

什麼是 XML?

蔡英德

靜宜大學資訊管理系

ytt sai@pu.edu.tw

摘要

XML 是目前的熱門話題之一。政府的電子公文是 XML 的應用，電子商務也要導入 XML 技術，XML 似乎已成為建構電子化社會的希望。本文就從 XML 的起源，功能及應用等方向來介紹此新技術。

前言

申請成績單是學校行政事務裡相當普遍的業務。以前申請人要填寫書面申請表後交予教務處處理；有了電腦後，大家利用文書處理軟體(Word Processor)完成申請表後再交予教務處處理。有了網路後，可利用 WWW 瀏覽器上網至指定的網站中完成申請。隨著 XML 新技術的發明，未來我們可能在任何地方於自己熟悉的文書處理軟體環境直接完成申請。也就是說，教務處可以經由 XML 制定一種表格的格式，其中也許要你填某種資料，你填的時候，用的是你所熟悉的文書系統，方便得很。精彩的是，填好送走，對方收到以後，對方的電腦就可以處理了。

再舉一個例子，甲公司和乙公司有來往，甲公司要向乙公司下定單，乙公司就可以用 XML 制定一種下單的格式，甲公司經由他們所熟悉的文書系統填好定單，乙公司到以後，並不需要人工鍵入所收到的資料，而可以交由電腦直接處理。

於 1986 年成為 ISO 8879 國際標準的標準通用標記語言 (Standard Generalized Markup Language, SGML) 是一種後設語言(Meta-language)，是可描述語言的語言。SGML 提供以標記(Markup)方式來描述文件的樣式外觀與內容結構的方法，進而使得文件的各個部分能被辨識及處理。以 SGML 來描述文件有利於電子資料交換和資訊的傳遞。美國國防部所推動的電腦輔助後勤支援 (Computer Aided Logistics Support, CALS) 計畫就是 SGML 的應用，國防部與國防工業廠商間使用 SGML 表示往來的相關文件來快速地交換資料、投標及獲得武器系統技術規格等。然而，因為 SGML 是通用型語言，在實際導入上有其困難

度，所以未被廣泛地全部採用。

目前的 WWW 網頁標準超文字標記語言 (Hyper Text Makeup Language, HTML) 是 1992 年所推出的簡化版的 SGML，可以描述文件樣式外觀和建立文字的超連結 (Hyperlink)。一般使用者只要運用簡單的標籤 (Tag) 設定就能設計出能互動的文件。由於 HTML 不允許自訂標籤和語法，對於使用其內定的標籤來描述文件的內容結構是相當困難且非常有限。近年來，原來以學術研究為目的的網際網路 (Internet) 已普及至全世界，愈來愈多的電子化應用 (如電子公文、電子商務) 將會透過網際網路來進行，如何建構出網路上共通的資料格式，如何讓各應用的資訊快速地流通等等，都是現在亟待解決的問題。很明顯的，HTML 還無法勝任此重大的任務，我們需要新的標準來達成此項目標。

1998 年 2 月，全球資訊網聯盟 (World Wide Web Consortium, W3C) 制定的可延伸式標記語言 (eXtensible Markup Language, XML) 1.0 (詳見[2]) 也是 SGML 的子集合，它不僅能和 HTML 一樣描述文件的外觀樣式，而且還能描述文件的內容結構。各應用文件的內容結構差異很大，可利用 XML 的自訂標籤功能來定義符合自己應用的標籤，進而清楚地描述文件的樹狀或階層狀的結構。另外，基於 XML 文件與其樣式表 (Stylesheet) 可分開處理與儲存，及文件裡的元件能與資料庫結合等優點，許多需要跨平台交換資料的應用都在開發以 XML 為文件格式的相關技術。目前多數的 WWW 瀏覽器都支援此項新技術。

