网上图书馆论文系统的设计与实现

Library paper system on net design and realization

高　龙　（北京信息工程学院　100101）

摘要：本文介绍针对图书馆的论文管理和服务的需要，采用标准 J2EE 平台，使用 XML 作为数据交换格式，采用 SQL Server 2000 作为后台数据库，来完成一个具有论文的采编、管理、检索、查阅等功能的网上图书馆论文系统的设计与实现。

关键词：J2EE　XML　全文检索

1　引言

随着图书馆领域的信息化普及，从而为图书馆系统提供了功能强大的信息处理平台，使得创建一套信息采集、信息加工、信息传输与信息存储高度自动化、共享化的信息管理系统成为必需了。网上图书馆论文系统就是根据现在图书馆论文管理和服务的需要来设计和实现的。

2　系统的分析与设计

网上图书馆论文系统负责论文的存储、管理、查阅等工作，面向管理员提供论文的采编、管理等功能，面向用户提供论文的检索、阅读等服务。整个系统包含用户管理，论文采编，论文检索，论文查阅等几部分。

本系统的设计是基于 J2EE 平台的。EJB 作为 J2EE 平台的核心组件之一，EJB 规范定义了一个可重用的组件框架来实现分布式、面向对象的商业逻辑，因此采用 EJB 来构建系统统一的业务逻辑的应用平台。XML 具有数据描述的特点，格式良好，与平台无关，便于在系统之间交换等特点，在本系统中采用 XML 作为基本的数据交换格式，例如论文的元数据信息就是用 XML 文档格式来进行数据交换。这样不仅使系统对外提供各种方式和平台的应用，而且使系统具有良好的可拓展性和互操作性，为进一步的分布式应用提供基础。

系统的后台数据库采用 SQL Server 2000，它不仅提供各种强大的数据处理功能，而且集成很多相关的服务。全文检索就是 SQL Server 2000 集成的服务之一。在系统设计过程中，考虑到开发一个全文检索服务的复杂性和艰巨性，就选用 SQL Server 2000 本身自带的全文检索功能，为系统中的论文检索模块提供全文检索服务。

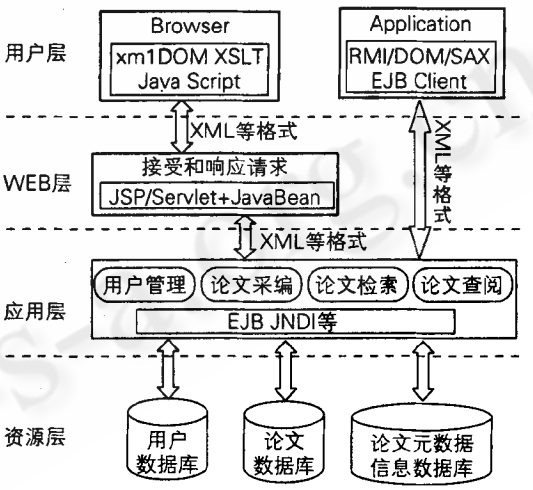

图 1　网上图书馆论文系统的系统框图

2.1　系统模块介绍

(1) 用户管理。用户管理提供个人信息的维护和身份验证的功能。

为了安全起见，用户注册的密码经 MD5 算法加密，最后以密文串的形式存入数据库中；在密码验证时，用 MD5 算法加密输入密码，把得到的密文串与数据库中对应的密文串对照，就可以实现间接的密码校验。这样做的好处是真正的密码只有注册者自己知

道,其他人(像管理者或黑客等)即使能从系统数据库中获取用户的信息,仍无法知道用户的密码。这样就从一定程度上保证了用户的安全。

(2) 论文采编。论文文件通常是 word,pdf 文档等,考虑到这些文档不是很大,为了便于管理和全文检索,就把它们存放到论文数据库中;同时出于论文检索、分类、管理等的需要,提取论文的元数据信息,把它们存入元数据信息数据库。图 2 就是采编模块在应用层上的流程图。

