

XML 技术在移动办公系统中的应用研究

The Research of XML in the Mobile Office System

吴海涛 潘金贵

(计算机软件新技术国家重点实验室 南京大学多媒体计算机研究所 南京210093)

Abstract With the rapid development of the technology of the mobile communication and the Internet, mobile office is a new technology in the application of the network. This paper describes the design and implementation of the mobile office software based on XML and the mobile network application, and emphasizes on the effectual using of XML. The result implies mobile office system decreases the flow of network, and keeps a few network connection, and many other advantages such as being propitious to the search of the information, open, extension, reusage.

Keywords XML, WAP, XML-QL, Encryption/Decryption

1. 引言

当前,在 Internet 上主要是以 HTML 形式传递信息,但使用 HTML 存在以下的缺点:

(1)HTML 的标签大多是设计用来呈现格局和外观的;它对布局、外观方面很擅长,却极度缺乏对内容,也就是对信息涵义的表达力;

(2)由于 HTML 是一个固定的元素集而受限制,它不可扩展,不支持自定义的元素;

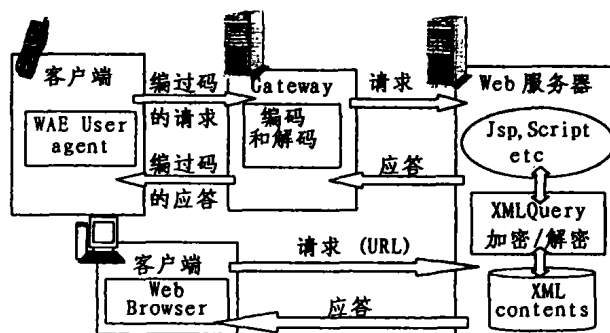
(3)它不具备任何高层次标记语言能力,它不能理解文档的概念,仅知道页面而已。HTML 不具备结构化,用于信息存储效果差。

XML (eXtensible Markup Language) 能克服上述的 HTML 的缺点,具有可扩充性、结构性和形式化等特点,能方便地定义文件类型,实现了文档的内容、展现和操纵的分离,因此,XML 渐渐成为公司内部、公司之间通过 Internet 交换信息的默认标准。为了方便信息的交换,公司内部或公司之间使用了公共的 DTD(Document Type Definition)来定义 XML 的格式。在我们设计和实现的移动办公系统 SKY Server 中,选择 XML 作为信息交换和处理平台。本文概要介绍 SKY Server 的设计思想和实现技术,着重讨论有效应用 XML 的方法。

2. 移动办公系统 SKY Server 结构

近年来,随着移动通讯技术、Internet/Intranet 应用的普及和发展,使得移动办公成为时尚,由于它具有迅速、快速、方便等优势,尤其为那些经常出差的营销人员所青睐。而互联网上的各种商品和技术信息越来越多,所以不管是商家还是个人都想通过一种简单有效的方式来充分利用这些网络资源。移动办公系统通过对积累的信息进行管理,提供使用接口和方法,从而解决了外出较多的营销人员与总部和客户沟通联络不便的问题。

SKY Server 系统以 XML 为信息交换和处理的平台。可以通过两种方法使用移动办公系统:XML 页面在线访问,通过 WAP 无线访问。



注: WAE (Wireless Application Environment)

图1 系统的结构

使用 WAP 访问时,页面是使用 WML 描述的。在字符集、语法构造、数据类型等方面 WML 都沿用或继承了 XML^[1,2]。因此,只需对 XML 描述的页面稍加改动,就能使用 WAP 访问了。

3. 移动办公系统中的信息查询

由图1知,随着服务器端堆积了大量的 XML 文件之后,需要在多个文件之间查找、交换数据,这一重要的应用称为 EDI (Interchange of Electronic Data)。我们可以把大量的 XML 数据看作是“数据库”,把 DTD(Document Type Definition)看作是数据库的“模式”,需要在系统中实现信息查询功能。在实现信息查询功能之前,先了解 XML 数据模型和 W3C (World Wide Web Consortium)的一种 XML 查询语言 XML-QL。

