

# XBRL应用中的核心技术

史振生 张爱芹

## 一、XBRL应用流程及相关技术分析

XBRL(eXtensible Business Reporting Language,可扩展商业报告语言)是基于元标记语言XML(eXtensible Markup Language,可扩展标记语言)技术建立起来的一套会计信息标准的集合,其应用的典型流程如图所示。

### 1. 制定XBRL分类标准

XBRL分类标准定义了财务报告中各要素的属性及其相互之间的关系,相当于一本会计要素“词典”,制定XBRL分类标准是XBRL应用的第一步。

在XBRL分类标准的制定中,需严格遵循XBRL国际组织颁布的XBRL技术规范编制。不过,XBRL不是一门“元”技术,而是由XML技术衍生而来,是XML在业务报告领域的扩展。XBRL包含了XML的核心内容,并引入了XLink和一些其它规范的定义。因此,XBRL应用的核心技术是XML。

2. 数据采集(录入、导入等)、验证,生成实例文档

将来自不同数据源的不同数据格式的数据,按照XBRL分类标准,映射

为XBRL格式的数据,并经过数据验证,生成符合分类标准的实例文档,是XBRL应用的关键一环。在这一过程中,除了使用文本编辑软件,手工编辑生成XBRL实例文档外,实例文档的生成主要利用XBRL实例文档生成工具,一般有三种方式:

一是采用GUI(图形用户方式)。开发一个图形用户界面编辑器,由用户将数据手工录入,再由编辑器按照XBRL分类标准和技术规范,加工生成XBRL实例文档。

二是开发数据转换接口。针对不同的数据源,依据XBRL分类标准,在源数据和XBRL实例文档间建立映射关系,通过该映射,程序自动将不符合XBRL分类标准的源数据转换为符合XBRL分类标准的目标数据(实例文档)。

三是改造原有系统,增加能够输出符合XBRL分类标准实例文档的功能。比如在现有会计软件、ERP软件和企业管理软件中,建立原系统数据与XBRL分类模式间的映射,自动生成XBRL实例文档。

### 3. 实例文档的传输

XBRL实例文档在Internet上,以Web服务的形式进行传输,XBRL实例文档传输过程中涉及的信息技术主要是网络技术,网络安全性问题会贯穿于XBRL应用的全过程。

### 4. 实例文档的存储

XBRL实例文档的存储策略是影响XBRL应用效率的一个重要方面。XBRL数据存储主要有:文件系统、关系型数据库、原生XML数据库三种方式。

### 5. 数据挖掘、分析

随着XBRL应用范围的扩大,XBRL数据量将急剧增长,如何在海量的XBRL数据中找到需要的信息,是XBRL应用进一步深入和发展的重要课题。XBRL数据挖掘和分析,将涉及数据库技术及数据挖掘技术。

### 6. 数据展示、发布

XBRL数据的展示和发布是提高XBRL数据有用性的重要一环。XBRL数据不仅用于信息交换,而且可以通过Web技术,将数据发布和展示到网站上,使得互联网用户可以通过链接和搜索引擎找到所需的信息。

## 二、XBRL的技术基础——XML技术

XBRL并不是一种原创技术,它是XML在财务报告信息交换方面的一种应用,即XBRL是由XML衍生出来的标记语言,XBRL技术的核心是XML。XBRL主要应用了XML技术和XLink技术。

### (一) XML

#### 1. XML的由来

XML是由万维网联盟(World Wide Web Consortium, W3C)于1998年2月发布的一种标准。它是SGML(Standard Generalized Markup Language,标准通用标记



图 XBRL应用的典型流程

语言)的一个简化子集,将SGML的丰富功能与HTML的易用性结合到Web的应用中,以一种开放的自我描述方式定义了数据结构,在描述数据内容的同时能突出对结构的描述,从而体现出数据之间的关系。这样所组织的数据对于应用程序和用户都是友好的、可理解的、可操作的。

## 2. XML的特点

(1) XML是自描述的。XML的优势之一是它允许各个组织、个人建立适合自己需要的标记集合来说明文档的内容。

(2) XML是一种元语言,允许开发各种不同专业的特定领域的标记语言。如XBRL即是XML在财务报告领域的一种应用;XML用于数学领域,产生了MathML标记语言;XML应用于化学领域,开发出了化学领域的CML标记语言;XML应用于电子商务领域,在汲取了EDI(电子数据交换)实践经验的基础上,产生了ebXML(基于XML的全球电子商务标准)等。

