

XML 应用研究

Research on the Application of XML

赵书良

(河北师范大学计算机科学系 石家庄050016)

Abstract In this paper, we express what's XML, what we can do by using XML, how the developing environments support the application of XML. We also give some application models of XML.

Keywords XML, XML-QL, DTD, Schema, CCS, XSL, Database

1 引言

HTML 的出现,使互联网迅速普及,由于其简单、易学、通用,从而人人都可以在 Web 上发布文档,随着网络的发展,人们已不满足于在互联网上展示信息,纷纷转向商务应用。但由于 HTML 过于简单的语法、结构,使它很难用于该领域,在这种情况下 XML 诞生了。

XML 的出现和当时的 HTML 出现一样,非常令人激动,以前的数据是以神秘的二进制方式存储的,而现在使用 XML,即使是普通人也可以阅读、创建、修改和控制自己的各种数据,因为现在数据是文本形式的。这种变革的意义是非常深远的。所以目前众多学者纷纷研究 XML,使其应用于自己的领域。

由于 XML 出现较晚,目前 XML 的应用较少,本文将作者多年对 XML 的应用研究写出来,以拓宽读者的思路,同时对 XML 有一个较全面的了解。本文介绍 XML 是什么;XML 能做什么;当今开发环境对 XML 的支持;几个 XML 的项目开发模型;最后对 XML 应用给出总结。

2 XML 究竟是什么?

XML (Extensible Markup Language) 是文档的数字表示形式,是 SGML (Standard Generalized Markup Language) 的一个简化子集,所以 XML 有着深厚的基础,同时与 HTML 同源,所以能很方便地用于 Web。

XML 是一种标记语言,用 Unicode 字符集,采用纯文本表示,XML 靠标记用数字化方法表示文档,把文档变成计算机能阅读的代码,以便计算机能帮我们存储、处理、查询、传输、打印它们。XML 具有很好的扩展性。

XML 文档 = 标记 + 字符数据。通过一系列机制 (主要是其 DTD——Data Type Definition 或其 Schema 机制) 使 XML 文档结构化、数据库化。这样很

容易验证文档数据的合法性,其他程序 (VB、C++、Delphi、脚本程序) 很容易提取 (查询) 文档中的数据,由于 XML 为文本文件,所以能直接利用 HTTP 协议在 Web 中传输,利用 CSS (级联样式表 Cascading Style Sheet) 或 XSL (扩展样式语言 Extensible Style Language) 能在浏览器中实现同一 XML 文档有多种显示形式,以满足 Web 的多种接入设备 (PC、手机等) 的显示要求,利用 XSL 可很方便地将 XML 文档翻译为 HTML 文档。

XML 将网络信息标准化,使得开发者和电脑易于辨认信息。XML 能创建不依赖于平台、语言或限制性格式化协定的开放数据。所以 XML 很适于作为异构系统的中间接口层。

3 XML 包含的内容

XML 共包含如下几部分内容:

1) 文档类型 (结构) 定义: DTD 或 Schema。用于定义文档结构,规范文档内容,是将来对文档查错及查询文档内容的基础。——相当于数据库技术中的表结构定义。

2) XML 文档内容: 包括标记和字符数据。——相当于数据库技术中的表的内容。

3) XML 文档显示格式: CSS 或 XSL。用于定义文档的显示格式。——相当于数据库的客户端应用软件。

4) XML 文档查询语言: XML QL。用于查询 XML 文档数据。——相当于数据库的 SQL。

5) XML 的链接语言: Xlink、Xpointer。用于实现 XML 文档的链接。与 HTML 的链接相似,但在功能上有较大的扩充,例如: 支持双向的链接、多目的地链接等。

4 XML 的应用领域

XML 可应用于如下方面: ①网络客户必须在不同

的数据库间传递信息的应用;②需要把大部分从网络服务器下载的数据放到用户端处理的应用;③相同的数据对于不同的用户需有不同的界面的应用;④网络情报供货商要把发现的信息精心裁减,并发送给不同的个人用户的应用;⑤在 Web 上发布数据,使 Web 数据更有效地管理;⑥自然语言翻译;⑦Web 商务应用。

5 当今技术对 XML 的支持

由于 XML 所具有的优越性,尤其是它所提供的平台无关性,使它非常适于作为应用的中间数据载体,这正好满足了当今的软件三层开发模式的需要,此时客户端和应用服务器间以及应用服务器和数据库服务器间以 XML 数据相互传递;同时,XML 又非常适于 Web 应用,所以世界上有影响的计算机公司纷纷支持 XML,并推出自己的产品(比如:Microsoft、IBM、Sun、Oracle、Inprise、Netscape 等)。目前主要的支持如下:

- Web 浏览器 Internet Explorer 4.0 及以上、Netscape 4.0 支持 XML 文档。

- 常用的开发语言中较新的版本基本上都支持 XML,提供了相应的类,使应用程序可很方便地动态生成 XML 数据(文档);同时提供 XML 解析器以检验文档的合法性;提供相应的函数或方法,以提取 XML 文档数据等。以下语言的较新版本都支持 XML:Java、VB(6.0)、VC++(6.0)、Delphi 5、C++ Builder。

- 一些数据库厂商也纷纷支持 XML,比如:Oracle、Microsoft、IBM。

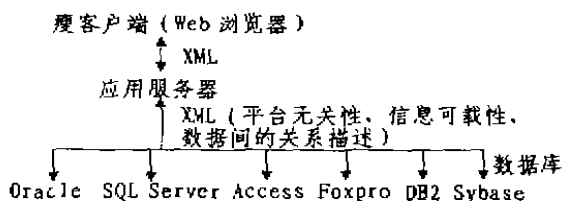
- Web 服务器方面,微软是坚决的支持者,其 IIS、ASP 都提供了相应的支持。

- 微软宣布其下一版本的 Office 支持 XML。

- 相应的 XML 开发工具不断出现。

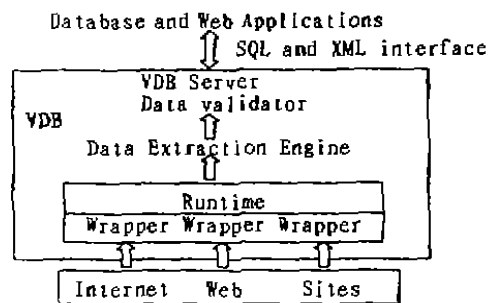
6 几个具体的应用开发模型

6.1 XML 在三层应用模式下的开发模型

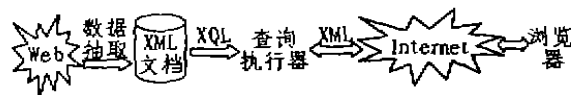


6.2 Internet 数据管理

• Web 信息的整理查询——虚拟数据库(VDB)技术 根据用户提出的查询要求,VDB Server 从相关的 Web 站点提取数据,以 XML 形式将查询结果返给用户,此模型还可用于 Internet 商店比较购物等应用环境。

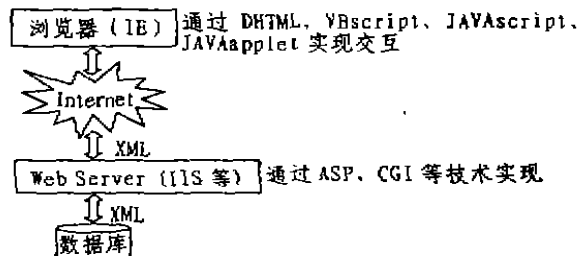


• Web 数据收集



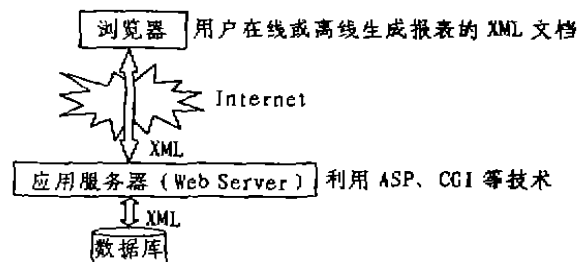
说明:由于目前 Web 页均为 HTML 页,而 HTML 只描述数据表现形式,不描述数据含义及结构,所以目前 Web 页数据抽取还没有好的方法,只能采用半自动化方式,需要人工介入。随着 XML 的使用,情况将会有较大的改观。

6.3 一个交互式 Web 站点的设计模型



6.4 基于 Internet 的远程报表系统的设计模型

该模型适用于税务报表、统计报表的上报等类似的应用。它采用目前先进的瘦客户端的三层设计方案。客户端简化到只有一个 Web 浏览器,用户报表采用 XML 组织数据,可在线或离线编辑(由于 XML 文档为纯文本文档,所以可用任意的文本编辑器编辑生成,加上有相应的 DTD 约束规范,能保证报表的正确性),通过浏览器将 XML 文档(报表)提交。采用这种模式可大幅度减少软件维护费用,报表格式的改变或软件升级,没必要将升级软件分发到路途遥远的客户手中,也没必要在客户端安装数据库软件。同时用户可



(下转第88页)

算法 利用 UML 分析模型计算功能点

1. [确定系统的边界]

循环 对每一个用例图

(1) 将用例添加到系统用例集;

(2) 将执行者添加到执行者集;

2. [计算文件的个数]

循环 对类图中的每个类

(1) 若该类是实体类,并且不是其它类的聚集类,也不是抽象类,则将该类的模板说明是 ILF,则将该类添加到 ILF 集;否则将该类添加到 EIF 集;

3. [计算事务的个数]

循环 对步骤1确定用例集中的每个用例

(1) 找出该用例对应的交互图

(2) 循环 对交互图中的每个消息

① 如果该消息添加到 EI,则将该消息添加到 EI 集;