XML 文件

接下來我們將以例子來簡介 XML 文件。如果你公司的訂單是由許多訂購的項目所組成，而每個項目又都有品名與數量兩個資料，那你可以這樣來描述一張訂單：

```
<訂單>
  <項目>
    <品名> Pentium III 866 </品名>
    <數量> 2 </數量>
  </項目>
  <項目>
    <品名> HP Laserjet 8100 </品名>
    <數量> 1 </數量>
  </項目>
</訂單>
```

上例中的<name>與</ name >分別是元件 *name* 的開始標籤與結束標籤，而在這兩標籤間的資料是元件 *name* 的內容。

那這些標籤要如何定義呢？我們可以利用資料型態定義 (Data Type Definition, DTD)來描述它們：

```
<!ELEMENT 品名 (#PCDATA)>
<!ELEMENT 數量 (#PCDATA)>
<!ELEMENT 項目 (品名, 數量)>
```

上例中定義的元件「品名」與「數量」都是任何不含標籤的資料(#PCDATA)，而元件「項目」包含了元件「品名」與「數量」。實際的 XML 寫法如下所示：

<pre><?xml version="1.0" encoding="BIG5"?> <!DOCTYPE 訂單 SYSTEM "order.dtd"> <訂單> <項目> <品名> Pentium III 866 </品名> <數量> 2 </數量> </項目> <項目> <品名> HP Laserjet 8100 </品名> <數量> 1 </數量> </項目> </訂單></pre>	<pre><?xml version="1.0" encoding="BIG5"?> <!-- order.dtd --> <!ELEMENT 品名 (#PCDATA)> <!ELEMENT 數量 (#PCDATA)> <!ELEMENT 項目 (品名, 數量)></pre>
---	---

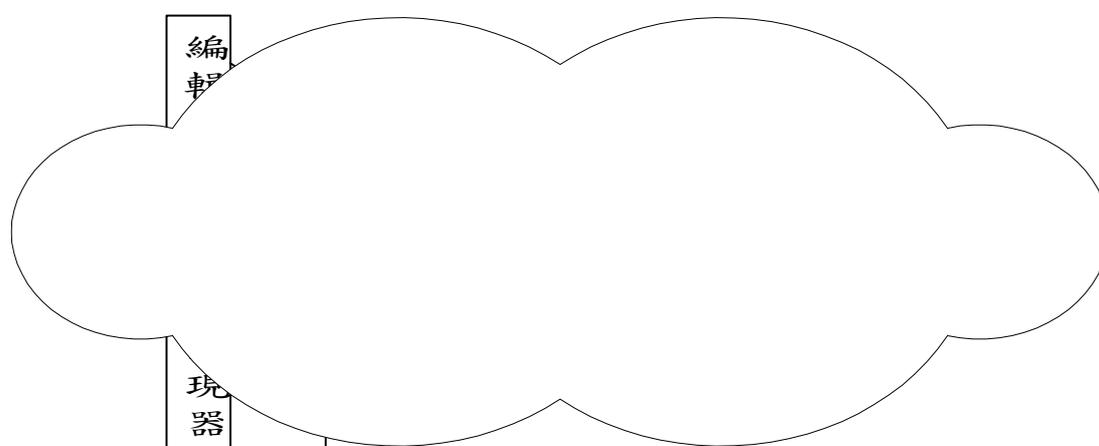
左方是描述訂單的 XML 文件，而右方是 order.dtd 的內容，是訂單所用到的 DTD 定義。這些元件內容的字型名稱、大小與顏色等樣式呈現方式則可能以下列敘述設定：

```
項目 { font-family: "Times New Roman";
      font-size: 12pt
    }
品名 { color: blue
    }
數量 { color: red;
      font-weight: bold
    }
```

XML 的樣式呈現可以 CSS(Cascading Style Sheets) 和 XSL(eXtensible Stylesheet Language)來描述。XML 也採用 DSSSL (Document Style Semantics and Specification Language)的基本規格，能使資料的排列能達到列印品質。

