

文章编号: 1007- 2985(2004) 01- 0052- 04

基于 XML 的分布式试题库系统的设计与实现

何昭青^{1,2}

(1. 国防科技大学计算机学院, 湖南 长沙 410073; 2. 邵阳学院数学系, 湖南 邵阳 42200)

摘要: 使用 Internet 高度分布式计算环境全新计算平台 .NET Framework 及 XML 语言构建并实现一个分布式试题库系统. 该系统有效地实现了分布式、异构题库资源的共享.

关键词: 分布式数据库; 远程教育; 试题库系统

中图分类号: TP311. 13

文献标识码: A

Internet 正在改变着人们的学习方式, 它已经成为人们获取信息、解决问题的重要途径. 由于传统的试题库系统中的数据信息一般是以数据表的形式存储在一些异构的关系数据库中, 使得它们各自为政难于实现共享. 即使在同一类数据库中存储的试题, 试题建模和构造方法也缺乏统一的格式标准, 相互之间访问也十分困难, 造成了大量小而散的题库重复建设和资源的冗余.

随着互联网络技术、分布式技术的迅速发展, 基于计算机网络实现无限共享(至少在理论上如此)为目标的分布式试题库系统正在兴起. 其关键技术是对这些分布、异构的题库资源进行标准化和优化处理. 这就需要一种语言, 能够将这些异构资源数据转化为统一的标准. 可扩展标示语言(XML, Extensible Marked Language)能胜任这一任务. 它具有简洁、高效、可扩充、跨平台的特点. 利用 XML 作为结构标示、传输及转化媒介, 可以极大地实现网络教育资源的共享.

笔者介绍的分布式题库系统是结合相关项目——军队现代远程教育项目而设计、开发的一个系统. 该系统充分利用了 XML 标准统一、简洁、自描述和易扩充的特征, 采用了基于 XML 的试题库建模与构造方法, 使应用程序能够高效访问异构系统的数据资源, 从而实现分布式题库资源的共享.

1 系统实现的关键技术

1.1 XML 技术

XML 是 Extensible Markup Language(可扩展标示语言)的缩写, 是 W3C 组织于 1998 年 2 月发布的标准. 它的 2 个先驱是 SGML 和 HTML, 因此, XML 既具有 SGML 的强大功能和可扩展性, 同时又具有 HTML 的简单性. XML 语言不仅能够表示文档内容, 而且可以表示文档的结构. XML 建立在容器模型的基础上, XML 文档是以说明语句开始, 后跟 XML 文档的 XML 元素, 每个 XML 文档必须有 1 个根元素, 这是一个包括文档中其它所有元素的元素. 一个 XML 元素是由开始标记、结束标记以及标记之间的数据构成的. 开始标记和结束标记用来描述标记之间的数据. 标记之间的数据被认为是元素的值. 在 XML 中规定: 所有元素必须有结束标记; 所有元素必须正确地嵌套, 不允许交叠元素; 所有特征值必须加引号.

同时, XML 还继承了 SGML 的大部分优点, 而且更加规范, 更易于使用. 可以说, XML 是一种新的协

收稿日期: 2003- 12- 12

作者简介: 何昭青(1964-), 女, 湖南省邵阳人, 国防科技大学计算机学院硕士研究生, 邵阳学院副教授, 主要从事现代远程教育和网络研究.

议,也是一种定义标记语言的元语言,还是一系列技术的结合. XML 主要有如下突出特点:(1)良好的可扩展性. XML 允许各个不同的行业根据自己独特的需要制定自己的一套标记,同时,它并不要求所有浏览器都能处理这成千上万的标记,同样也不要求一个置标语言能够适合各个行业各个领域的应用.它允许开发者创建他们自己的数据规则,有效地创建可被用于多种应用的 可扩展的 标志集.使用几个附加的标准可以向核心的 XML 功能集增加样式、链接、和参照能力.(2)内容与形式的分离. XML 中信息的显示方式已经从信息本身中抽取出来,放在了 样式单 中,这样就使得 XML 具有良好的自描述性,能够描述信息本身的含义甚至它们之间的关系.(3)开放性和跨平台性. XML 有着很强的 表述 信息的能力,并且以文本格式存储和传输,便于用作各种不同系统之间的交流媒介,是一种非常理想的 网际 语言.(4)互操作性. XML 可以在多种平台上使用,而且可以用多种工具进行解释. XML 支持用于字符编码的许多主要标准,允许在各种操作系统平台上使用.

