

数字化校园网信息资源组织模型与实现

Organization Model and Implement on Information Resource for Campus-Network

李 晓 李立新 邱玉辉

(西南师范大学计算机与信息科学学院 重庆400715)

Abstract It is key for successful digital-campus to develop and make full use of the information resource. The approaches to information resource organization directly work on information service supported. This paper, which is based on study of information resource organization of the existing campus-network, presents a new model (Xcontent) based on content exchange for "User Demands First" and proposes an idea about how to implement based on J2EE and distributed database. This method has been applied successfully to the construction of the website of Southwest China Normal University.

Keywords Campus-network, Information resource, Organization, Model, Implement

1. 问题引入

Internet 的出现和发展,使得人类全部的信息资源以前所未有的程度和方式在全球内相互联通。信息资源即为在 Internet 上公开发布的网页和在线数据库的总和^[1]。在数字化校园实施过程中,如何开发利用信息资源,满足用户对信息的个性化需求,是信息化建设取得实效的关键。

基于 Internet 的不同应用系统有不同的信息资源组织方式^[2],如提供产品支持的站点按面向任务的模式组织资源,在线购物系统则按产品目录分类。信息资源组织方式与应用系统提供的信息服务直接相关,也对 Web 站点的主页和一级页面内容设计产生影响。

数字化校园网信息资源的组织,即是基于“用户信息需求第一”观点,对信息资源进行统筹,系统地分析、设计和实现,以能快速地把内容提供给不同用户,并在用户界面级提供个性化的服务。然而,现有的数字化校园网站在信息资源的组织方式存在着不足,无法满足“用户信息需求第一”,这主要表现在:

第一,目前网站大多以面向发布的方式组织信息资源,基本以 html 形式响应用户请求,但 html 文档天生的缺陷——只有显示标记而无数据组织关系的描述,使得信息交换实现困难。

第二,校园网站信息资源组织多以主题频道和内容的的方式进行,如系科简介频道,BBS 频道等,各频道各自管理,各主题也仅是简单的集成。这种组织方式存在着投资重复,网站更新、维护不便,资源浪费等问题。

第三,信息安全难以得到保证。不同用户对信息拥有不同的访问权限,这在目前的信息资源组织方式中实现较困难。

第四,信息的搜索在 html 或半结构化数据中进行,查准率和查全率不高。

最后,也是最重要的一点,难于全面支持用户对信息的个性化需求。尽管有的网站也区分不同的用户群^[3,4],但因其粒度过大,不能算是真正意义上的个性化支持。

数字化校园网建设的最终目的是实现信息内容交换,以

能快速地把内容提供给不同用户,使得用户无论是在内容方面,还是在用户界面的呈现方面,都能满足其个性化的需要,即“用户信息需求第一”。为此,本文围绕如何开发利用信息资源,提出了基于内容交换的、新的校园网信息资源组织模型(Xcontent),并给出了基于 J2EE 平台和分布式数据库的实现方案,以高效合理地利用信息资源,为用户提供个性化服务。

2. 信息资源组织模型

2.1 目前信息资源的组织和集成

现有校园网站信息资源的组织模式如图1所示。网站由多个主题频道组成,各主题频道围绕一个特定主题,如系科简介、BBS 论坛,进行组织。特定主题通常下设多级子主题,子主题又可由多个不同主题构成,一个简单的子主题(即有且仅有一个简单主题)则由一个动态或静态的网页构成。

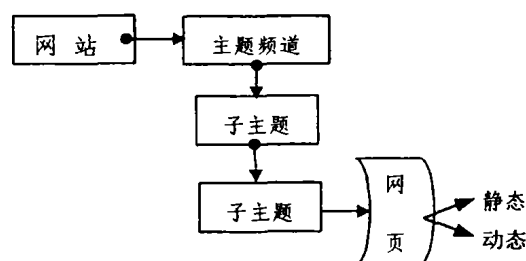


图1 现有校园网站信息组织

这种组织方式的特点是:第一,有多个组成信息资源的构成元素,如主题频道元素,内容元素;第二,各元素不同构,实现复杂;第三,各主题之间没有适当的集成方法,只是简单将它们放在一起,以“千人一面”的方式为用户提供信息,不能满足不同用户对信息的不同需求。

2.2 新的信息资源的组织模型——Xcontent

Xcontent 是我们提出的基于内容交换的、新的信息资源组织模型,其结构是面向对象的和层次的。

Xcontent 采用信息资源元作为构成信息资源的唯一元素,以 XML^[5]形式组织信息内容。这种模型解决了同构和集

李 晓 讲师,博士研究生,主要研究方向为网络应用、人机界面,分布式人工智能。李立新 副教授,博士研究生,主要研究方向为电子商务,分布式人工智能。邱玉辉 教授,博士生导师,主要研究方向为网络应用、电子商务、机器学习、分布式人工智能等。

