行政院國家科學委員會補助專題研究計畫成果報告

※※※※※※※※※※※※※※※※※※※※※※※※※
※　　　　　　　　　　　　　　　　　　　　　　　　※
※　　　　　　　　　　　　　　　　　　　　　　　　※
※ XML 文件處理技術的研究和以 XML 爲基礎之網際網路應用的發展 ※
※　　　　　　　　　　　　　　　　　　　　　　　　※
※　　　　　　　　　　　　　　　　　　　　　　　　※
※※※※※※※※※※※※※※※※※※※※※※※※※

計畫類別：■個別型計畫　　□整合型計畫

計畫編號：NSC 89－2213－E－004－001－

執行期間： 88 年 8 月 1 日至 89 年 7 月 31 日

計畫主持人 ：陳 正 佳　　國立政治大學資訊科學系

共同主持人 ： 無

本成果報告包括以下應繳交之附件：

□赴國外出差或研習心得報告一份

□赴大陸地區出差或研習心得報告一份

□出席國際學術會議心得報告及發表之論文各一份

□國際合作研究計畫國外研究報告書一份

執行單位：國立政治大學資訊科學系

中 華 民 國 89 年 10 月 15 日

行政院國家科學委員會專題研究計畫成果報告

XML 文件處理技術的研究和以 XML 爲基礎之網際網路應用的發展

The Study of XML Document Processing and the Development of some XML-based Internet Applications

計畫編號 ：NSC **89-2213-E-004-001-**

執行期限：88 年 8 月 1 日至 89 年 7 月 31 日

主持人：陳正佳　　國立政治大學資訊科學系

計畫參與人員：王俊強、黃立昇、 涂富祥 政大資訊科學系

一、中文摘要

XML [3] 是一個通用的資料結構定義語言，可用以定義各種結構化資訊的內容與表達格式，其原設計目的是爲了定義網際網路應用中各種交換資訊的結構與格式。XML 也是一種通用的資訊交換格式。所有依 XML 定義的結構化資訊都必須以 XML 規定的統一語法格式表示。也因此，以 XML 爲基礎的網路架構，爲所有應用程式在資料層次上建立了最基礎的互通性 (interoperability)。

資料格式的統一僅是系統互通的必要條件之一，除此外尚需各種可處理 XML 資料的相關前後端系統。本計劃即是針對 XML 架構下的各種相關規格，處理技術，發展工具與應用領域，做一研習與了解。

應用研習經驗，我們發展了兩個以 XML 爲基礎的 JAVA 應用系統。首先我們以 XML 定義一個標記語言稱爲 JCML，目的是作爲原先 JAVA 定義的二進制目標檔格式的另類表示法。其好處是文字式的 JCAML 內容可使程式設計師方便的直接檢視，修改與產生 JAVA 目標檔。我們爲 JCML 實做了一些輔助程式，可讓使用著輕易地檢視，產生與修改 JCML 的內容，以及在兩種表示法間做格式轉換。其次，我們實做了一個以 XML 爲表達格式的 JAVA 應用程式說明文件產生器。其好處是利用額外的其他 XML 工具，我們可動態的由此基本文件抽取眞正需要的部分並以最合適的包裝呈現給讀著。另外我們嘗試利用 JAVA 2D 實做一個圖形瀏覽器，可呈現 SVG 格式的向量式圖形。SVG 是一個以 XML 定義的向量圖形語言，目前尚在 W3C 研訂之中，將是未來的主流向量圖形語言。最後，在理論上我們證明 W3C 所制定的 XML 文件轉換語言 XSLT 可以計算所有部分遞迴函數 (partially recursive functions) 因此 XSLT 和其他通用程式語言一樣，具有相同的計算與表達能力。

關鍵詞：可延伸式標記語言、XML Namespace， DOM、 SAX、JDOM、XSLT、XPATH、Java byte code、doclet、SVG、JCML

Abstract:

XML[3] is a markup language recommended by W3C in 1998 intended to define the structures and formats of messages transported over the internet. XML is also a data representation format; it requires all message defined by it be encoded in a unifying syntax. XML-based internet architecture establishes a basis on the interoperability of web applications by unifying all data format they receive and transmit.