3.1 XML 数据模型的分析

XML 数据模式与传统的关系和面向对象数据的模式不同,表现在:允许先出现复杂结构的 XML 数据,后定义 DTD;DTD 可以随着数据的不断更新而变化;DTD 是用于描述数据的结构信息(如元素出现的顺序,出现的次数),而不是对数据进行强制性的约束(如 PCDATA 可以是数字,也可以是字符)。因此,XML 数据与半结构化的数据非常相似。

为了方便信息查询,我们把 XML 文档转换成半结构化

吴海涛 硕士研究生,研究方向为 XML 技术,电子商务和网上信息检索。潘金贵 教授,博士生导师,从事中间件,Agent 技术及多媒体远程教育等的研究。

数据模型。转换规则如下:XML 标识转换成结点标识;元素属性列表转换成属性名/原子值(attribute-name/atomic-value)序列;对于 XML 文档中 IDREF 类型的属性 i,或者 IDREFS 类型的属性部件 i,增加一个引用子结点<label, eid>,其中 label 是相应的属性名,eid(其 ID = i)是结点号。

这样,在大量的 XML 文档中抽取数据,交换数据,合并数据就可以借助在半结构化数据研究中取得的一些成果。我们参考了在半结构化数据研究方面有一定代表性的工作

```
<!ELEMENT fruitBill ( fruit+, garden+ )>
<!ELEMENT fruit (breed)>
<!ATTLIST fruit FID ID #REQUIRED>
<!ATTLIST fruit price CDATA #REQUIRED>
<!ATTLIST fruit producer IDREF #REQUIRED>
<!ELEMENT breed (#PCDATA)>
<!ELEMENT garden (brand)>
<!ATTLIST garden GID ID #REQUIRED>
<!ATTLIST garden address CDATA>
<!ELEMENT brand (#PCDATA)>
<fruitBill>
  <fruit FID= "001" price= "2.00"
    <breed>apple</breed>
  </fruit>
  <fruit FID= "002" price= "4.00"
    <breed>banana</breed>
  </fruit>
  <garden GID= "100" address = "Nanjing Hankou
    road 12#">
    <brand>Gold Garden</brand>
  </garden>
</fruitBill>
```

Stanford 大学的 Lore 系统^[4]。

在 XML 数据模型中,把 XML 文档用带标签的有向图——XML 图来描述成数据模型。一个 XML 图包括:每一个节点用对象标识符 OID 代表的图 G,用元素的 tag 来标注 G 的边,用属性值的集合标注 G 的节点,用值(字符串)标注 G 的叶节点,用一个特殊的节点表示根(Root)^[3]。例如:存在 XML 文档 Handbill-fruit.xml,对应的 XML 图如图2所示。

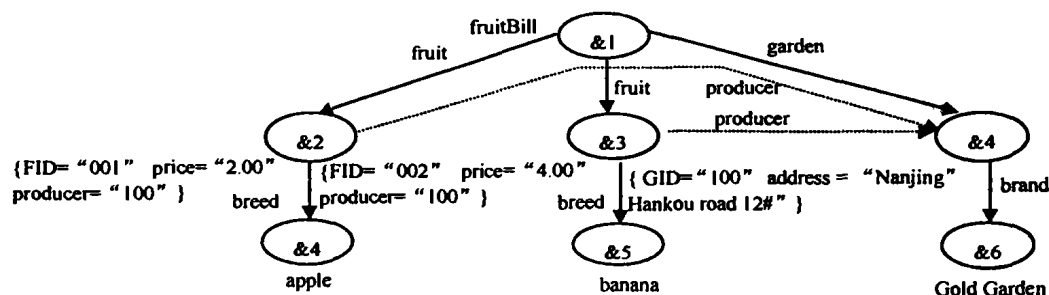


图2 Handbill-fruit.xml 对应的 XML 图

图中节点内的 &1、&2 等是用来表示 eid。需要说明的是上图中有两类“边”，使用 tag 名标识的实线边，使用属性名标识的虚线边。去掉虚线边就是“字面”(literal)上的树状结构；保留虚线边，去掉 {} 中的 IDREF 属性就是“语义”(semantic)上的图状结构。

