

基于元数据的目录服务体系研究与实现

卫佳蕴, 孙 莉, 朱吉翔

(东华大学 计算机科学与技术学院, 上海 200051)

摘 要:随着网络技术和计算机技术的迅猛发展,世界各国对于电子政务建设日趋重视,电子政务在国外蓬勃发展,而我国的电子政务起步比较晚,经济发展呈现不平衡态势。针对政务信息化建设中出现的“信息孤岛”等问题,在现有的政务信息资源注册管理模型基础上,提出了基于元数据的目录服务,分析研究了政务信息资源目录服务体系的框架。并提出了基于元数据的政务信息资源的目录服务的实现方法,最终在上海市水务局实现试点应用。

关键词:电子政务;元数据;信息资源;目录服务

中图分类号: TP311

文献标识码: A

文章编号: 1673-629X(2008)04-0042-03

Design and Implementation of Catalog Service System Based on Metadata

WEI Jia-yun, SUN Li, ZHU Ji-xiang

(School of Computer Science and Technology, Donghua University, Shanghai 200051, China)

Abstract: With the development of the network as well as the computer technology, the development of E-government has been emphasized lots by the nation all over the world. As the result, E-government developed quickly and prospectively abroad. However, the degree of the development in E-government in our nation seems stay in a primary phase especially comparing with the foreign countries while the governments in all levels have devoted greatly in both material and human resources. To solve the “isolated information island” problem, has pointed out the catalog service based on metadata. By analyzing the frame of the catalog service, mentioned the structure of E-government information catalog service system. Then the situation of such implementation in Shanghai Water Bureau was introduced.

Key words: E-government; metadata; information resource; catalog service

0 引言

目前,电子政务在国外蓬勃发展。然而,由于我国电子政务起步比较晚,经济发展呈现不平衡态势,虽然各地政府加大对于政务系统的开发投入,然而,和国外相比,各地政务系统还是处于闭门单干的状态,政务信息资源利用依然很困难,各地各部门间互联互通、信息共享、协同工作的水平还很低。

针对这一问题,提出了基于元数据^[1]的政务信息资源的开发、利用和交换方式。将政务信息提高到信息资源的高度,将元数据引入政务领域,从各政务部门按照一定的标准采集、抽取信息资源进行注册管理,并且通过目录服务体系与交换体系的方式加以利用,从而解决目前普遍存在于政务业务系统间的信息孤岛

问题。其中目录服务是整个系统中一个重要的组成部分。

文中在现有的政务信息资源注册管理模型^[2]基础上,提出了基于元数据的政务信息资源目录服务体系的框架并加以实现。

1 政务信息资源目录服务体系框架

目录服务从其功能上来讲应该具备资源的标识、检索、权限管理等功能,向社会公众提供公共资源信息,向政务部门提供共享利用的交换资源信息。目录服务系统建立在信息资源注册管理(系统)的基础上,是资源注册管理的应用和服务。

根据政务信息资源共享和应用的特点,需要把政务信息资源分为公共信息资源和政府内部交换信息资源:公共信息资源是指安全限制分级和用户分级最低的普通用户可以访问的目录中的公共信息资源,政务部门内部交换信息资源是指具有一定的安全限制的、政务部门用户才可以访问目录中的公共资源和内部交

收稿日期:2007-07-06

基金项目:上海市科研基金(03DZ05015)

作者简介:卫佳蕴(1983-),女,上海人,硕士研究生,研究领域为数据库技术、面向对象分析与设计;孙 莉,硕士生导师,副教授,研究领域为数据库技术、面向对象分析与设计。

换信息资源。当然对于政府部门用户也是需要进行分级控制,图1即是政务信息资源目录服务系统框架。

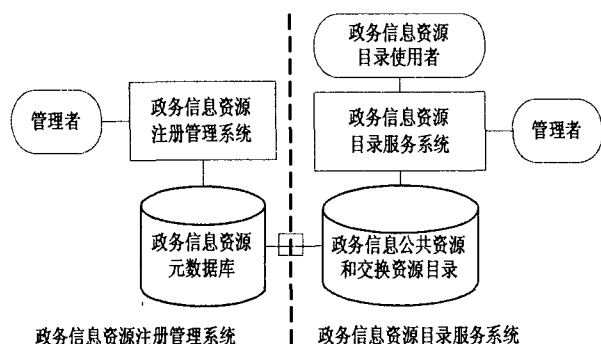


图1 政务信息资源目录服务系统框架

目录服务体系存在三种网络模型,分别为层次结构目录服务网络模型,分布式目录服务网络模型,分散集中式的目录服务网络模型。根据政务信息资源公开和政务部门之间松耦合的特点,文中提出的目录服务系统的设计和实现基于 Web 和 Web Service 技术。