This project is mainly concerned with the study of related specifications, document processing techniques, development tools and applications of this new language.

Based on the experience learned from the study, we developed two XML applications for Java: (1) we defined an XML-based language called JCML to provide an alternative representation for binary Java class file format. Compared to the original format, XML-based representation has the advantage of being much easier to create, modify, and uncover the content of a java class. Auxiliary

programs have also been developed to help operate on JCML and to help make conversion between both formats. (2) We developed an XML doclet that, when called back by the javadoc program shipped with Sun's Java 2 platform, can produce Java API Documentations in XML format instead of the fixed standard HTML format. XML-based Documentations are much more flexible in that, through pervasive XML tools such as XSLT processors, we can easily extract only parts of the documentation that we really want to view and adorn them in a format which, when rendered, would make readers feel more comfortable.

We also tried to develop a viewer program that can render 2D graphics encoded in SVG format, which is XML-based and is currently under development by W3C. Finally we proved that the transformation language XSLT can compute all partially recursive functions and hence is as powerful as all other general purpose programming languages.

Keywords: XML, XML Namespace, DOM, SAX, JDOM, XSLT, XPATH, Java byte code, Doclet, JCML, SVG

二、緣由與目的

網際網路的技術發展，使得網路應用由以往的資訊專業領域，進展到目前的一般大眾生活。各式各樣的網路應用，從單純的網頁漫遊與資料尋取，到複雜的電子商務與企業網路往來，已悄然而迅速的佔據了大部分的網際網路頻寬。然而現今的全球資訊網依然存有一些架構上的限制，使得某些重要網路應用，如電子商務或夥伴企業往來，在實際系統的開發上，存有一些困難。其主要原因之一是因爲當初發展全球資訊網(World Wide Web)時，主要的目的是希望提供給使用者一個透過網際網路，可快速尋找，瀏覽與抓取相關資料的一個網際網路應用架構。在此架構下的主要互動是使用者與 WEB 伺服器，而目的是經由瀏覽器呈現給給使用者一個親善的資訊閱讀環境。也因此所設計的 HTML 網頁語言，基本上是一種可在原始資料上加註呈現方式的標記語言。透過瀏覽器解譯，HTML 可爲人與網路提供優良的親善介面。 但是當網路應用由人機間的資料尋取進展至 電子商務所需的應用程式與應用程式間資訊交換時，HTML 的優點反而成爲機器間訊息交換的巨大阻礙，其癥結在於額外的資訊呈現方式記載對程式並無任何用處，反而會因此而使得相關資料粹取變得複雜。

然而什麼才是下階段，以機器間互動爲主的應用型態所需的訊息交換語言呢? 爲了因應此種需要，W3C (World Wide Web Consortium)早在 1996 年即開始成立小組，著手規劃此一語言與相關架構。經過兩年發展，終於在 1998 年二月正式推薦一個稱爲 XML (eXtensible Markup Language，可延伸式標記語言) 的新式語言。XML 與 HTML 都是源自 SGML 的所謂標記語言(Markup Language)，允許在資料內容上加註各種標記;然而不同於 HTML 的是，XML 在資料內容上加註的是功能(或語意)標記而非呈現方式標記。此外由於不可能有一組固定數目的功能或語意標記可用以標註網路上所有開放式資料，因此 XML 允許各領域或事業的訊息供需雙方，依據其資訊種類與函意的不同，以及資訊處理的效率需要，而自訂不同標記集。每一標記集即構成一個所謂的標記語言。這些標記語言除了使用不同集合的標記符號外，XML 規格要求他們都必須以符合 XML 規範的語法格式表示資料，就此而言 XML 是一種通用的訊息交換格式;然而就自訂標記集的功能而言，XML 實際上也是一個通用的資料結構定義語言。