3.2 XML-QL 查询语言

XML-QL 是针对 XML 的一种查询语言,类似于 SQL 的 SELECT-WHERE 结构并且借助了半结构化数据查询语言的研究成果。它可对 XML 文档进行查询、构造、转换和集成。XML-QL 与 XSL (Extensible Stylesheet Language) 是有区别的,尽管 XML-QL 与 XSL 共享一些方法,XSL 主要是用来描述 XML 文档的风格和显示的,而 XML-QL 则更注重数据内部的操作。通过对 XML-QL 的语句分析,可以把 XML-QL 的语句划分成三个部分:模式部分、条件表达式部分和结果部分。XML-QL 与结构化查询语言的最大的区别在于在查询条件中加入了模式匹配,查询的结果一般用 XML 形式表示^[5]。

XML-QL 的功能十分强大,表现在:除了具有选择和抽取数据,改变 XML 文档结构,合并多个 XML 文档的数据,对文档中的某个 element 进行排序、成组操作之外,还能进行嵌套查询,通过支持 tag 变量进行结构查询。例如:把上述的 Handbill-fruit.xml 中 fruit 元素与 garden 元素调换一下,可以通过如下的 XML-QL 实现。

```
WHERE (fruitBill)
  (fruit) (breed) $ b /breed /fruit } 模式
  (garden) (brand) $ s /brand /garden }
```

```
</fruitBill> IN "Handbill-fruit.xml" } 条件表达式
CONSTRUCT (result ID= produce($ s, $ b))
  (garden) (brand) $ s /brand /garden } 结果(XML 形式)
  (fruit) (breed) $ b /breed /fruit
</result>
```

查询的一个结果是:

```
(result ID= produce(100,001))
  (garden GID= "100" address = "Nanjing Hankou road 12#"
    (brand) Gold Garden /brand)
  /garden)
  (fruit FID= "001" price= "2.00" producer= "100"
    (breed) apple /breed)
  /fruit)
</result>
```

3.3 XML-QL 语言的实现

XMLQuery 是本系统中使用 Java 实现的 XML-QL 语言的子集。实现的主要思想是,借助于 Lore 的 XML 数据模型,先把 XML 文档转化成树状 XML 图,再把查询转化成对树的操作,最后把查询的结果子树表示成符合要求的 XML 形式,实现的流程如图3。

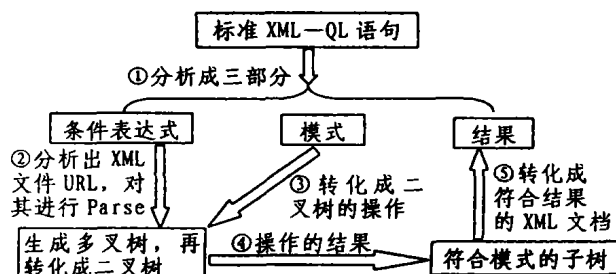


图3 XML-QL 语言实现流程

4. 信息的加密

由于存放在服务器上的 XML 文档是纯文本形式的,任何知道服务器 XML 文档路径的用户都可以通过浏览器访问该页面,获取信息。有些信息是可以公布的,而有些信息是公司内部的机密,是不可以对外公布的;有的信息对 A 是可以获取的,而对 B 是保密的。很多 XML 文档中既包含可公布信息,又包含不可公布的信息,就需要对 XML 中那部分不可公布的信息进行加密;XML 文档中既包含 A 可以获取的信息,又包含对 A 保密的信息,就需要对 XML 中那部分对 A 保密的信息进行加密。常用的加密方法有:

①子树加密。是 Imanura^[7]与 Simon, LaMacchia^[8]提出来的。它们的共同特点是:把子树序列化成文字的表示,使用象对称加密这样的加密算法对文字进行加密,用加密的结果去替换未加密的部分。

②服务器端存取控制。是 Damiani 等^[9]、Bertino 等^[10]和 Ku-do, Hada^[11]提出来的,实现的方法是:存放在服务器端 XML 被解析成 DOM 树;如果客户端有请求,那么请求者必须先被授权,服务器对授权进行检查;根据检查的结果对 DOM 树进行标注;根据标注的结点对树剪枝(把该客户授权以外的结点从树中删除,这里的结点包括树中任何结点);最后把剪枝后的树序列化。