(3) XML支持对文档内容的验证。XML文档的结构和内容必须符合一定的文档规则,要遵守基本的XML语法,另外,还要受到一定模式的约束,以方便地验证文档的有效性。

(4) 支持互联网上的快速搜索。由于XML文档的结构和含义可以通过模式文件获得,因此很容易在XML文档中进行搜索。在网络上,如果网页是XML格式的,则搜索会更高效。而且,不仅可以搜索数据,还可以在搜索中加入与数据相关的上下文信息,这样就形成了更精确的搜索机制。

## 3. XML的应用

(1) 商务的自动化处理。XML丰富的标记可以描述不同类型的单据,例如信用证、保险单、索赔单以及各种发票等。结构化的XML文档发送至Web的数据可以被加密,并且很

容易附加上数字签名。ebXML(基于XML的全球电子商务标准)即是XML应用于电子商务领域的成果,且XML有希望推动EDI(Electronic Data Interchange,电子数据交换)技术在电子商务领域的大规模应用。

(2) 信息发布。信息发布在企业的竞争发展中起着重要作用。使用XML技术,服务器只需发出一份XML文件,客户可根据自己的需求选择和制作不同的应用程序处理数据。加之XSL(eXtensible Stylesheet Language,扩展样式表语言)的帮助,使广泛的、通用的分布式计算成为可能。

(3) 智能化的Web应用程序和数据集成。XML能够更准确地表达信息的真实内容,其严格的语法降低了应用程序的负担,也使智能工具的开发更为便捷。来自不同应用程序的数据也能够转化到XML这个统一的框架中,进行交互、转化和进一步的加工。

### (二) Xlink技术

Xlink是XML Link Language的简称,它定义了一个在XML文件中创建超链接的标准方式。Xlink用于创建XML文件中的超链接,XML文件中的任何元素都可以以Xlink的形式表现出来。它支持简单链接和扩展链接,且提供多目标、双向链接等。

## 三、XBRL应用中的数据库技术

XBRL数据的存储离不开数据库技术。目前,XBRL应用采用的存储方式主要可以分为文件存储方式、数关系型数据库存储方式和原生XML数据库存储方式。

### (一) 文件存储方式

XBRL一般以实例文档的文本文件的形式直接存储在计算机的文件系统中。XBRL实例文档即是XML文件,也是自描述文档,这种文件完全可以数据存储方式来存储数据。但是,这种“简易”存储方式,由于缺少数据库管

理系统的高效存储组织、索引、事务处理、触发器、数据完整性和多用户管理等功能,在对大量XBRL数据进行处理时,其效率和安全性是比较差的,无法满足业务高性能存储的要求。

### (二) 关系型数据库存储方式

关系型数据库是当今应用最广的一类数据库,其技术也是最成熟的。基于关系数据库的XML数据存储,将XML数据分解到若干关系表中存储,充分利用关系型数据库的已有资源和成熟技术来管理XML数据。在关系数据库模式下,XML查询操作转化为一系列关系查询操作,利用关系查询处理和优化机制来响应查询。不过,由于XML数据大都是半结构化或者非结构化的数据,而关系数据都是结构化数据,两者之间存在本质上的差异,因而将XML数据分解到关系存储需要经过映射,这种映射有时会导致一些信息的丢失。

### (三) 原生XML数据库存储方式

一般认为,满足以下三个条件的XML数据库才能称之为原生数据库:1.为XML文档定义了一个逻辑模型,XML数据的存储和查询都是基于这个模型。2.将XML文档作为逻辑存储的基本单位,就像关系数据库将元组作为它存储的基本单位一样。3.不要求只能使用某一特定的底层物理模型或某种专有的存储格式。

这一定义表明,原生XML数据库的核心在于模式,其与支持XML的关系数据库的本质区别是逻辑模式不同。不过需要强调的是,支持XML的关系数据库与原生XML数据库之间的界限已越来越模糊。很多传统的关系数据库都增加了原生支持XML的功能,而原生XML数据库也增加了对文档存储在外部关系型数据库的支持。■

(作者单位:中国银行总行财务管理部 河北经贸大学会计学院)

责任编辑 刘忻