② 如果该消息的版型是 EO,则将该消息添加到 EO 集;

③ 如果该消息的版型是 EQ,则将该消息添加到 EQ 集;

4. [计算文件的复杂度]

1) 循环 对步骤2确定 ILF 集中的每个类

① 将该类的每个属性计算为该类的的一个 DET;

② 将聚集该类的每个子类(包括子类的聚集后代类)计算为该类的的一个 RET;

③ 由该类的 RET 和 DET 的个数确定该类对应 ILF 的复杂度值和加权因子;

2) 循环 对步骤2确定 EIF 集中的每个类

① 将该类的每个属性计算为该类的的一个 DET;

② 将该类的每个祖先类计算为该类的的一个 RET;

③ 由该类的 RET 和 DET 的个数确定该类对应 EIF 的复杂度值和加权因子;

5. [计算事务的复杂度]

循环 对步骤3确定的 EI、EO 和 EQ 集中的每个消息,由该消息模板说明中的 FTR 和 DET 的个数确定该消息对应事务的复杂度值和加权因子;

6. [计算系统的功能点计数]

对步骤2和步骤3确定的每个功能域,将其复杂度加权因子进行求和,得到的值便是未校正的功能点计数。

5 应用

将该方法用到早期的分析模型上,我们可以估算出项目的大小,将估算值作为 COCOMO2.0 模型的输

入值可以估算出项目开发的成本和持续时间等。而将该方法应用到后期的设计模型上,我们可以收集本次功能点的度量数据,以找出本次估算偏差的原因,并为以后的项目估算提供历史经验数据。

功能点估算的准确性依赖系统需求的准确性及完整性,因此本方法计算结果的准确性也依赖于 UML 分析模型的准确性和完整性。

参考文献

- 1 Fettecke T. Mapping the OO-Jacobson Approach into Function Point Analysis, 1998. <http://user.cs.tu-erlin.de/~fettecke/papers>
- 2 David, Improved Function Point Definitions. <http://www.ifpug.com/Articles/index.htm>
- 3 Ashley N. Measurement as a powerful software management tool. McGraw-Hill International(UK) Ltd. 1995
- 4 Cantor M. Object-Oriented Project Management with UML. Wiley, 1998
- 5 Jacobson I. Applying UML in The Unified Process. <http://www.umlchina.com>
- 6 Jacobson I, Booch G, Rumbaugh J. The Unified Software Development Process. Addison Wesley, 1998
- 7 邱仲藩等译. UML with Rational Rose 从入门到精通. 电子工业出版社, 2000
- 8 黄柏素, 梅宏译. 软件工程——实践者的研究方法. 机械工业出版社, 1999
- 9 刘超, 张莉. 可视化面向对象建模技术——标准建模语言 UML 教程. 北京航空航天大学出版社, 1999

(上接第95页)

通过浏览器查看各种统计报表。

结论 Java 依靠其虚拟机及预编译字节码等技术,能创建跨平台的应用。XML 走的是另一条路,它发展了一种对象传输协议,将基于网络的信息标准化,使得开发者和电脑易于辨认信息。XML 能创建不依赖于平台、语言或限制性格式化协定的开放数据,XML 得到了微软公司的大力支持。

XML 使文档数据库化,为文本文件、便于程序自动生成、提取数据、网络传输。

XML 不能代替数据库,当数据量大时,XML 文档将变得很大或文件将会很多,这时管理效率将成问题。当数据量少时,可代替部分数据库。其实 XML 和数据库侧重点不同,数据库侧重数据管理,而 XML 则侧重于提供不依赖于平台、不依赖于语言的开放数据。

一般认为 XML 和 JAVA 也不冲突,在很多方面增强了 Java。Java 能创建不依赖于平台的应用(是语言级的),而 XML 能提供(创建)不依赖于平台的数据

(是数据级的)。

由于 XML 的优越性能,它将成为异构环境下的通用语言(数据或协议转换);在 Web 领域,它将弥补 HTML 的不足(在短时间内不能代替 HTML,由于 HTML 简单易用,非常适于信息发布,而 XML 更注重数据内容(含义)表示),所以 XML 在 Web 领域将会有更多的应用。由于 XML 的加入,Web 将变得功能更强大、更完善,这时 Web 数据将变得容易管理。

XML 技术虽然发展迅猛,但其应用目前仍处于初级阶段。相信在近几年内将会出现大量的不同领域的各种应用。

参考文献

- 1 Goldfarb C F, 张利, 等译. XML 实用技术. 北京:清华大学出版社, 1999
- 2 冯延晖, 等. XML 完全手册. 北京:中国电力出版社, 2000
- 3 王海燕, 孟小峰, 王珊. 基于 XML 的 Web 信息查询系统 XWIS: 结构与实现. 计算机科学, 2000, 27(10. 增刊)