③编号加密。是 Christian Geuer-Pollmann^[6]提出来的。实现的方法是:把 XML 文档转化成 DOM 树;把每个结点编

号并把位置信息(父结点,左兄结点,右兄结点)储存下来;为每个结点/位置信息对构造私钥;根据存取控制列表或者加密要求为每个客户构造钥匙包,该钥匙包中存放客户允许存取的结点的密钥,再对钥匙包进行加密。在存取 XML 文档的时候,客户取得钥匙包,对钥匙包进行解密,对所有的文档元素解密,把原来的树根据结点的位置信息构造起来。

本系统中由于实现了 XML-QL,所以对上述服务器端存取控制进行一些改进,来实现系统的加密功能。实现的思想类似数据库中的视图。实现的方法是:对客户进行分类,对每类客户构造存取控制列表;根据存取控制列表构造用户可见的视图,记录视图与 XML 文档的映射关系;客户从视图中读取数据,隐蔽了未授权给用户的数据;写入数据时,根据映射关系写入 XML 文档。

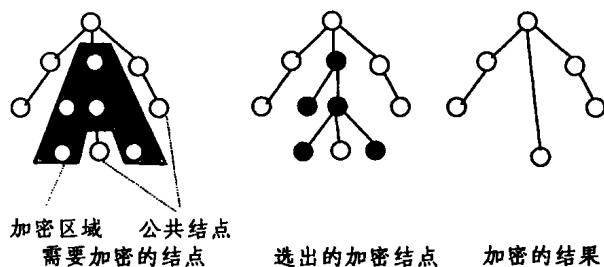


图4 使用 XML-QL 加密的示意图

XML-QL 加密方法与常用加密方法的比较:

模型	子树加密	服务器端存取控制	编号加密	使用 XML-QL
对整个子树加密	可以	可以	可以	可以
仅对属性值加密	可以(依据具体要求)	可以	可以	不可以
子树中非加密结点是否可见	否	是	是	是
是否需要安全的服务器	否	是	否	是
是否自动进行存取/加解密	否(需要样式表)	是	否(需要样式表)	是
是否需要额外的传输安全	否	是	否	否
加入新内容后是否需要重新加密	否	仅当新加入存取控制列表	是	仅当加入新的客户类

5. 系统的实现与运行效果

SKY Server 系统采用 JDK 1.2 在 Turbo Linux 环境下实现,客户端在线部分使用 Windows 2000/NT 来进行测试, WAP 部分使用可上网的 PDA 测试。系统使用后提高了信息处理的速度,大大方便了工作人员相互沟通,提高了工作效率。

参考文献

- 1 WAP Architecture (Version 30-Apr-1998) Wireless Application Protocol Architecture Specification. <http://www.wapforum.org>
- 2 WAP WML (WAP-191-WML 19 February 2000) Wireless Application Protocol Wireless Markup Language Specification Version 1.3. <http://www.wapforum.org>
- 3 XML-QL: A Query Language for XML (Submission to the World Wide Web Consortium 19-August-1998). <http://www.w3.org/TR/1998/NOTE-xml-ql-19980819>

- 4 Goldman R, McHugh J, Widom J. From Semistructured Data to XML: Migrating the Lore Data Model and Query Language <http://www-db.stanford.edu/lore>
- 5 孟小峰,曹巍,王珊. Web 查询技术研究. 计算机科学, 2001, 28 (2)
- 6 Geuer-Pollmann C. XML Encryption and Access Control A comparison and 2nd encryption model Draft 29. 10. 2000
- 7 Imamura. Another proposal of XML Encryption, Takeshi Imamura (Mon, Aug 14 2000)
- 8 LaMacchia S. XML Encryption strawman proposal. Simon and Brian LaMacchia Ed (Wed, Aug. 2000)
- 9 Damiani, et al. Design and Implementation of an Access Control Processor for XML Documents. Ernesto Damiani, Sabrina De Capitani di Vimercati, Stefano Paraboschi and Pierangela Samarati
- 10 Bertino, et al. Controlled Access and Distribution of XML Documents E. Bertino, S. Castano, E. Ferrari, M. Mesiti
- 11 Hada K. XML Document Security and e-Business applications. M. Kudo and S. Hada, Nov. 2000