

● 孙雨生 (武汉大学信息资源中心 湖北 430072)

基于 workflow 技术的数字图书馆互操作问题研究

摘 要: 数字图书馆的关键是解决各个信息仓储间的互操作性问题, 以实现跨仓储的统一检索。因此互操作问题是数字图书馆研究的核心问题之一。本文首先分析了数字图书馆的互操作问题, 然后介绍了 workflow 技术及解决互操作问题的方法, 最后提出了一种基于 workflow 技术的数字图书馆互操作问题解决方案。

关键词: 数字图书馆; 管理系统/workflow 技术

Abstract: The key problem of digital library is to solve the interoperability of different information repositories so as to realize unified retrieval. Therefore, interoperability is one of the kernel problems in the research on digital library. This article first analyzes the interoperability problem of digital library, then introduces the workflow technology and the method to solve the interoperability problem. Finally, it gives a solution to solve the interoperability problem of digital library based on workflow technology.

Keywords: digital library; management system/workflow technology

因特网的出现不仅促进了信息资源的迅速增长, 而且加剧了与用户信息需求难以得到满足之间的矛盾, 并最终导致一种新型信息资源组织和整合机构——数字图书馆的出现。

一般认为, 数字图书馆是一个网络环境下数字化的分布式信息资源体系结构, 基本特征是数字化的信息资源、网络化的存取环境和分布式的信息资源管理。但是现存的数字图书馆系统却基本上都属于网络环境下的集中式信息资源管理与服务系统, 特征是信息资源的集中式采集、信息资源的数字化加工和存储、通过网络来提供信息检索和服务^[1]。客观地讲, 这种数字图书馆模式并不适合现在及未来的信息资源环境, 无法从根本上解决信息资源不断增长和用户需求难以得到满足之间的矛盾: 一方面信息资源的客观分布性使得没有任何一个机构能够只依靠自己的力量就建立起用户需要的所有数字化信息资源或服务, 同时用户不断变化和日益明显的个性化需求也要求数字图书馆建立起对其业务流程不断地进行调整和优化的动态机制, 以有效地整合分布式信息资源; 另一方面则是分布式数字图书馆便于根据应用需要和存取方便的原则来配置信息资源, 在满足成员图书馆的局部自治要求的同时实现信息资源的共建共享。因此, 在网络信息环境下, 比较理想的数字图书馆模式是分布式数字图书馆。

所谓数字图书馆互操作问题是指数字图书馆内部各组件之间或数字图书馆之间交换与共享文档、查询和服务的能力^[2]。一般情况下, 用户信息需求的满足需要查询数字图书馆的多个信息仓储才能完成, 但是在网络信息环境下, 由于信息仓储的信息组织方式、处理方式及查询方式的差异性, 数字图书馆的互操作问题日益突出^[3]。同时,

不断变化的用户需求也要求数字图书馆建立一种能够不断优化业务流程并对业务内容进行调整的动态机制, 并在对分布式资源进行整合的基础上实现数字图书馆业务自动化, workflow 技术恰好就是这样的技术, 它能够在计算机支持下实现整个或部分业务过程的全自动化或半自动化, 满足各种机构优化业务流程和调整业务内容的要求。

1 数字图书馆的互操作问题分析

数字图书馆必须有效解决互操作问题, 实现对分布式资源的有效整合, 即要求数字图书馆能够有效地屏蔽各信息仓储间的差别, 提供一致的检索界面和检索技术, 由系统自动执行跨仓储的检索, 进行不同的信息格式、检索方式等方面的转换, 同时要能够满足数字图书馆业务规则和业务内容不断调整的要求, 形成一种有效的资源整合机制, 主要是解决好以下几方面问题^[4]:

1) 解决分布性和异构性的问题, 实现数字图书馆间的互操作。能够将分布、异构的资源整合成一个对用户透明的、统一的资源网络, 用户只需提出自己的需求, 数字图书馆就能够自动进行分布式信息资源的收集、整理, 并将最终的结果反馈给用户。

2) 支持各个信息仓储的局部自治性, 保证整个系统的开放性。要能够容纳现有的信息系统, 持续支持各个系统的自主建设与发展, 即不需要对各个信息系统的资源组织模型进行大的变动, 并有效保证对知识产权资源使用的本地控制, 同时能够实现资源的共建共享。

3) 满足数字图书馆的动态变化要求, 实现业务处理的自动化。由于成员馆自身的变化, 如业务规则和流程的

变化,用户需求的变化等原因,数字图书馆一方面要有适应变化的能力;另一方面要将这些变化迅速反映到现实系统中,达到对分布式数字资源的有效组合和整理。

2 工作流技术

所谓工作流技术是指在计算机支持下实现整个或部分业务过程的全自动化或半自动化。在这些业务过程中,文档、信息或任务按照预定的一组规则在所有参与者之间传递,以实现整个业务过程的自动化,其中过程是指一组相关联的流程或活动,它们相互协作,来实现组织的一个目标。该技术的最大优点是将应用逻辑与过程逻辑分离,在不修改具体功能的情况下,通过修改过程模型改变系统功能,完成对部分过程或全过程的集成管理,有效地把人、信息和应用组合起来^[5]。

