

XML 基础上的多媒体数据技术研究

张彦军

(暨南大学, 广东 广州 510610)

摘要: 数据的转换与传输技术是 DMB EPG 前端系统的核心技术。通过深入研究 DMB EPG 的协议、技术背景和工作原理,探讨了 DMB EPG 前端系统的数据转换技术,并设计实现了基于 XML 的 DMB EPG 数据转换技术,即提出了一种基于 Schema 模式的数据转换方法。

关键词: 数字多媒体广播; 电子节目指南; 多媒体对象传输

中图分类号: TP393

文献标识码: A

文章编号: 1674-7720(2011)15-0060-04

Multimedia data technology research based on XML

Zhang Yanjun

(Jinan University, Guangzhou 510610, China)

Abstract: Data conversion and transportation is the most important technology of DMB EPG front system. This paper firstly analyzes and studies a series of protocols, backgrounds and principles of DMB EPG. Then the data conversion of DMB EPG front system is discussed, and the data conversion based on XML and is designed and implemented. Additionally, a data conversion method based on schema is presented.

Key words: digital multimedia broadcasting; electronic program guide; multimedia object transmitting

采用 XML 格式来描述 EPG (Electronical Program Guide) 的数据理由有:(1)标准问题。XML 是描述结构化数据的一个非常好的标准。(2)扩展性。由于 EPG 数据类型随着广播电视的发展必然会有新的数据类型出现,而 XML 非常有利于未来扩展向后兼容。(3)随着 XML 技术的不断发展成熟,关于 XML 技术的相关开发工具日益丰富,为以后从事 EPG 相关开发及功能扩展提供了便利。

为此,引出本文所涉及的一般数据转换为 XML 格式数据的问题,这里的一般数据指无结构数据或关系数据库中的结构数据。在研究了已有的数据转换模式的基础上,本文提出了一种基于 Schema 模式的将一般数据转换为 XML 文档的方法。

1 DMB EPG 系统

数字多媒体广播 DMB (Digital Multimedia Broadcasting)是在 DAB(Digital Audio Broadcasting)基础上发展起来的,DMB 充分地利用了 DAB 能在高速移动环境下可靠接收信号的技术优势,在功能上将原来传输单一的音频信息扩展为数据、文字、图形与视频等多种载体,即多媒体对象传输 MOT (Multimedia Object

Transmitting)。DMB 将数字化了的音频、视频信号及各种数据业务信号,在数字状态下进行压缩、编码、调制、传输等处理,可实现高质量传输、很高的功率效率和频谱效率,同时兼具多媒体特性,提供容量大、效率高、可靠性强的数据信息传送。从 DAB 到 DMB,意味着从数字音频广播到数字多媒体广播的跨越,使任何数字信息都可以用一个数字化的平台系统来传递,其系统可以为用户提供包括音频、视频在内的综合视听信息服务和娱乐享受^[1]。

电子节目指南 EPG 用来提供包括音频和数据业务在内的节目列表信息,也是一种用户选择业务、节目及其相关内容的机制。DMB EPG 系统可以分为前端系统和后端系统,前端系统就是通过广电网络发送 EPG 信息的发送端,后端系统指接收 EPG 信息的接收终端,主要是移动手机。EPG 系统的关键是数据格式规范,通过规范化的数据格式协议,使所有不同种类支持该协议的接收终端都能正确接收并展示信息。

2 DMB EPG 系统的数据规范解析

DMB EPG 系统的数据规范也就是其定义的数据类型的结构规范,它有一套专门的标准,全称是 Digital

Audio Broadcasting XML Specification for DAB Electronical Programme Guide^[2],它是 ETSI(European Telecommunications Standard Institute)的技术规范。由于 DMB 是在 DAB 的基础上发展起来的,所以其 EPG 标准目前仍采用 DAB EPG 数据规范,文中的 DMB EPG 也可称为 DAB EPG。

DMB EPG 的数据可以分为 3 个主要组成部分,分别是业务数据 (Service Information)、节目数据(Schedule Information)和节目的分组信息(Group Information)。一个 Service Information 下可以有多个 ensemble,一个 ensemble 下可以有多个在其上播送的 Service。一个 Schedule information 下面可以有多个节目表 Schedule(描述一个或多个 Service 下的节目信息),一个 Schedule 描述多个节目,每个节目又可以包含多个节目事件 (Event)。一个 Group Information 包含多个 Group,一个 Group 下又可以出现多个 Group,而且可以出现多层 Group。这更像一个网状结构,每个 Group 子节点可以隶属于多个 Group 父节点,最底层的节点是节目 Programme。

