

文章编号: 1002-0268 (2005) 09-0119-04

XML 技术在交通综合数据库中的应用研究

李瑞敏, 陆化普, 史其信
(清华大学交通研究所, 北京 100084)

摘要: 交通综合数据库是交通综合信息平台的核心和基础。本文首先分析了交通综合数据库的功能以及其所存储的信息类型, 分析了 XML 技术在异构数据源之间进行数据共享和交换方面的优势以及目前许多国家和地区对 XML 在交通数据交换中的研究状况, 在此基础上提出了基于 XML 技术的交通综合数据库的系统结构, 重点研究了利用 XML 实现数据交换的理论与技术途径, 通过在实际研发工作中的应用, 结果表明 XML 技术在交通综合数据库中有良好的应用前景。

关键词: 交通综合数据库; XML; 数据共享

中图分类号: U491.1⁺12

文献标识码: A

Application of XML Technology in Transport Integrated Database

LI Rui-min, LU Hua-pu, SHI Qi-xin

(Institute of Transportation Engineering, Tsinghua University, Beijing 100084, China)

Abstract: Transport Integrated Database (TID) is the base of the Transport Integrated Information Platform. This paper describes a theoretic study and practical development of an XML-based approach to achieve TID function. The authors analyze the function of TID and the data type stored, and the advantages of XML in terms of data sharing and exchange, also the development condition of XML applied in transport data sharing and exchanging fields, and present the system architecture of XML-Based TID, with the theory and technology approach of implement of TID using XML. The application of XML in practical development, proves that it is feasible and successful of XML in TID. Although the research appears limited, a key goal of it was to ensure that the Transport Integrated Information Platform developed could be extended to other transport-related data resources.

Key words: Transport Integrated Database; XML; Data Sharing

交通综合信息平台的主要功能是对来自各个不同 ITS 数据源的数据进行采集、存储、管理和分发, 实现对 ITS 数据的充分有效地利用; 通过对来自不同数据源的数据进行融合挖掘, 产生有用的信息和知识, 为各个 ITS 子系统提供信息支撑; 实现交通信息的增值服务等。交通综合信息平台主要由以下几部分组成: 异构数据源、数据交换层、综合数据库以及基于交通综合信息平台开发的应用系统等。

1 交通综合数据库

1.1 功能

交通综合数据库是实现交通综合信息平台的基础和前提。交通综合数据库的主要功能是: (1) 从一个或多个数据源中采集数据和数据目录并将其存储在一个集中的数据库中; (2) 能够对存储的数据进行数据质量检查、错误通知以及存储等操作; (3) 能够实现对数据

收稿日期: 2004-06-14

基金项目: 科技部十五科技攻关项目 (2002BA404A20B)

作者简介: 李瑞敏 (1979-), 男, 山东莱州人, 博士, 研究方向为智能交通系统等。(liruimin00@mails.tsinghua.edu.cn)

库的维护和运行管理功能以及提供在其基础上进行交通综合信息平台应用子系统的开发等。其目标是在尽可能少地影响各系统自治性的基础上,构造用户所需要的全局数据库,从而支持交通综合信息平台的各类应用以及各数据库之间灵活的信息交换和共享。

对于一个城市或地区而言,并不是需要所有相关部门或应用系统的数据都进行共享,有些数据只有在某个部门或系统内部才有作用,而另外有一些数据出于安全或者保密等原因的考虑,也不能进行完全的共享和交换。因此交通综合数据库按照“只取所需”的基本原则,存储交通综合信息平台所需要的信息以及一些需要进行交换和共享的信息,从而避免了在综合数据库中维护各数据源全部数据的需求。

1.2 存储内容

本文分析的交通综合数据库主要是为城市交通范围内的交通综合信息平台服务的,用于支撑日常的交通管理以及交通服务并且为交通规划、交通决策提供数据基础和信息支撑,因此,其存储内容主要以交通管理相关数据为主,兼有公交、市政等方面的信息。具体如表1所示。