2.1 工作流管理系统的体系结构及运行机制

工作流管理系统(Workflow Management System, WFMS)是一个运行计算机软件程序的系统,它完全定义、管理和执行工作流,实质就是通过管理工作活动的顺序,调用与各工作活动步骤相关的人工或IT资源来实现业务过程的自动化,系统中软件的执行顺序用计算机信息表达方式描述的工作流逻辑来驱动。体系结构见图1,其中用户接口系统是工作流管理系统和应用系统提供同用户之间的交互式接口;应用系统处理工作流过程中特定的应用数据;数据库管理系统负责存储和管理所有的工作流数据;通信系统提供连接其他工作流管理系统的通信服务;操作系统向所有其他各层提供最基本的系统功能调用。

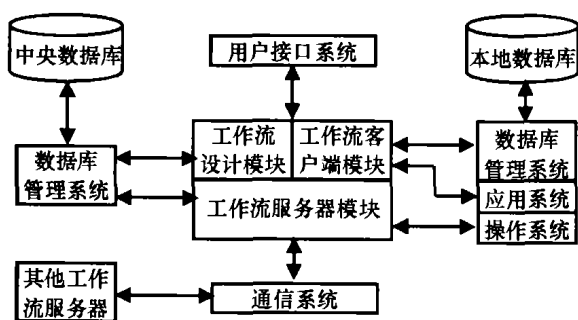


图1 工作流管理系统的整体结构

如图1所示,工作流管理系统主要包括工作流设计模块、工作流服务器模块和工作流客户端模块三部分^[6],各部分的具体功能是:①工作流设计模块负责为定义工作流语义提供支持,保证工作流语义的语法正确性,设计好的工作流数据保存在中央数据库中,供工作流服务器使用。其中工作流设计引擎是核心组件,它提供标准的工作流设计过程。②工作流服务器模块是整个工作流管理系统的中枢,完成工作流的执行功能。它通过操作工作流数据库中的数据来完成工作流活动的调度,实现各工作流之间的通

信,管理群体工作成员的协作信息和协作活动,共同完成业务处理过程,同时它对系统的所有工作流实例进行监控,保证工作流的正确运行。它的核心是工作流服务器执行引擎,主要功能是处理工作流事件。③工作流客户端模块为工作流的执行和管理提供底层的终端用户服务,它可以同工作流执行服务器进行通信,执行模块向用户提供交互式访问功能,用户借此来完成他们在工作流应用中的任务。

2.2 工作流技术中互操作问题的解决方法

工作流互操作问题的核心是标准化问题,与之相关联的是建立一个开放的多协议与多种类中间件环境与执行构架来满足数字图书馆的互操作要求,而要满足“成员馆间工作流”则需要建立可与不同组织模型实现交互(包括安全策略与模型,业务规则等)、支持包含在动态与自适应工作流中的多重参与组织模型。这种在各个成员馆之间运作的虚拟流程可以通过在支持流程离散分段时创建的工作流定义集来实施,同时为了避免在实现端到端的业务流程操作中创建自动化孤岛,需要在不同工作流产品间实现消息交换,达到流程互操作与集成,从而驱动与管理整个数字图书馆流程的运转。在工作流技术中,互操作问题的解决主要是基于以下两个标准来实现的^[5]:

2.2.1 WfMC 工作流互操作标准 它定义了一个满足工作流执行服务端到端交互的简要协议,支持在一个执行服务管理的工作流中的活动通过另一个执行服务管理的子过程来实现。标准提供了仅在工作流流程级的互操作,流程实例内部在跨越业务域边界时并不需要了解,这样就可以实现分布式的流程整合。其中,工作流过程是一套属性集,定义了工作流过程的关联操作关系,在过程生命周期中产生结果的表示,以及表示当前过程状态的属性。该标准定义了一个最小状态模型,包括状态的运行、完成及中断,同时还定义服务请求方向服务供给方发送的服务请求内容,控制与监视远程过程的执行。

2.2.2 OMG 工作流管理工具标准 它是基于 WfMC 标准建立的,定义基于业务对象的分布式工作流应用框架,说明完成工作流流程组件的互操作,监视工作流的执行,并解决联合工作流组件到工作流流程资源中的有关问题。其基本模型对工作流服务请求器和工作流服务供给器进行了特征描述,主要接口如下:工作流请求器接口,用于提出作业请求。可以将它的兴趣注册到工作流,提供流程可以使用的操作。工作流流程接口,执行作业请求通常委派其到另一实体,它提供操作作业请求的执行,并获得其状态。工作流流程管理器接口,提出一个工作流定义,作为一个工作流的工厂,允许工作流请求器在工作流流程创建时注册。工作流活动接口,提供了一个适配器,把存在的业务对象集成到工作流中,与工作流请求器接口一起协

同,并允许主工作流与另一工作流进行交互。另外,标准还定义了作业序列处理和流程审核接口。

总之,我们通过工作流设计工具来改变流程定义,满足数字图书馆的动态变化性要求,通过增加应用系统中的应用数据来满足数字图书馆的扩展性要求,通过工作流技术中的用于解决互操作问题的标准来解决数字图书馆的互操作问题,满足其业务规则的变化和业务流程的调整,实现数字资源的有效整合。