资源目录服务可以由发现服务、访问服务和管理服务组成。

其中,发现服务是指用户通过政务信息资源目录服务系统发现和定位政务信息资源的元数据。发现服务也是目录服务系统必须实现的功能。

访问服务是提供客户获取政务信息资源的服务。访问服务可以分为二种类型:

(1)直接访问,提供客户端一个句柄,当这个句柄被客户端使用时,可以提供数据给客户端。

(2)代理访问,提供客户端订购数据的方法,以外部的其它方式提供数据访问。

需要指出的是,访问服务只限于定位链接,对链接的公共资源的下载由资源提供者支持实现,交换资源则由“交换中心体系”支持实现。具体的交换方式将在下文中进行阐述。

目录管理服务为目录系统提供维护和管理目录系统的功能。此功能可以通过系统的后台程序实现。

2 政务信息资源目录服务系统的实现

2.1 政务信息资源编目与目录条目格式

编目是将注册的政务信息资源元数据,按照资源分类索引和目录条目格式要求,形成资源记录信息导入到政务信息资源目录树中,构成可共享利用的政务信息资源目录。目录条目是供资源使用者发现、检索和开发利用资源而建立的,目录条目中不包括对资源注册和管理方面的信息。政务信息资源目录条目一般应包括资源标识和内容方面的描述元数据^[3,4]、资源获取和应用方面的元数据、资源安全和资源分类方面

的元数据等。如:资源目录标识、资源名称、资源出版日期、关键词、资源摘要、资源使用限制、资源访问中介、资源分类类目名称、资源类型、资源字符集、资源安全限制分级、资源安全保密期限、资源语种、在线资源链接地址等,具体应用在目录服务系统中的条目格式由目录服务系统管理机构根据实际应用的需求制定。

政务信息资源目录条目是供资源使用者发现、检索和开发利用资源的信息,目录条目中不包括对资源注册和管理方面的信息。通过参考目前流行的 DC^[5] 标准,本系统中的政务信息资源目录条目由以下数据项组成:资源目录标识、资源名称、资源替代名称、资源出版日期、资源更新日期、资源负责单位、关键词、资源摘要、资源使用限制、资源访问中介、资源分类类目名称、类目编码、分类标准、资源类型、资源字符集、资源安全限制分级、资源安全保密期限、资源语种、在线资源链接地址。

2.2 目录服务系统功能设计

目录服务系统的详细功能包括:目录发布,目录查询,资源定位链接,目录维护,日志分析,服务监控,用户反馈,权限管理等。

2.2.1 目录发布

目录发布是目录服务系统根据目录服务的要求,从政务信息资源元数据库中导出注册成功的政务信息资源目录信息,生成资源目录条目,并导入到政务信息资源目录库中,供用户访问。在本规范中规定目录发布的目录数据表示格式采用 XML 格式^[6]。

2.2.2 目录查询

目录查询是目录服务系统通过 HTTP 或者消息访问接口提供用户查询目录信息的功能。目录查询支持用户按照三种方式进行查询:

- * 按照资源分类编目进行浏览查询;
- * 按照单项条件进行查询,如:按资源名称、资源标识、资源发布时间、资源关键字、资源类型等;
- * 按照组合条件进行查询。

2.2.3 资源定位链接

目录服务系统提供的资源定位链接功能是指资源目录条目中包含了获取资源的信息,如果信息资源是在网上(公网或政务网)发布的信息,目录服务系统提供访问资源的 URL 链接。如果资源所属的部门还没有提供在线访问或交换服务的方式,用户可以通过目录条目中的资源负责单位信息和资源的使用限制信息,离线方式获取资源。

为了实现消息访问目录服务,系统部署了一个目录访问 Web 服务,作为对外的接口,接收和应答来自用户应用系统的消息,并且把用户的目录查询请求提

交给数据库访问中间件,得到查询结果后,Web 服务再以应答消息的形式把查询结果返回给用户系统。

通过第三方软件 TCP Monitor(TCP Monitor 在下文交换技术的实现中会有比较详细的介绍)可以截获通信的 SOAP 包,这里截取其中部分片段进行说明:

```
<egxml:mpdu>
<egxml:request type="4">
  <egxml:catalogId>byLaw</egxml:catalogId>
  <egxml:catalogSysId>Law</egxml:catalogSysId>
</egxml:request>
</egxml:mpdu>
```