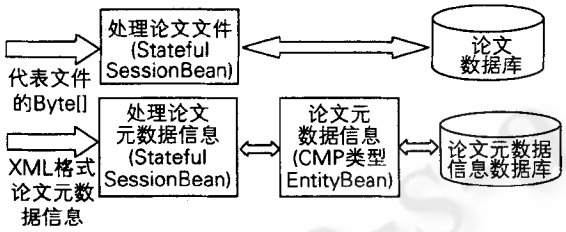

图 2　采编模块在应用层上的流程图

由于 EJB 不能访问文件系统,而且对外接口的参数的数据类型必须是串行化的,因此 EJB 不能以文件类型作为方法的参数类型来接受论文文件。只有在处理论文文件的 SessionBean 中采用以 byte[ ]作为远程方法的参数类型,来实现接受论文文件。当客户程序(JSP/Servlet 或应用程序)调用该远程方法,EJB 把接收的 byte[ ]用 FileInputStream 类的方法 read(fileByte,0,fileByte. length - 1) 和 DataOutputStream 类的 write(fileByte,0,fileByte. length - 1) 读到文件流 FileInputStream 中,然后用 PreparedStatement 的方法 setBinaryStream(1,fis,(int) fileByte. length - 1) 把论文文件存入数据库中。

(3) 论文检索。为了使用用户高效,合理的使用论文资源,因此论文检索是论文系统中很重要的一部分。该模块提供两种检索方式:

① 高级组合检索。通过论文的各元数据项的检索关键词的逻辑组合,对论文元数据信息数据库中的元数据信息进行匹配检索。

② 全文检索。通过检索关键词逻辑组合对论文数据库中的论文文件进行全文检索。

在高级检索中,因为涉及到多个论文元数据项的关键词的组合,出于灵活考虑,把这些关键词用自描述的 XML 格式的字符串来进行封装,当负责高级检索的 SessionBean 接受到该字符串后,用 Java DOM 来解析它,把这些关键信息构造为访问数据库的检索条件。在全文检索中,为了使检索智能化地支持逻辑组合检索,采用专门的程序(例如 JavaBean)来处理用户按不确定的逻辑组合规则把关键词组合成的字符串,按照自定义的一些逻辑组合规则,把这些关键词和逻辑组合关系构造为访问数据库的检索条件。

(4) 论文查阅。论文查询是向用户提供论文的元数据信息和论文文件本身。

在 CMP 类型的 EntityBean 中采用 EJB - QL 语句(例:SELECT OBJECT(h) FROM FileInfo AS h WHERE h. IDNumber = ?1),来实现查找论文的元数据信息,并把它封装到一个 CMP 类型的 EntityBean 中;SessionBean 通过引用该 EntityBean,通过调用它的方法来获得论文的元数据信息,并把这些元数据信息封装到 XML 文档格式的字符串,转发给请求元数据信息的程序。

对于论文文件,仍然是在 SessionBean 中从数据库中获得文件流,然后把它变成 byte[ ],对外部程序提供以返回值为 byte[ ]类型的方法。当外部程序获得 byte[ ],再根据文件名和格式类型,把 byte[ ]读到文件中去。

2.2　系统的访问方式

面向用户和管理员,用户层提供两种访问方式,一种是基于浏览器的,是属于 B/S 结构下的访问方式;另一种是面向应用程序的,是属于 C/S 结构下的访问方式。两种访问方式各有各自的特点和优势。

(1) 基于 B/S 结构下的访问方式。由于 B/S 结构在成本投入,数据安全性,网络限制等方面的巨大优势,B/S 结构已得到了广泛的应用。在 B/S 结构中,Web 层分担了部分的数据处理和提供一定安全性。在本系统中,Web 层不仅负责了密码加密,检索关键词的组合,出错处理等工作;而且维护着浏览器和 web 服务器之间的会话状态(通过 HTTPSession 来实现的),使系统获得了一定的安全性;还在浏览器中把论文上传至服务器的实现过程,Web 层在接受论文文件时用到第三方的文件上传组件—Fileupload(注:Apache 的开源项目),来实现接受批量的上传论文文件。