3 基于工作流技术的数字图书馆互操作问题解决方案

基于工作流技术的数字图书馆互操作问题解决方案主要应该考虑两方面的因素:首先是系统底层基础通信机制,这关系到系统各个部分以何种方式进行互操作;其次是系统各个组成部分之间的协作过程,即具有不同功能、相互独立的软件模块的互操作过程,也就是工作流的运转过程,这一部分会经常发生变化。前者我们依靠 WFMC 工作流互操作标准和 OMG 工作流管理工具标准来解决,后者我们则通过分布式的工作流管理系统来实现。

虽然数字图书馆要求工作流管理系统能够集成并调用外部应用程序以完成自动化活动中特定的任务,但是任何 WFMS 实现都不可能具有对所有潜在应用的调用处理逻辑,因此就需要通过应用代理层来封装将被调用的各类应用,并为工作流提供统一标准的接口。这是因为应用代理具有自治能力,可与其他代理进行交互,能对环境变化做出响应并且可以基于目的采取行动^[7]。所以基于工作流技术的数字图书馆系统结构分为基本功能体层、应用代理层、工作流系统层三个层次,最后由分布式 WFMS 来接受用户的服务请求,并将最后的结果返回给用户(如图 2 所示)^[8],其中分布式 WFMS 的结构如图 3 所示^[9]。

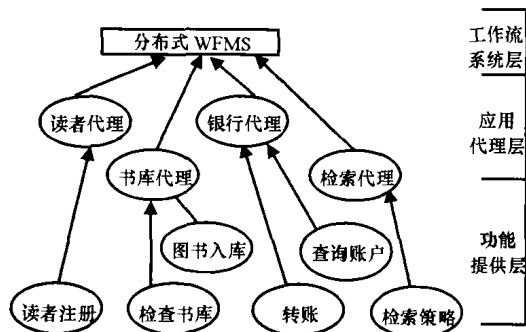


图 2 基于工作流技术的数字图书馆系统结构图

图 2 中的基本功能体层主要是提供系统最基本的功能。功能代理层一方面将基本功能体层的基本功能体映射成代理的私有方法;另一方面则通过代理层的规则来确定本代理所管理的基本功能体的前后执行次序(业务过程的

流程);工作流系统层为多代理之间的协作提供一个环境(包括代理之间的通信),并根据环境确定代理活动的前后次序,进而与功能代理层所确定的业务子流程一起形成数字图书馆的整个业务流程。这样我们就建立了一个开放的多协议与多种类中间件环境与执行构架,并通过工作流管理系统实现分布式的流程整合和工作流组件间的互操作,同时满足传统图书馆局部自治和共建共享的要求,实现数字图书馆业务过程自动化和动态调整。

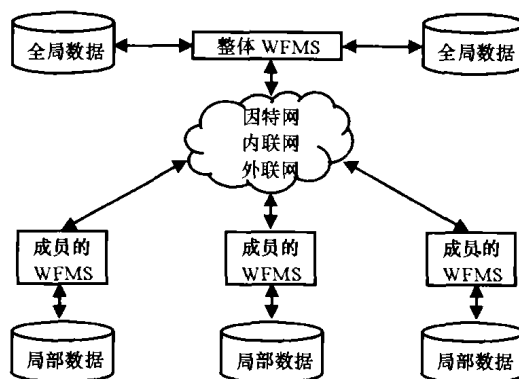


图 3 分布式 WFMS 结构图

4 结束语

一般而言,解决数字图书馆的互操作问题应当以中间件技术来实现分布式数字图书馆的成员馆内部及相互间系统的互联与互通,解决系统之间的异构问题;利用工作流技术来控制数字图书馆的业务流程,将业务规则和数字资源的变化及时反映到系统的流程中去,以满足系统的动态变化要求。□

参考文献

- 耿骞,汤艳莉.多 Agent 技术与分布式数字图书馆系统框架.中国图书馆学报,2003(5):52~56
- 郎庆华.略论数字图书馆互操作性的技术实现方式.现代情报,2003(6):57~58
- 王军等.数字图书馆的体系结构.情报学报,2000(6):564~572
- 张晓林.分布式数字图书馆机制.情报学报,2002(1):63~70
- 侯宏仑等.电子商务中的工作流技术与标准化问题研究.中国机械工程,2002(4):301~306
- 林成栋,冷劲松,张超.工作流管理系统的体系结构研究.中国管理科学,2002(3):56~60
- 封明玉,赵政.基于工作流和 CORBA 技术的可重构信息系统研究.微型机与应用,2002(1):36~38
- 董福壮,罗伟其.XML 在工作流管理系统中的应用.计算机工程与应用,2003(33):223~226
- 张艳,史美林.工作流技术在虚拟企业信息处理中的应用.清华大学学报(自然科学版),2001(7):90~93

作者简介:孙雨生,男,1980年生,硕士生。

收稿日期:2004-11-19