表1 交通综合数据库存储数据类型

类型	内容
空间地理数据	包括与城市交通相关的空间数据以及空间属性数据,例如道路、路口、停车场、交通枢纽、大型场站等城市路网元素以及特征建筑物等的相关数据,如名称、坐标、图层属性、几何尺寸、渠化参数等数据。
交通基础设施数据	除交通空间地理数据外的硬件基础设施数据,包括交通通信设施、交通检测设施(线圈检测器、视频检测器等)、交通控制设备(信号控制器、电子警察、视频监控等)、交通渠化设施(各类标志、标线等)、其他交通设施(如可变信息板等)等。
交通管理业务数据	包括警力配置、管理装备、车辆驾驶员基本信息、违章、事故类基本信息等。
城市交通状况数据	包括实时交通流数据、施工、事故、拥堵、气象、停车场等各类相关数据;以及交通信号控制、视频监控、电子警察、警力配置、信息发布、设备故障等实时数据。
历史数据	主要包括某段时间内(如2年内)有关交通流、事故、违章、设备运行状态等历史数据。
其他相关数据	主要包括公共交通类相关信息(如公交线路、运行时刻表、换乘信息等),气象信息,与交通服务相关的其他黄页信息(如火车、航空、加油站、宾馆等)。

尽管该交通综合数据库的设计是为城市交通综合信息平台服务的,但是由于目前管理上的一些限制,很难实现将公安交通管理部门的数据与其他部门的数据进行实时共享与交换,因此,在实际应用中,笔者针对公安交通管理部门的综合交通信息平台的交通综合数据库进行了研究。

2 XML 技术及其应用

2.1 XML 技术特点

XML (eXtensible Markup Language 可扩展标记语言)是一种元语言,是一种用于描述其他语言的语言。XML的创建是为了在应用程序之间或者应用程序的组件之间能够共享数据,确保在通过网络交换数据时具有良好的可靠性和互操作性。XML具有如下特点^[1]:

(1) XML的语法允许开发者使用普通的文本来自行定义任意复杂度的结构,允许不同系统之间跨平台交换异构的结构化数据,可以有效地解决新旧系统平台、不同应用系统平台之间或者不同数据源之间的数据共享与交换问题。

(2) XML文档只是表示数据本身,实现了数据的结构和显示的分开,为基于Web的应用提供了一个描述数据和交换数据的有效手段。

(3) XML将信息内容和表现形式的合理分开提高了XML信息的可理解性、可交换性和可重用性。

(4) XML提供客户端的包括机制,可以根据多种来源集成数据并将其作为一个文档来显示。

(5) 采用XML来进行信息集成无需对原有的系统做大量的改造即可继续使用等。

2.2 目前的交通领域数据交换技术

EDI是近年来在交通运输领域应用较多的数据交换技术。EDI是为支持事务处理而在异构的平台或应用间用电子格式进行数据交换的一种集成方式。企业间通过共同制定的协议和开发专用接口,将不同公司企业应用或商业运营的数据进行交换和共享^[2]。

然而EDI在应用中存在着如下一些不足,如:新组织加入时,必须制定新的协议和编写新的专用接口;使用固定的数据格式,限制较为严格,缺乏弹性;使用增值网络,需要专用的软硬件系统及专业技术人员,运营成本较高。因此EDI的大规模普及应用受到了极大的阻碍,且应用领域非常有限。

同时,随着ITS的发展,为了在ITS各子系统之间进行有效数据交换与共享,各个国家和地区开始制定新的数据交换标准。

DATEX是目前欧洲用于交通数据交换和数据库开发的标准,包括了DATEX数据字典和信息交换规范。由于DATEX提供了一个标准的,公有的数据交换方式,因此在欧洲获得广泛的应用,然而依然存在着一些问题^[3]。美国ITS Architecture则是美国ITS领域信息和数据交换的标准。

2.3 XML在交通数据交换中的应用

XML本身所具有的特征使其成为目前信息资源集成的很好的技术和工具,也为交通综合信息平台

交通综合数据库的实现提供了可行的解决方案。XML是为分布式环境所开发的,非常适合于ITS的应用环境。同时其数据格式是公用的,可以为多个用户所使用。XML在数据交换等方面所体现出来的优势使人们开始将其应用到交通信息共享与交换中,并取得了良好的效果。