成问题,从根本上解决了开发过程中的重复投入等问题,易于扩展。此外,也易于与其他资源(如用户)进行集成;一方面,通过引入用户权限,对信息进行不同程度的屏蔽,保证了信息资源只能由授权用户访问,这在数字化办公系统中是很重要的。另一方面,可根据用户信息偏爱模型,基于XML的特点,采用类似XSL的方式来解析XML,实现用户界面的个性化。

下面给出本文对Xcontent的相关定义。

定义1 信息资源元 info_unit 是组成信息资源的基本的、最小的元素,具有如下属性:

$\text{info_unit} ::= \langle \text{id}, \text{name}, \text{author}, \text{xml}, \text{xslUrl}, \text{dbServer}, \text{tblName}, \text{X_bkt} \rangle$

其中: id--信息资源元标识; name--该信息资源元名称; author--创建该信息资源元的创建者姓名; xml--该信息资源元内容的XML表示; xslUrl--解析xml文件所在URL; dbServer--该信息资源元所在数据库服务器; tblName--该信息资源元的数据表名称; X_bkt--信息资源元属性扩展基,通过增加“名-值”对实现。属性扩展基的存在保证了信息资源元定义的灵活性。

定义2 信息资源组织模型 $Xcontent = (INFO, R)$ 。其中: $INFO = \{\text{info_unit} \mid \text{info_unit} \text{ 为信息资源元}\}$; 若 $n(\text{INFO}) = 0$, 则 $R = \emptyset$; 否则 $R = \{H_{INFO} \mid H_{INFO} \text{ 是 INFO 上某个二元关系}\}$ 。

H_{INFO} 为如下描述的二元关系:

i. 在 INFO 中存在唯一的被称为网站 webSite 的信息资源元,在 H_{INFO} 下无前驱;

ii. 若 $INFO - \{\text{webSite}\} \neq \emptyset$, 则存在 $INFO - \{\text{webSite}\}$ 的划分 $INFO^{(m)}$, ($m > 0$), 对任意 $j \neq k$ ($1 \leq j, k \leq m$), 有 $INFO^{(j)} \cap INFO^{(k)} = \emptyset$, 且对任意的 $1 \leq i \leq m$, 存在唯一的信息资源元 $\text{infounit}_i \in INFO^{(i)}$, 有 $\langle \text{webSite}, \text{infounit}_i \rangle \in H_{INFO}$;

iii. 对应于 $INFO - \{\text{webSite}\}$ 的划分, $H_{INFO} - \{\langle \text{webSite}, \text{infounit}_1 \rangle, \dots, \langle \text{webSite}, \text{infounit}_m \rangle\}$ 有唯一的划分 $H_{INFO}^{(m)}$ ($m > 0$), 对任意 $j \neq k$ ($1 \leq j, k \leq m$), 有 $H_{INFO}^{(j)} \cap H_{INFO}^{(k)} = \emptyset$, 且对任意的 $1 \leq i \leq m$, $H_{INFO}^{(i)}$ 是 $INFO^{(i)}$ 上的二元关系, ($INFO^{(i)}, \{H_{INFO}^{(i)}\}$) 是一个符合本定义的 Xcontent 的子结构。

3. Xcontent 的实现

考虑到应用数据的安全性、负载的平衡和系统的可移植性,采用了基于J2EE^[6]平台和分布式数据库^[7]的多层分布式结构来实现对信息的组织,如图2所示。这是因为一方面, J2EE 平台提供了多层的分布式应用模型、组件再用和一致化的安全模型;另一方面, J2EE 使得平台独立。

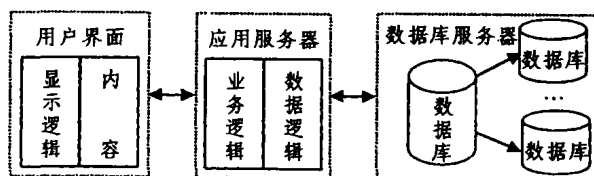


图2 Xcontent 的实现

在客户端用户界面,区分内容和显示逻辑。静态内容可从html或xml文件获得,动态内容则通过与应用服务器的业务逻辑交互实现;显示逻辑则是通过对获得信息资源元对象的xml属性的解析实现。显示逻辑和内容分开的优势在于,通过