本系统采用了 XML Schema 语言进行题库建模. W3C 认定的文档验证模式有 2 种: DTD(Document Type Description) 和 XML Schema. 其中 XML Schema 是 2001 年 5 月正式发布的 W3C(万维网联盟)的推荐标准,成为全球公认的 XML 环境下首选的数据建模工具. XML Schema 完全使用 XML 作为描述手段,具有很强的描述能力、扩展能力和处理维护能力,而且具有一致性、扩展性、易用性、规范性及互换性等出众的性能.

1.2 .NET 软件框架

.NET Framework 是基于 Internet 高度分布式计算环境并以简化应用程序开发为目的的全新计算平台. .NET Framework 提供了稳定、安全的跨程序语言和平台的通用语言执行环境以及应用程序可以调用的类函数库. 其目的是让 Web 开发应用程序与 Web 服务的建立更为简单,使 Internet 上的各应用程序之间可以使用 Web Server 进行沟通.

1.3 ASP.NET 技术

Microsoft 新近推出的 ASP.NET 不是对 ASP 版本上的更新,而是一门全新的技术,是对 ASP 如何提供动态 Web 开发环境跳跃性发展. ASP.NET 几乎完全基于组件和模块化,用户使用的每个页面、部件和 HTML 单元都是一个运行时间的组件对象. ASP.NET 提供了许多基础服务以协助程序员建置互连网应用程序. ASP.NET 是建置、管理、部署 Web 服务的最佳平台,是 .NET 的基础. 在 Visual Studio.NET 上可以开发 ASP.NET 应用程序.

1.4 ADO.NET 数据访问

ADO.NET 是一个以 .NET Framework 为基础的全新数据操作模型. 它是专门为 .NET 平台上的数据存取操作而建置的.

ADO.NET 是一种与数据来源无关的新一代的数据访问技术. ADO.NET 技术是一个基于标准的、面向创建分布式资料共享应用程序的编程模型,是 ADO(ActiveX Data Objects) 技术的延伸和发展. 它带来了比 ADO 技术更为优越的互操作性、可维护性、可编程性和更出色的性能. ADO.NET 的主要对象有 Connection, Command, DataReader, DataSetCommand 及 DataSet 对象. DataReader 提供程序读取数据的接口; DataSetCommand 执行 SQL 指令、开启表; DataSet 提供给程序存取数据的接口.

ADO.NET 中的许多对象是从 ADO 技术中进化而来的,例如 Connection 和 Command; 也有许多对象是全新的,例如资料阅读器(DataReader)、数据集(DataSet)、数据视图(DataView)和数据适(DataAdapter)等等. 为了将数据访问和数据操作分离开来,通过使用这些类来获得数据访问服务; 因此,对于基于 XML 的数据源,采用 .NET 的 ADO.NET 的数据访问技术设计 Web 应用程序,可以实现对 XML 数据源的高效访问.

2 系统的体系结构

本系统采用基于 XML 客户和服务器的(C/S)交互模式分布式试题库系统的体系结构,实现最广泛的数据共享. 和传统的 C/S 交互不同的是,基于 XML 的 C/S 交互模式为客户自定义试题筛选方案提供了极大的自由度,可以真正实现个性化和智能化的组卷. 分布式试题库系统的结构包括以下 3 部分: 分布式存储的远程数据层; 以数据集为核心的远程业务层; 本地的业务层和数据层以及可以实现交互式信息传递的客户端. 其逻辑结构如图 1 所示.