目前美国、欧洲、日本都在进行将XML应用到交通信息领域中的标准化的工作。欧洲的TRIDENT项目是为了克服DATEX的一些不足而由欧洲ITS组织ERTICO实施的,其研究目标就是建立一种不同交通方式不同运营者之间交换和共享数据所需要的通用的和可重复使用的机制。同样美国的SAE ATIS项目的主要内容也是研究将ITS数据字典转换到XML格式。日本也组织了Road Web Markup Language项目进行相关研究。

3 基于XML的交通综合数据库

交通综合信息平台的系统结构采取的是B/W/S结构,即客户端、Web服务器和数据库服务器3个层次。其中交通综合数据库位于数据库服务器的层次上,是该层次的核心。在此不去详细介绍交通综合信息平台的各个层次(另有专文介绍),而重点介绍其中与交通综合数据库有关的部分。如图1所示为交通综合数据库的逻辑体系结构。

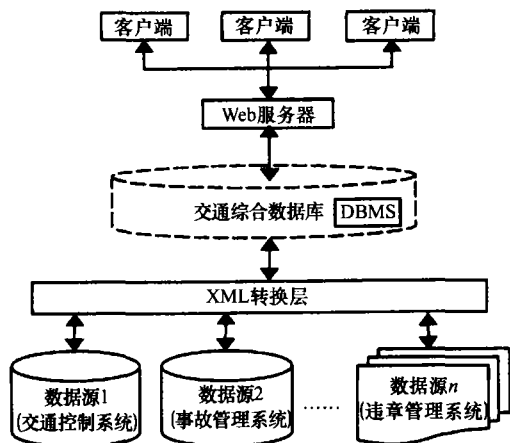


图1 交通综合数据库体系结构

图1中,各个数据源作为交通综合数据库中数据的来源,其构成多种多样,且具有分布性、异构性等特点。对城市交通管理而言,主要包括交通控制系统、事故管理系统、违章管理系统、业务管理系统、信息服务系统、气象、市政等方面的数据源。交通综合数据库支持的数据源类型包括关系型数据库、实时数据库、Web数据库及XML文档等。

XML转换层实现对异构数据源的转换与复制功能,其对各个数据源的请求采用的是标准的XML格式,并将其转换为各数据源的标准查询语言,提交给相应的数据源进行处理。处理完成后,还将返回结果重新封装成为XML格式的数据,并进行发送。XML转换层把各类不同数据源的数据结构定义以DTD的形式统一表现,其提供的底层数据类型和数据结构信息为信息的集成提供了一种数据规格,同时也支持数据源结构的改变能够及时通知应用层。

Web服务器主要是提供基于交通综合数据库开发的应用系统的运行环境等条件,而客户端则是用户访问这些系统的人机接口,在此不再细述。

4 数据交换

在研究过程中,笔者采用Oracle9i的数据库系统环境建立了交通综合信息平台的综合数据库,采用Oracle8i数据库系统建立了城市交通信号控制系统的数据库,下面分析从信号控制系统的数据库向交通综合数据库的数据集成过程及关键点。

从信号控制系统数据库向交通综合数据库的数据转换主要经过以下4个步骤:(1)数据变化记录:利用大型关系型数据库的触发器机制在信号控制系统数据库中埋入触发器,当其数据有变化时(如新增、修改、删除)触发器在信号控制系统数据库中记录相应的数据变化;(2)打包抽取:根据配置要求或在网络不忙时将此变化数据打成专用的XML数据包,并将此数据包放在公共数据传输平台上;(3)数据传输:将数据包稳定可靠地传送到数据中心;(4)数据入库:当数据到达数据中心后,解析程序负责分析其数据类型及其完整性、有效性,对安全完整的数据按照一定的规则将其分别更新到综合数据库的不同的表中。在此过程中有两个关键点:数据库结构映射到XML文档、XML文档映射到数据库结构。

数据库结构映射到XML文档主要进行如下步骤:(1)确定进行映射的范围;(2)根据不同主题将不同数据表的内容对应的生成一个XML文档的元素;(3)将每个简单的字段映射为元素的属性;(4)对于较为复杂的字段映射为元素的子元素;(5)直到将所有进行转换的数据表映射完毕。在此过程中有一些问题需要注意,例如^[4]:数据库中表名和列名的表示,由于某些数据库对其不分大小写,因此在实现的过程要将XML文档中的节点和属性的名称记录下来,以便在转换时保持XML文档的正确性。以及除去无用的最上层的根节点、处理空值等操作。