浏览器出于它的安全性和通用性的考虑，它在本地的处理能力，处理的针对性等方面还不够，因此在出于这些方面的考虑时，可使用 C/S 结构下应用程序来访问系统。

（2）基于 C/S 结构下的访问方式。鉴于 C/S 结构下的客户端应用程序具有很强的本地处理能力和应用处理的针对性强，以及本系统的需要，实现了 C/S 的访问方式。采用 C/S 结构后，原来在 Web 层和浏览器中的数据处理和实现都在客户端的应用程序中实现，加上应用程序本地处理的优势，不但提高了系统的处理效率，而且为客户端的处理进一步拓展提供了条件。

由于与 EJB 交换的数据大都是 XML 格式的，加上 XML 的自描述性，极大的方便了客户端应用程序的处理。客户端的应用程序与业务逻辑的应用平台的 EJB 通信是使用 RMI，RMI 在数据处理等方面有很大优势，但是它要求应用程序是基于 Java 平台的，更关键的是它不能透过防火墙的限制。这是系统在 C/S 结构下的缺陷之一。

2.3 数据库的全文搜索功能

全文检索功能在 SQL Server 7.0 中开始引入。全文检索的核心引擎建立在 Microsoft Search（MSSearch）技术上。SQL Server 2000 的全文检索功能包含两个基本的组件：全文索引器和四个 Transact－SQL 操作。全文索引器是用于创建和填充全文目录，而全文目录是保存在 SQL Server 数据库的外部，被 microsoft 搜索服务维护和管理。四个 Transact－SQL 操作的全文检索分别是 COTAINS，FREETEXT，CONTAINSTABLE，FREETEXTABLE。前两个用在 WHERE 语句中，后面两个是行集功能，用在 FROM 语句中，这些操作可进行查询全文目录，只不过后面两个还向 SQL Server 返回键值（key）和排位值（rank）。

设置全文检索的具本步骤如下（括号内为每步所调用的存储过程名称）：

（1）启动数据库的全文处理功能（sp_fulltext_datebase）；

（2）建立全文目录（sp_fulltext_catalog）；

（3）在全文目录中注册需要全文索引的表（sp_fulltext_table）；

（4）指出表中需要全文检索的列名（sp_fulltext_column）

（5）为表创建全文索引（sp_fulltext_table）；

（6）填充全文索引（sp_fulltext_catalog）。

考虑到系统的效率，本系统采用增量填充。

2.4 系统性能优化

本系统的系统性能是一个需要考虑的问题，本系统在应用层和数据库方面做了一些性能优化工作。

在应用层的 EJB 设计时，合理的使用各种类型的 EJB，比如使用 CMP 类型的 EntityBean，可以允许容器进行优化；在方法设计上遵循粗粒度原则，把多个方法合并，减少方法调用次数，来提高系统性能。

在数据库的访问方面，通过设置数据库连接池，数据源（DataSource）等来提高访问效率。在全文检索方面，考虑到数据库填充全文索引要花费很大的系统开销，安排填充时间在数据库处于非高峰活动期间。

3 结束语

本系统基于标准 J2EE 平台，采用 XML 相关技术，后台的数据库采用 SQL Server 2000，较好的实现了一个网上图书馆论文系统，并提供较为强大的基本服务功能，还为进一步拓展提供了基础。系统还有些方面有待进一步改进：SQL Server 数据库提供的全文检索的功能有限，可以进一步的考虑专业的全文检索产品；提供深层的，增值的信息服务等。网上图书馆论文系统作为图书馆的子系统之一，负责论文的管理，检索等工作，具有相当不错的实用价值。

参考文献

1 Rod Johnson 著，魏海萍等译，J2EE 设计开发编程指南，电子工业出版社，2003.7。

2 Brett McLaugblin 著，刘基诚译，Java 与 XML（第二版），中国电力出版社，2004.2。

3 许苓、陈康，SQL Server2000 中的全文检索，程序员，P98，2005.2。

4 李晨阳、焦海星，创建高性能的 J2EE 应用系统，计算机系统应用，P10，2005.2。