本計劃主要目的即是針對 XML 架構下的各種相關規格，處理技術，發展工具與應用領域，做一研習與了解。此外我們希望能應用研習經驗，發展一些以 XML 爲基礎的 應用系統。

三、結果與討論

本計劃計有以下幾項結果

首先，我們詳細的研習了 XML 的相關規格與處理技術，使我們具備開發 XML 相關應用的能力。 除了 XML1.0 規格本身外，

我們也充分的研習並了解 XML 應用系統架構下的其他必要部分與相關規格。此部份至少包括：XML 名稱空間 (XML Namespace[7])、XML 連結語言 (XML Linking Language, Xlink [9])、XML 指標語言 (XML Pointer Language Xpointer[11])、XML 路徑描述語言 (XML Path Langugae, Xpath[10])、XML 樣式描述語言 (Extensible Stylesheet Language, XSL [4])、XML 格式轉換語言 (XSLT[15]) 以及 XML 綱目語言 (XML Schema[12, 13, 14])。在 XML 應用程式介面上，主要的有 DOM [1, 2]、SAX[8] 與 JDOM[6]。

然後，我們發展了兩個以 XML 爲基礎的 JAVA 應用系統。我們以 XML 定義一個標記語言稱爲 JCML，目的是作爲原先 JAVA 定義的二進制目標檔格式的另類表示法。其好處是文字式的 JCAML 內容可使程式設計師方便的直接檢視，修改與產生 JAVA 目標檔。我們爲 JCML 實做了一些輔助程式，可讓使用著輕易地檢視，產生與修改 JCML 的內容，以及在兩種表示法間做格式轉換。很顯然的，此種型態的 XML 應用是非常廣大的，幾乎適用於所有以二進碼表示結果的所有應用系統，例如其他程式語言的目標碼，動畫檔，EDI 資訊或網路訊息等。

另外，我們用 XML 定義一個 JAVA 程式說明文件標記語言，並據此發展一套可以由原始檔產出此格式說明文件的 Doclet。當使用此 XML Doclet 以替代 JAVA 2 上原有的標準 HTML Doclet 之後，我們可以使用 javadoc 產出 XML 格式的 JAVA 說明文件。利用 javadoc 產生的標準 HTML 格式說明文件最大缺點在於其內容與呈現格式已爲固定，以致讀著有時會抱怨其內容過於複雜以致無法快速掌握整理程式架構，然而有時卻又會抱怨其內容不夠完整以致無法知悉程式眞正如何進行。XML 格式的說明文件完全克服了此種缺失。利用額外的其他簡單 XML 工具程式，我們可動態而輕易的由此原始 XML 文件抽取讀著眞正需要檢視的部分，並依讀著喜好與需求，以各種不同的的包裝呈現給讀著。

此外，我們嘗試利用 JAVA 2D 實做一個圖形瀏覽器，可呈現 SVG 格式的向量式圖形。SVG 是一個以 XML 定義的向量圖形語言，目前尚在 W3C 研訂之中，將是未來的主流向量圖形語言。SVG 功能非常強大複雜，可以表示文字、圖形、線條、顏色、字型、動畫，可群組圖形也可使用樣式表，具事件處理與互動功能。由於時間與人力限制，我們僅實做其中的圖形文字群組等簡單部分，其目的在藉此學習如何解譯複雜的 XML 文件。

最後，在理論上，我們證明 XML 文件轉換語言 XSLT 可以計算所有的部分遞迴函數 (partially recursive functions)，因此 XSLT 和其他通用程式語言一樣，具有相同的計算與表達能力。XSLT 是 W3C 制定推薦的一個規則式 XML 文件轉換語言。可用以將 XML 文件轉換成其他 XML 或非 XML 格式文件。由於其描述方法迥異於一般程式語言，因此在 Mailing list 上有人問及 XSLT 的計算複雜度問題，也有人提出一些希望以 XSLT 解決的非 XSLT 應用範疇的問題。事實上早有人點出 XSLT 應是 Turing-complete，然而卻未有正式或嚴謹的證明或說明。我們因此嘗試驗證是否可以用 XSLT 的文件轉換功能來計算所有的部分遞迴函數，答案果然是肯定。也因此有關 XSLT 的計算複雜度問題基本上是沒有意義的。