定义Web模板,可快速建立相似的或个性化的网页。

应用服务器或者说中间层基于EJB^[8]实现,业务逻辑用一个JavaBeans类管理用户的输入输出,输入被发送到Enterprise Beans类进行处理。Enterprise Beans是一组可重用的组件,用于建立分布式应用程序。实现业务逻辑的组件封装了整个业务,通过调用数据逻辑,实现对数据库的访问。

对数据库的访问由数据逻辑负责。考虑到管理的方便性,设置了一个中心数据库来管理各分布式的数据库。多个分布式数据库的存在,一方面使得不同安全级别的信息可以分开存放;另一方面也避免了数据负荷过重。

4. 原型系统

本文提出的信息资源组织模型Xcontent和实现方案已成功地应用于我校校园网站建设,原型系统已在Linux服务器上顺利运行。

原型运行环境:服务器Linux环境支持,分布式数据库采用MySQL。

原型开发环境:应用服务器即中间层开发采用J2EE平台, EJB模型,处理业务逻辑和数据逻辑;客户端采用JSP^[9],处理内容的获取和显示逻辑。

图3是基于Xcontent的实现流程,而图4是Xcontent与用户资源集成的一个个性化界面,登录用户info是一个具数字化办公权限的用户。

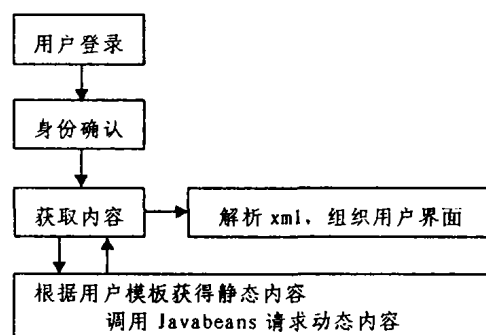


图3 基于Xcontent的实现流程

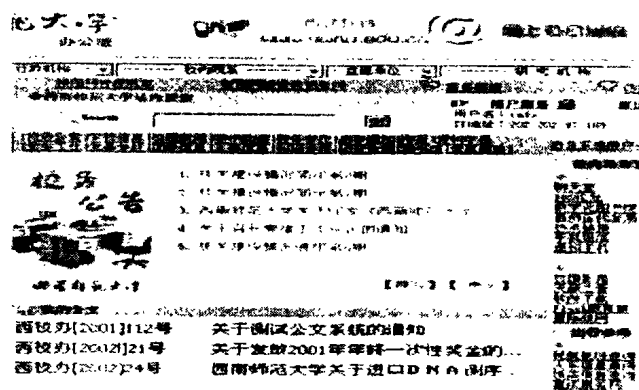


图4 Xcontent与用户资源集成结果

结束语 信息资源的开发利用在数字化校园建设中有着举足轻重的作用,信息资源的组织方式直接影响为用户提供信息服务。本文提出的基于内容交换的、新的信息资源组织模型(Xcontent),已被证明是有效的,易于与其他资源(如用户)进行集成。现在,我们正在完善该方案。

(下转第131页)

综合分析,我们不难得出结论,采用中间件建立的 Web 应用程序是我们在 Windows 平台下面最好的选择。因此,我们在业务层选择与采用的开发技术为:①ASP 作为服务器前台连接 COM 的脚本。②ADO 对数据提供程序提供了一个应用水平的界面。③OLE DB 对数据提供程序提供系统级的界面。④MTS 以达到缓冲池、实时作用、过程分离,提高组件应用效率以及事务处理特性。⑤COM 封装商业逻辑,增强处理效率、安全性和垂直应用能力有助于接口统一,提高开发效率。

4 COM 技术特征与实现

4.1 COM 的原理与技术特点

COM 是以 Microsoft 定义的组件体系结构标准,是一个基于对象的编程模型。它解决了基于组件系统的互操作性、语言独立性、透明性、版本等基本问题,定义了一个二进制标准。COM 建立了一套软件形态标准和接口标准,标准规定了组件如何为它的接口构建虚拟表以及用户,如何通过该虚拟表间接调用接口函数。传统的方法所开发的应用软件往往是一种独立的整体性系统,开发周期长、扩展功能差,应用系统严重依赖于操作系统和特定的网络服务,因而开放性很差。

COM 中定义了客户应用程序和服务对象的连接机制,连接建立后客户和对象可直接通讯而不需 COM 的介入。COM 组件服务器以进程中、本地、远程三种方式执行,用户不必知道自己正调用哪一个服务器。COM 的核心部件是接口,接口的定义和功能保持不变,COM 组件开发者可以改变接口功能,为对象增加新功能,用更好的对象来代替原有对象。对建立在组件基础上的应用程序几乎不用修改,大大提高了代码可重用性。软件组件与其他可复用软件模块的区别在于,它既能够在设计时使用或进行修改,也能够以二进制执行模块时使用或修改。一个以二进制形式实现的软件组件能够有效地嵌入其他开发的组件之中。与面向对象编程语言不同,组件技术独立于语言和面向应用程序,只规定组件的外在表现形式,而不关心其内部实现方法。软件开发技术组件化,使设计开发利用可复用的组件,组装构造新的应用软件系统,来实现高效通用、可扩展的软件结构。