EPG 的数据是以 XML 文件格式存放的,它的数据结构采用 XML Schema 格式定义,EPG 的 XML 文档模式主要有 3 个:Common data types:epgDataTypes_11.xsd,Schedules:epgSchedule_11.xsd,Service Information:epgSI_11.xsd。

3 一般数据转换为 XML 文档的方法

3.1 基于模板的数据转换方法

基于模板的转换方法并非事先定义好 XML 文档与其他文档之间的映射关系,而是在 XML 文档中嵌入一些可执行的命令。这些指令在转换过程中被系统识别和执行,执行的结果被替换到指令所在的位置,从而生成目标 XML 文档。基于模版的数据转换方法的架构图如图 1 所示。以关系数据库数据为例,为了从数据库中获得节目信息,并将节目信息用 XML 文档表示出来,可以定义以下的一个模板:

```
<? xml version="1.0">
<ProgrammeGroup>
<Programme>
<SelectStmnt>
SELECT      shortName , mediumName , longName , KeyWords
From  Programme
</ SelectStmnt >
</ Programme>
</ProgrammeGroup>
```

如果生成 XML 文档时,系统扫描这个模板,当遇到<SelectStmnt>指令的时候,系统识别出这是一个可执行指令,于是调用指令执行程序执行该指令。指令执行后的结果为:

```
<? xml version="1.0">
<ProgrammeGroup>
<Programme>
<shortName>Apple Pie</shortName>
```

```
<mediumName>Apple Pie version 2</mediumName>
<longName >Apple Pie version 2 From America </
longName>
<Keywords> Apple</Keywords>
</Programme>
</ProgrammeGroup>
```

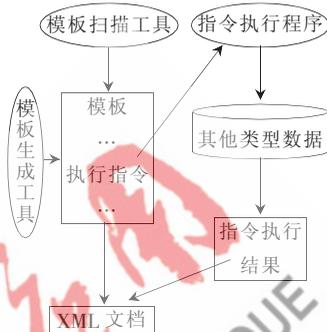


图1 基于模板的数据转换方法架构图

3.2 基于模型的数据转换方法

基于模型的数据转换方法用事先定义好的数据模型来映射 XML 文档结构与其他格式数据的结构之间的关系。以数据库为例,一个最为简单的模型就是将文档结构定义为如下的模型:

```
<database>
<table>
<row>
<column1>...</column1>
<column2>...</column2>
</table>
</database>
```

把数据库数据转换成 XML 文档时,只要把一个表或一个查询结果的数据插入到相应位置即可;而把 XML 文档数据转换成数据库数据时,只要把内容插入到相应的表中即可。另一种常用的模型是将 XML 文档的结构定义为一棵数据对象树,根据规则将文档的层次结构转换为树状结构(通常是把文档中的元素定义为树的节点)。这种模型对于 XML 文档与面向对象数据库和层次数据库之间的转换是非常方便的。当与关系数据库进行转换时,可以利用传统的“对象-关系”映射技术来实现。图 2 为基于模型的转换方法的架构图。

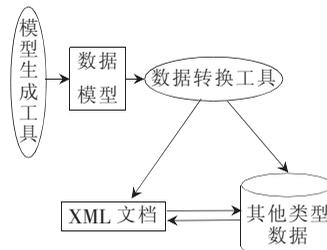


图2 基于模型的数据转换方法架构图

3.3 基于 Schema 模式的数据转换方法

基于模板的转换方法的好处在于转换的步骤比较

《微型机与应用》2011年第30卷第15期

简单,只要给出模板,就可以快速地生成相应的 XML 文档,如图 2 所示。不足之处在于,它只适合将其他类型的数据转换为 XML 文档,对于反向的转换就无能为力了。此外,基于模板的转换方法关键是要生成大量合理的模板,为此,系统需要为用户提供一套生成模板的工具以及相应的指令执行程序。对于数据库数据,可以借用数据库管理系统方便地生成指令执行程序;而文本数据和 OCR 数据指令执行程序的编写则需要大量的工作^[3]。

基于模型的转换方法由于有数据模型的支持,转换工作相对比较简单,并且可以完成 XML 数据与其他格式数据之间的双向转换。但是模型的引入使得 XML 文档的结构受到了一些限制,即一个 XML 文档必须符合模型所规定的结构,才能将 XML 文档转换成其他类型的数据,而从其他类型数据转换得到的 XML 文档也具有某种结构特点。所以,基于模型的转换方法的关键是设计一个灵活的映射模型,使得对 XML 文档结构的限制尽量地少^[3]。