四、計畫成果自評

本計劃申請時的主要目標有二。一爲研究 XML 的相關規格與處理技術。此部份的工作基本上已達成當初申請目標。另外當初申請時還提出一些構想，希望能發展一些 XML 工具並應用 XML 技術能發一些網路應用。發展 XML 工具的考量之一是鑒於當初申請計劃時ＸＭＬ才剛出現不久，相關的工具還很欠缺以致不易發展應用。然而 XML 的發展進度實在非常神速驚人，當我們還在研習 XML 時，就不斷的看到各種ｘｍｌXML 相關工具，如語法分析器，編輯器，文件轉換器與各種資料庫的 XML 前端包裝等商業或開放軟體工具被發展公佈出來。由於相關工具及原始程式碼已能方便取得，重新實做的意義不大，因此我們乃稍改方向，將重點放在 XML 的應用上。我們以爲我們在應用 XML 於 JAVA 工程上的努力

與結果應可彌補原計劃未被實現的部分。另外爲實現發展網路應用的原始規劃，我們亦將以 Servlets 的型態將本計劃的應用以網路服務的方式發佈，讓使用者可經由網際網路叫用這些 XML 應用程式。

五、參考文獻

[1] Document Object Model(DOM) Level 1 Specification,Version 1.0 W3C Recommendation 1 October, 1998. http://www.w3.org/TR/REC-DOM-level-1

[2] Document Object Model (DOM) Level 2 Core Specification Version 1.0, W3C Proposed Recommendation 27 September, 2000. http://www.w3.org/TR/DOM-Level-2-Core

[3] Extensible Markup Language (XML) 1.0. (second edition) W3C Recommendation 6 October 2000. Editors: Tim Bray, Jean Paoli,C.M.Sperberg-McQueen, and Eve Maler http://www.w3.org/TR/REC-xml

[4] Extensible Stylesheet Language (XSL) Version 1.0, W3C Working Draft 18 October 2000, http://www.w3.org/TR/ xsl.

[5] Java and XML, Brett McLaughlin, O'Reilly & Associates, 2000, ISBN 0-596-00016-2, http://www.oreilly.com/catalog/javaxml

[6] JDOM, http://www.jdom.org/

[7] Namespaces in XML, W3C Recommendation 14 January 1999,Editors: Tim Bray, Dave Hollander,Andrew Layman. http://www.w3.org/TR/REC-xml-names

[8] SAX 2.0: The Simple API for XML, Friday 5 May 2000, David Megginson, http://www.megginson.com/sax/

[9] XML Linking Language (XLink) Version 1.0, W3C Candidate Recommendation 3 July 2000.http://www.w3.org/TR/xlink

[10] XML Path Language (XPath) Version 1.0 W3C Recommendation 16 November 1999, http://www.w3.org/TR/xpath

[11] XML Pointer Language (XPointer) version 1.0, W3C Candidate Recommendation 3 July 2000.http://www.w3.org/TR/xptr

[12] XML Schema Part 0: Primer, W3C Candidate Recommendation, 24 October 2000,Ed. David C. Fallside. http://www. w3.org/TR/xmlschema-0/

[13] XML Schema Part 1: Structures,W3C Candidate Recommendation, 24 October 2000.Ed. Henry S. Thompson, David Beach, et.al.http://www.w3.org/TR/xmlschema-1/

[14] XML Schema Part 2: Datatypes, W3C Candidate Recommendation 24 October 2000, Ed. Paul V. Biron, Ashok Malhotra. http://www.w3.org/TR/xmlscheam-2

[15] XSL Transformations (XSLT) Version 1.0 W3C Recommendation 16 November 1999, Ed.James Clark.http://www.w3.org /TR/ xslt·