2.1 分布式存储的远程数据层子系统

分布式存储的远程数据层子系统,由在不同的节点服务器上分散存储的数据源组成.存储媒介可以是具有 XML 功能的数据库和零散的 XML 格式的试题文档.具有 XML 功能的数据库包括 XML Enabled Database(支持 XML 的数据库,简称为 XEDB)和 Native XML Database(纯 XML 数据库,简称为 NXDB).前者是在传统的关系数据库、面向对象数据库或关系对象数据库中增加了对 XML 的支持,数据库内部仍然采用原有的表格格式存储,当用户进行 XML 查询或其它处理时,数据库将这些数据重新组合并转化为 XML 格式输出;后者则不同,它存储 1 个 XML 文件的时候,会创建 1 个基于 XML 的模型,其中包括多层嵌套,完整地保存 XML 的分层结构.

2.2 远程业务层和本地业务层子系统

该系统由 2 层组成 远程业务层和本地业务层.远程业务层是在远程服务器上实现与数据库交互的 Web 应用程序,相对于数据库又称为中间件,它通过一定的接口技术(如 ODBC, JDBC 和 OLE DB 等)实现对数据库访问(如选取、插入、删除和更新等)的操作.此外,远程业务层还负责各个节点服务器之间的通信以及向各自的客户端提供信息和服务等任务.本地业务层是试题库系统应用程序存放的服务器,其功能面向 2 个方面:一方面作为本地服务器面向客户端提供应用服务;另一方面,作为节点网络服务器,实现与其它节点服务器的信息交流与资源共享.

2.3 本地数据层及透明化的客户端

本地数据层除了用来存放本地试题库的信息外,还存放那些从远程数据库中获得的长期有用的试题资料(虽然未必马上用得上).本地数据库的内容能够通过一定的算法实现智能化动态更新,存放本地大多数用户经常访问的数据.这样,在大多数情况下就无需再分别访问各个网络服务器节点来查找试题,只需访问本地数据库就可以组建 1 份试卷,只有那些本地数据库中不存在的试题才通过节点服务器来查找,节省了系统的开销.而所有这一切都被本地服务器屏蔽掉了,客户在测试的时候并不知道哪些试题来自本地服务器的,哪些试题来自远程节点服务器.

3 系统的实现机制

本系统使用的是基于 XML 的 CS 数据交互模式,它是 CS 间通信的一种高效的工作方式.其实现机制为:首先在客户机对 XML 数据打包,然后以 XML 数据包为单元发送给服务器.服务器在接收到客户机发送过来的数据包后,马上处理数据,并做出相应的响应:服务器先创建 1 个 XMLDOM 对象,然后将 Request 对象中的数据载入,并通过 XMLDOM 对象开始访问 XML 数据.例如,在获得 XML 数据的访问入口后,首先做的事情是对 XML 文档进行验证,一旦通过验证,就可以利用 DOM 接口对 XML 数据进行分析处理(如将提取的信息来更新数据库中的记录).服务器在处理完数据之后,构造响应消息并返回消息给客户机.消息的形式可以是纯文本、HTML 页面、XML 文档或嵌有 XML 数据岛的 HTML 页面.客户机接收到消息后执行其它操作,从而结束 CS 间的 1 次通信周期.下文是具体实现步骤.

(1) 客户机构造 1 个 XMLDOM 对象,作为发送 XML 数据的载体.XMLDOM 作为数据包的载体其数据来源可以是任何 XML 文档或是 XML 文档的片段(如 XML 数据岛),甚至可以是使用 loadXML 方法接收用

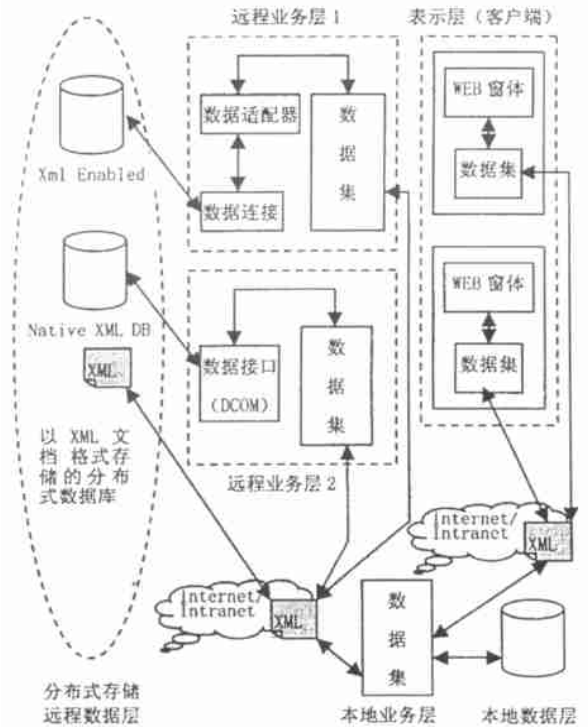


图 1 基于 XML 的分布式数据库系统的分层逻辑结构模

户输入信息后动态生成的 XML 文档. 部分语句如下:

Set docSubmit= CreateObject(Microsoft.XMLDOM) 创建一个 XMLDOM 对象.

docSubmit. async= False 设置同步通讯.

docSubmit. loadXML 用户输入的信息.

(2) 客户机创建一个 XMLHTTP 对象, 该对象包括多种方法和属性, 可以发送 XML 数据到服务器上的应用程序(如 ASPX 页面), 同时准备接收响应信息:

Set poster= CreateObject(Microsoft.XMLHTTP) 创建 XMLHTTP 对象.

对象创建后调用 Open 方法对 Request 对象进行初始化, 语法格式为:

poster. open http- method, url, async, userID, password.

(3) 客户机将 XML 数据包转载到 XMLHTTP 对象上并发送给 ASPX 页面:

poster. send XML- data 调用 Send 方法发送 XML 数据.

(4) 服务器执行 ASP.NET, 并创建 1 个服务器端 XMLDOM 对象来接收 XML 数据:

displayArea. innerHTML = poster. responseText 接收 Resoponse 消息.

(5) ASP.NET 把数据包转载到服务器端的 XMLDOM 对象上. 程序略.

如果服务器采用 XML 文档作为返回客户机的消息, 客户机就可以利用智能程序对结构化的消息进行分析, 更准确地理解服务器所要表达的信息内容.

4 结语

基于 XML 的分布式题库系统涉及到 XML、数据库、数据建模、WEB 应用和教育考试理论等诸多领域, 是一项复杂的系统工程. 基于 XML 的分布式题库系统目前已在实验室运行, 由于系统使用 XML 作为整个分布式试题库信息流通的“世界通用”的语言, 因此, 它能兼容不同的数据库结构和平台, 能充分实现因特网上资源共享. 各模块功能明确, 调用关系科学, 系统整体效率也高. 由于 XML 的标准还在进一步完善当中, 尤其是 XML 在试题库中的应用尚处于起始阶段, 所以这一系统在目前情况下还有许多要解决的技术难题, 比如对 XML 数据的转化和访问的效率还不理想, 针对 XML 数据的算法优化还有待于进一步挖掘. 伴随着 XML 诱人的应用前景, 随着 XML 技术的进一步成熟, 基于 XML 的分布式的试题库系统必将取代现有的试题库系统.

参考文献:

- [1] ERIC von der Vlist. Using W3C XML Schema[EB/OL]. http: www. vschool. net. cn xml. 2001.
- [2] 佚名. 构建面向 CRM 的数据挖掘应用[M]. 北京: 人民邮电出版社, 2001.
- [3] 徐振航, 刘莉芹. XML 和数据库[EB/OL]. http: www. pbookvn. com. 2002.
- [4] 王凯明. 在一个 DataSet 中储存多个数据表[EB/OL]. http: stady. 51. net webnet aspnet. 2002.
- [5] 方翔, 袁国栋, 李伟生. 建立特殊索引实现 XML 文档的查询优化[J]. 计算机工程与发展, 2002, 28(3): 114-115.

The Design and Accomplishment of a Distributed Database System on Test Question Based on XML

HE Zhao-qing^{1,2}

(1. School of Computer Science, National University of Defense Technology, Changsha 410073, China; 2. Department of Mathematics, Shaoyang College, Shaoyang 422000, Hunan China)

Abstract: Through the completely new computer platform under the highly distributed computer circumstances of Internet Framework and XML, a system of distributed test questions database is constructed, and the system make it come true to share the distributed, different Database on test questions.

Key words: distributed database; long-distance education; test question databased system