其意义为删除指定目录(catalogSysId)下的目录结构(catalogId)

其中元素意义如表 1 所示

表 1 元素意义表

| 参数名称 | 参数含义 |
|--------------|---------------------|
| type | 用于指定进行目录删除操作 |
| catalogId | 用于指定更新目录的唯一标识 |
| catalogSysId | 用于指定更新目录所属目录系统的唯一标识 |

2.2.4 目录维护

目录服务系统还应该提供目录系统的维护功能,包括目录结构树的维护和目录条目信息的维护。当资源的分类体系发生变化时,目录服务系统要调整目录结构。当资源条目信息需要修改、删除、添加和更新时,目录服务系统应该支持管理员通过后台实现这些操作。除了对目录结构和内容的维护以外,目录系统还应该提供辅助的系统功能,主要包括:

- * 日志分析:根据元数据查询日志,统计访问中心网站的次数,统计不同政务信息资源元数据的查询次数;

- * 服务监控:监控目录服务器、网站的运行;

- * 用户反馈:管理用户的反馈意见,并和提供者进行协调;

- * 扩展目录服务系统的分类编目方案,根据不同的应用需求提供多个目录索引,并根据用户的特点提供个性化目录服务。

2.2.5 权限管理

权限管理是指给不同的用户不同的访问目录系统的权限,以保证用户可以正确地访问资源目录信息,并且保护目录系统目录信息的安全。根据用户的划分和目录服务系统安全管理的需求,把用户分为 3 类:第一类是公共用户,可以访问公共的信息资源目录信息;第二类是政务部门用户,可以访问公共的信息资源目录信息和有限的交换资源目录信息;第三类是目录服务

系统管理者,负责对目录服务系统的管理、目录信息的更新和维护等工作。用户在使用目录服务系统时,系统会根据用户的权限确定用户可以执行的操作和过滤,阻止用户的非法操作,并且会根据用户的权限和目录条目中资源保密属性过滤用户执行目录查询的结果。

2.3 目录服务系统实现

目录信息存储在关系数据库中,Web 服务作为接收和处理用户资源查询、目录管理和资源访问请求。Web 服务接口提供用户通过消息实现目录服务,HTTP 接口提供用户直接通过浏览器查询和管理目录服务。HTTP 接口一般是目录体系中必须实现的,而消息访问接口可以根据各地的实际情况选择实现。目录服务系统的框架结构如图 2 所示。

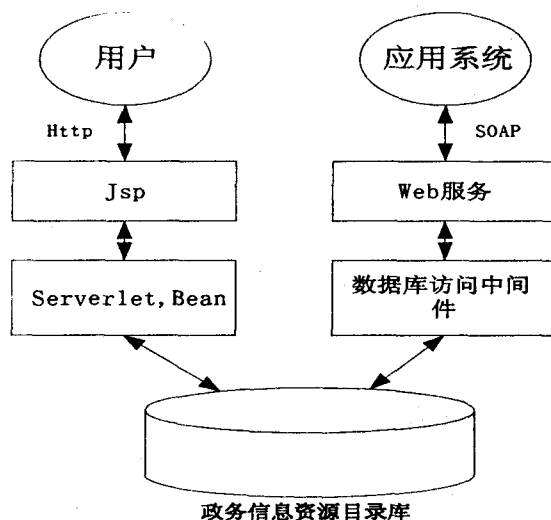


图 2 目录服务系统的框架结构

文中在“政务信息资源统一描述、注册管理和目录服务标准与原型系统”项目的基础上,在“政务目录服务与交换系统”原型系统中实现了上述两种目录服务方式,HTTP 目录访问方式采用 Java + tomcat 开发,目录服务器端部署 Java Bean 和 Servlet 处理用户的请求和完成与目录库的交互,前台提供 JSP 页面作为 HTTP 接口,并提供用户访问目录服务系统的界面。

3 结束语

文中依托了上海市科委科研基金项目(“政务信息资源统一描述、注册管理和目录服务标准”,编号:03DZ05015)。根据我国电子政务的发展特点和电子政务实际需求提出了基于元数据标准的电子政务目录服务的设计与实现,并且以上海市水务局为试点单位,根据其水务行业的需求,进一步完善了其电子政务的建设。

(下转第 48 页)