4.2 COM 组件的工作过程

COM 是借助接口来输出它提供的服务,接口是 COM 对象提供服务的基本单元。因此一个 COM 对象中可以提供多个不同的接口,并且不同的接口提供了不同的方法和属性。COM 对象的接口都由 Class 类来定义。在 COM 中是使用 Windows API 的 CoGetObject 建立对象,只需要再输入机器名称的参数即可在远程指定的计算机中建立对象,并且取得指定接口的信息。

在 Windows 应用程序中建立 COM 对象,主要是利用 COM 执行时的函数库所提供的 API 来实现。在面向对象的编程语言中,编译器使用 Windows API 的 CoCreateInstance 来建立 COM 对象。事实上在一般的工具或中介服务器中(如:SQL Server 等),建立 COM 对象的标准方式是使用 Windows API 的 IClassFactory。其主要流程如图3所示。

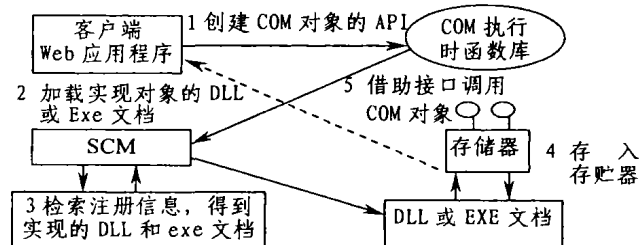


图3 调用 COM 的主要流程

由于 COM 必须注册到系统,让加载 COM 的管理器 SCM(Service Control Manager)找到它的注册信息才能确定调用一个 DLL 文档提供 COM 接口,因此注册组件是非常重要的。执行 COM 的两种方式中,其中一种就是 DLL 组件,称为 IN-Process-Server。这些组件加载到执行它的程序内存当中,其调用方法像调用一般的程序和函数一样非常快速。在执行的时候,应用程序通过 COM 对象公有 Iunknow 接口作为父接口,调用内存里面的 COM 对象接口。由于只需注重实现具体事务逻辑的代码编写而不需要过于关注底层实现的手段,因此,组件对象模型的这个特点大大地方便了开发人员,我们在实现系统的组件设计中加以广泛应用。

结束语 基于 DNA 的思想进行建模和构建电子商务平台,其健壮性、友好性非常突出。系统采用多层体系结构和 COM 组件技术,将事务逻辑处理层与表示层分开,具有较好的可伸缩性。客户端完全基于浏览器方式,界面友好、操作与维护方便。系统利用 Windows 2000 和 SQL Server 等多级权限控制,进一步确保系统的安全性。利用 MTS 进行组件管理,提高了系统性能。分布对象技术与其它技术在多层体系结构基础上的融合是其不断发展的方向。

参考文献

- 1 李世杰. DNA&Web 数据库应用与剖析. 北京:科学出版社,2000
- 2 潘爱民. COM 原理与应用[M]. 北京:清华大学出版社,2000
- 3 李维. Delphi5 ADO/MTS/COM 高级编程. 北京:机械工业出版社,2001
- 4 PaPa J, Shepker M, 等. SQL Server 7.0 Programming Unlesshed [M]. 前导工作室译,北京:机械工业出版社,2000
- 5 McGrath S. XML 应用实例. 北京:人民邮电出版社,2000

(上接第133页)

参考文献

- 1 中国互联网络信息资源数量调查报告. <http://www.cnnic.net.cn/>,2001
- 2 Dobroth K,McInerney P,Smith S. Organizing Web Site Information: Principles and Practical Experience, SIGCHI Bulletin,2000, 32(1):23~26
- 3 <http://www.fudan.edu.cn/>,2001(旧版)

- 4 <http://www.swnu.edu.cn/>,2002
- 5 www.w3.org/XML/,2001
- 6 Perrone P J,et al. J2EE 构建企业系统——专家级解决方案. 清华大学出版社,2001
- 7 Reese G. JDBC 与 JAVA 数据库编程(第二版). 中国电力出版社,2002. 03
- 8 EJB2.0. <http://java.sun.com/products/ejb/>, 2001
- 9 JSP Whitepaper. <http://java.sun.com/products/jsp/>, 2002