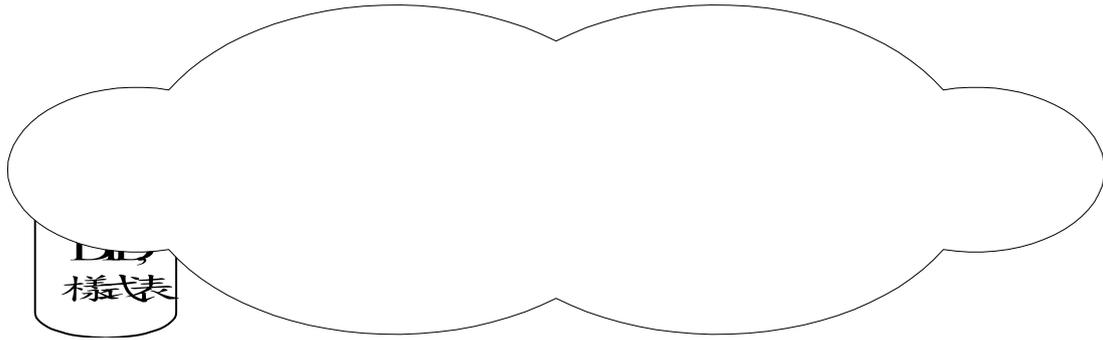
XML 應用

以 XML 建構電子資料交換的平台還需要剖析器(Parser)，呈現器(Viewer)和編輯器(Editor)等軟體的建置。剖析器是檢查 XML 文件是否依規定的標籤，語法與語意描述之；呈現器負責將符合規定的 XML 文件依其指定的外觀樣式顯示出來；編輯器是提供輸入文件的環境，並能輸出符合規定的 XML 格式的文件描述。下圖是一個可能應用於網路資料交換的 XML 環境：



從網際網路進來的資料經解密後還原成 XML 文件，該文件可藉由剖析器確定無誤後交予呈現器顯示其所表示的文件，而使用者能經由編輯器將輸入的文件匯出成 XML 格式後再經加密程序送出。這裡的樣式表可以是事先設計好並放於各應用端中，也可以放於某特定的網路遠端上。另外，XML 文件上記載的資訊可在剖析或編輯時與資料庫互動。

行政院研究發展考核委員會(研考會)現在正積極推動電子公文(詳見[3])，要將各單位間的公文傳遞都改以 XML 格式的公文傳遞，藉以提高政府行政效率達成無紙化辦公室的目標。此系統架構如下所示：



研考會定義了多種制式公文的 DTD 及版面樣式，並都先放於各使用端內，各使用單位的文書處理器（如 Microsoft Word）須要能匯入 XML 文件並顯示出來，還要能將輸入的文件匯出成正確的 XML 文件。所以此文書處理器須具有剖析器呈現器與編輯器的功能。為了確保各單位間 XML 公文傳遞的正確性，研考會要求只能使用通過匯出入 XML 公文之功能認證的文書處理軟體。另外，此處的前置處理器負責與其他單位和國家檔案交換局連線，並加/解密進出的 XML 文件。這樣類似的架構適用於有標準文件往來的環境。

結語

XML 技術將是建立電子化政府，電子商務等願景的希望，許多相關技術與工具似乎還未成熟，仍需要進一步確定與開發。還有一個大問題是如何讓廣大的 HTML 設計師轉換成 XML 設計師呢？W3C 也於 1999 年訂定了 XHTML（參考 [1]），是 HTML + XML 的語言，能讓 HTML 設計師在 HTML 環境中使用 XML。這是一個帶動 XML 應用的新技術，也是邁向效率的、便利的資訊社會之新機會。雖然 XML 許了一個跨平台資料交換的機會，但還須制定各種標準文件或資料的格式，這是社會各界應一起繼續努力的方向。

參考文獻

- [1] XHTML 1.0, <http://www.w3.org/TR/xhtml1/>
- [2] XML 1.0, <http://www.w3.org/TR/REC-xml>
- [3] 文書及檔案管理電腦化作業規範（八十八年修訂版），行政院研究發展考核委員會，<http://www.rdec.gov.tw/doc/edoc/index.htm>