300, Taiwan.

[‡] Washington High School, Taichung County 411, Taiwan.