当用户在未登录的情况下登录系统的时候,SecurityFilter 过滤器会检测用户的系统登录情况,用户没有登录的情况下,访问应用系统会自动把页面跳转到系统登录界面。

2) SessionObject 登录用户基础权限信息类。

SessionObject 类,保存了用户登录的信息,同时保存登录后用户的身份信息等到 Session 中。

2.3.3 公共服务包 common

common 包里是系统公共核心类,为系统提供常量服务、消息服务、配置文件服务以及分页服务等。其主要有以下几个类,简介如下:

1) Config 类:读取 applicationConfig.properties 配置文件信息。

2) Constants 类:常量文件类。

3) Message 类:读取 applicationMessage.properties 自定义信息类。

2.3.4 工具包 util

util 包是应用框架的所有工具类,主要为系统提供 intelliWeb 与 VO 类型转化、文件、算法、唯一键值、字符和日期工具等服务。其中的几个关键类介绍如下:

1) DateUtil 日期工具类和 UniqueKeyUtil 产生唯一 KEY 类。

DateUtil 开发接口,主要是针对重复使用的 DateTime 日期时间类。UniqueKeyUtil 开发接口,产生唯一 Key 值类。

2) StringUtil 字符串截取工具类。

StringUtil 开发接口,主要是针对字符串的操作类,比如去掉空格,截取部分字符,替换等等操作。

3) BeanUtil 转换工具类。

BeanUtil 开发接口,主要是 DataSet 与 VO、VO 与 DataSet 互相转换的接口类。开发接口主要有:recordToVO(把 DataSet 转为单个 VO 对象),toSingleVO(把 DataSet 转为单个的 CustomDataset 对象),listToDataset

(把集合对象转为 CustomDataset 对象)等等接口。

2.3.5 异常包 exception

这个包中只有一个 BaseException 类,它是所有异常类的父类,所有异常类都要继承此类,并对此类进行扩充实现对异常的处理。

2.3.6 应用业务类包 webapp

webapp 包实现了部分常用的应用,包括权限管理、登录管理、日志管理、员工管理、机构管理、菜单管理和角色管理等应用开发中的常用模块,并对应到 auth, login, log, employee, inst, menu 和 operator 等包来实现,同时它们严格按照三层结构,区分出 Web 层、业务逻辑层和持久层,从而从上到下实现松耦合。

3 结 语

文中设计并实现的基于 intelliWeb, Spring 和 Hibernate 的应用框架经过一些开发人员与项目组的评估和使用,并在他们提出的宝贵意见的基础上进行了一些必要的改进,目前框架稳定、性能优越、可扩展性强,表现出相当高的实用价值和推广价值,下一步工作就是要将这个应用框架规范化、标准化,以实现大面积的推广。

参考文献:

- [1] 席晓峰,吕良双,逯 鹏.使用 J2EE 框架技术构建可重用的 Web 应用[J].计算机工程与应用,2005,41(6):208-214.
- [2] 沃尔斯,布雷登巴赫.Spring in Action 中文版[M].李 磊,程 立,周悦虹译.北京:人民邮电出版社,2006.
- [3] 陈天河.Hibernate 项目开发宝典[M].北京:电子工业出版社,2006.
- [4] 穆福森,吴观茂.基于 Struts + Spring + Hibernate Web 应用开发框架技术[J].电脑知识与技术,2006,2:81-82.
- [5] 李英军,吕 建,王宏琳.面向对象应用框架在油气勘探领域的应用研究[J].软件学报,1999,10(4):349-354.

(上接第 44 页)

文中根据政务工作的特点,在现有的政务信息资源注册管理模型基础上,提出了基于元数据的政务信息资源目录服务体系的框架并加以实现。最终完成该目录服务系统在上海水务局的试点应用。

参考文献:

- [1] 陈艳红,盛子刚.DC 元数据与网络信息检索(综述)[J].河北科技师范学院学报,2005,19(2):73-76.
- [2] 朱吉翔,孙 莉.电子政务信息资源注册管理模式的研究

[J].计算机工程与应用,2006,42(5):29-30.

- [3] 张道顺.元数据与银行业信息资源管理整合[J].中国金融电脑,2005(11):33-37.
- [4] 李莲花,刘 渊.基于元数据的网络资源管理系统的建设研究[J].信息技术,2002(11):42-44.
- [5] Dublin Core Metadata Initiative (DCMI)[EB/OL].2007-06. <http://dublincore.org/>.
- [6] 赵 阳.基于 OAI 的元数据收割系统结构框架分析[J].现代图书情报技术,2004(6):11-15.