交通综合数据库接收到 XML 文档后,按以下步骤在数据库中生成相应的数据表:(1)对树状结构的 XML 文档按深度有限的原则从树叶向树根搜索;(2)每碰到一个内部节点(除去叶节点和根节点),查看数据库中对应的抽象数据类型是否已经存在,若不存在,则创建对应的抽象数据类型;(3)重复上步直到根节点的直接子节点(次根节点);(4)创建基于次根节点所对应抽象数据类型的对象,该对象就是根节点对应的对象;(5)将 XML 文档中各元素的属性值填入对象中的相应域^[5]。

在完成数据交换过程之后,交通综合数据库的一个主要任务就是数据处理问题,包括数据融合、数据挖掘等过程。数据处理问题不是本文的核心,将另文进行分析。

5 应用开发

下面结合一个实例来说明如何用 XML 实现数据库之间的数据交换。

在信号控制系统的数据库中创建有关于路口交通流量的数据表:

```
create table TBL-FLOWDATA
(
  DEVICEID NUMBER (8) not null,
  FLOWDATE DATE not null,
  HOUR NUMBER (2) not null,
  MINUTE NUMBER (2) not null,
  LANEID NUMBER (2) not null,
  LANETYPE NUMBER (1) default 1 not null,
  FLUX NUMBER (3) not null,
  LANOCP NUMBER (3) not null,
  SPEED NUMBER (3),
)
```

按照上面实现数据交换的步骤,首先将数据库中的某条信息产生类似如下结构的 XML 文档:

```
<? xml version = " 1.0" encoding = " GB2312"?
>
< FLOWDATA >
  < DEVICEID > 10003 < /DEVICEID >
  < FLOWDATE > 2004 - 2 - 3 < /FLOWDATE >
  < HOUR > 7 < /HOUR >
  < MINUTE > 5 < /MINUTE >
```

```
< LANEID > 4 < /LANEID >
< LANETYPE > 1 < /LANETYPE >
< FLUX > 10 < /FLUX >
< LANOCP > 4 < /LANOCP >
< SPEED > 20 < /SPEED >
< /FLOWDATA >
```

交通综合数据库接收到 XML 文档后按照上面实现数据交换的步骤将 XML 文档中各元素的属性值填入数据表的相应域中,这样数据库中的数据通过 XML 文档作为中间文件实现了交换。

6 结论

在智能交通系统的发展过程中,不同应用系统之间的数据交换及共享问题是交通领域一直希望解决的问题,而各种数据源之间不同的数据格式、不同的操作系统平台等因素使得该目标的实现非常困难。XML 技术的出现使得对于各种交通信息的标准化描述成为可能,同时,数据库中的数据在被转换成 XML 文档之后可以直接在 Web 页上进行发布,在未来 XML 将成为建立交通综合数据库而进行的数据交换的一种有效方式。本文重点研究了基于 XML 技术来实现交通综合数据库的数据集成理论与技术途径。随着数据仓库技术的发展,未来交通综合数据库将向着数据仓库的方向发展,相关研究表明,XML 在数据仓库的建立过程中可以发挥同样的作用^[6,7]。

参考文献:

- [1] Chuck White, Liam Quin, Linda Burman, 著,周生炳,宋浩,肖伟,等译.XML从入门到精通[M].北京:电子工业出版社,2002.
- [2] 罗国福,施法中.基于XML和移动代理的企业资源信息集成方法研究[J].计算机应用研究,2004,21(3):36-38.
- [3] Kaltwasser J, Hübner D. European Traffic Information Backbone: The Datex Lesson. ITS Conference Torino2000 [C].2000.
- [4] 任庆东,苏斐,李井辉.利用XML实现异源数据库中的数据交换[J].计算机应用研究,2001,18(12):129-130.
- [5] 徐周,黄上腾.基于XML实现数据库间信息交换的方法[J].计算机工程,2001,27(4):177-179.
- [6] 鲍军鹏,张丽霞,刘晓东,沈钧毅.一个基于XML数据仓库系统的设计与实现[J].计算机工程与应用,2004,40(6):220-222.
- [7] 孙焕良,李彤,吕立,张晓山.基于XML技术的数据仓库多维数据模型[J].小型微型计算机系统,2002,23(11):1306-1309.