由于 DMB EPG 的 XML 规范是基于 Schema 模式的,为此,本文在已有的数据转换模式的基础上^[3-5],提出一种基于 Schema 模式的数据转换方法,它本质上也是一种基于模型的数据转换方法,只是其文档结构是根据 Schema 模式文件来确定的。基于 Schema 模式的数据转换方法的过程如下:

(1) 首先根据 Schema 模式文件生成一棵元素节点树,树的每个节点对应 Schema 模式文件中每一个元素,树的根节点对应模式文件的根元素,每个节点的结构包含属性集、子元素集以及节点值 CDATA,这三部分都可能为空。每个属性包含属性名、属性类型、属性值,其中,Schema 模式文件定义了属性名和属性类型。每个子元素的结构都由模式文件定义,其中每个子节点的出现次数也都由 Schema 模式文件定义。复杂数据类型的节点一般不存在 CDATA 部分。

(2) 利用 Schema 解析工具(如开源的 xmlBeans)生成的原始的、模糊的元素节点树,之所以称为原始模糊树,是因为树的具体结构还没有确定,只是一个树的模糊框架。主要由以下几个方面构成其不确定性:

① 元素出现的次数引起的不确定性。例如,规定出现 0 次或 1 次,那么该元素是否出现则在后面的数据插入转换过程中确定。

② 几个元素只能有一个出现,也就是 Schema 中 Choice 选项,最终哪个元素出现也由后面的数据插入转换过程中确定。

(3) 数据映射。数据的转换必须要有相应的数据源,这就要确定 XML 文档的元素节点与数据源之间的映射关系。

① 对于非结构化的数据,需要建立一个相应的数据源数据与树节点的映射表。

② 对于结构化的数据源,如关系数据库,可以建立如下的数据映射关系:

一棵元素节点树就是一个 Schema 所包含的元素之间的关系树,它的构成如下:

- 元素树的每一个节点对应 Schema 中的一个元素。
- 每个节点包含以下主要信息:属性列表、子元素列表和其他信息。
- 属性列表包含了该元素的所有属性,每个属性是一个三元组(属性名、属性值、属性类型)。
- 子元素列表包含了该元素的所有子元素,每个子元素对应一个新的节点。
- 除了表节点和内容模式节点以外的节点称为字段节点,它只对父节点所对应的表中的一个字段。
- 内容模式节点不对应数据库中的任何对象。

在实际应用中,数据库的相应表结构正是根据 Schema 模式文件来设计的,这就建立了相应的映射关系,在数据转换过程中,只要找到数据库数据表与树节点的映射关系即可。

(4) 数据转换,XML 文档树生成。前面已经提过,最初的元素节点树是原始的模糊树,而最终的数据转换过程正是 XML 文档树的生成确定过程。数据转换在基于前面原始模糊树的基础上,根据映射关系表找到每个节点对应的数据库中的数据表,并通过查询获取数据。然后将该节点的相应属性赋值,并确定子节点及每个子节点出现的次数。

基于 Schema 模式的数据转换方法的框架如图 3 所示。

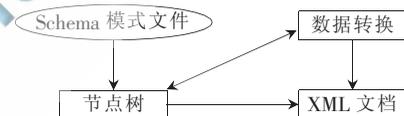


图 3 基于 Schema 模式的数据转换架构图

4 基于 Schema 模式的 EPG XML 文档转换实现

DMB EPG 系统需要生成三类 XML 文档,分别表示业务信息 SI、节目信息 PI 以及组信息 GI。规范这三类信息的 Schema 模式文件共有三个,分别是 epgSchedule_11.xsd、ServiceInformation_11.xsd 和 epgDataTypes.xsd。其中,epgSchedule_11.xsd 规范了 PI 和 GI 的树结构,ServiceInformation_11.xsd 规范了 SI 的树结构,而 epgDataTypes.xsd 定义了一些基本的数据类型。下面以生成 SI 的一个 XML 文档为例,描述业务信息的 XML 文档的转换过程。

首先,利用 XML Schema 解析工具(如 XMLBeans 等)解析 Schema 模式文件 ServiceInformation_11.xsd,得到一棵业务信息 SI 的 XML 文档框架树。根据这棵生成 Schema 模式树,建立如下的数据库表结构,建立表的过程的同时确定模式树与数据库数据的映射关系。图 4 中的每个表对应元素节点树的每个复杂节点,表中每个除了主键

