

# 基于网格的数字图书馆 应用系统框架构建

张继东 著

经济日报出版社

本书是国家自然科学基金项目“基于语义网格的数字图书馆应用模型研究”（项目批准号：70803009）的成果之一

# 基于网格的数字图书馆 应用系统框架构建

张继东 著

经济日报 出版社

## 图书在版编目 ( CIP ) 数据

基于网格的数字图书馆应用系统框架构建 / 张继东著.  
北京: 经济日报出版社, 2009.5

ISBN 978-7-80180-999-5

I . 基… II . 张… III . 数字图书馆 - 研究 IV . G250.76

中国版本图书馆 CIP 数据核字 ( 2009 ) 第 064148 号

---

### 基于网格的数字图书馆应用系统框架构建

---

作 者	张继东
责任编辑	汤雪梅
责任校对	刘 涛
出版发行	经济日报出版社
社 址	北京市宣武区白纸坊东街 2 号 ( 邮政编码: 100054 )
经 销	全国新华书店
印 刷	成都新千年印制有限公司
开 本	880×1230mm 32 开
印 张	7
字 数	175 千字
版 次	2009 年 5 月第一版
印 次	2009 年 5 月第一次印刷
书 号	ISBN 978-7-80180-999-5
定 价	20.00 元

---

版权所有 盗印必究 印装有误 负责调换

## 前 言

随着数字图书馆的研究和发展，越来越多的资源和服务也日益增加。由于这些数字图书馆系统形成的时期各不相同，因而系统内的资源类型和格式也各异。用户为了能尽量查到需要的资料，不得不在不同数据库中寻找，还要熟悉各种平台的检索和收藏特点。这就造成了图书馆用户即使面对着浩瀚的信息资源，却无法查准、查全自己所需资源的困境。同时，图书馆的各种网上服务也逐步兴起，但是用户却不习惯在不同的系统之间切换，使得服务的效率降低。

数字图书馆是一项持续的工程，数字图书馆之间必须是交互的、开放的，所以数字图书馆所采用的技术平台必须是开放的、持续发展的和规范的。数字图书馆界近年来开始考虑解决这种资源和服务系统处于

分散、孤立存在的情况。研究者们根据其他领域的类似研究成果，提出了多种资源整合模型。由于互联网和计算机技术的日新月异，对数字图书馆的服务要求也日益突出，服务系统的兴起也同时带来了与资源类似的整合要求。

当前数字图书馆领域研究的热点之一就是，针对分布式的、多样化的信息资源，提供一种支持不同信息系统之间互操作以及这些系统间无缝交换和共享信息资源与服务的解决方案，并由此构造一种逻辑上的集成信息服务机制，从而形成基于集成信息服务的数字图书馆。然而，由于各个数字图书馆存在的系统和资源上的异构性，给信息系统集成与信息资源共享带来了很大困难。另外，随着数字资源日趋丰富，数据规模逐渐扩大，这就会给我们传统的检索工具带来很大压力，性能上的瓶颈也将成为一个不可避免的问题。

网格技术是当前计算机技术发展的一个重要方向，它把用通信手段连接起来的资源无缝集成为一个有机整体，给用户提供一种基于互联网的新型计算平台。网格环境下的资源地理分布极广，类型和数量巨大，动态变化，且工作在异构平台上由不同的管理策略控制，其复杂性对信息服务提出了不一般的要求，也使得对于网格信息服务的研究成为一项比较有意义且具有挑战性的工作。

作为网格技术的一个重要分支，信息网格就是要利用现有的网络基础设施、协议规范、Web 和数据库技术，为用户提供一体化的智能信息平台，其目标是创建一种架构在 OS 和 Web 之上的基于 Internet 的新一代信息平台和软件基础设施。在这个平台上，信息的处理是分布式、协作和智能化的，用户可以通过单一入口访问所有信息。信息网格追求的最终目标是做到服务点播 (Service On Command) 和一步到位的服务 (One Click is Enough)。

本书重点研究和讨论的问题是，利用信息网格技术来解决数字图书馆的互操作问题及检索上的性能瓶颈。本书在仔细研究数字图书馆互操作协议及网格技术的基础上，提出了数字图书馆网格应用系统的框架模型，详细介绍了框架中使用的关键技术，并且给出了这个框架模型的详细解决方案。

本书是作者主持的国家自然科学基金项目“基于语义网格的数字图书馆应用模型研究”（项目批准号：70803009）的成果之一。

本书研究和讨论的重点在于：

(1) 深入研究和讨论信息网格技术，将网格技术与数字图书馆工程结合在一起，试验证明此方案是可行的。目前国内的网格研究都是基于计算网格的，对于网格在数字图书馆中的应用还是空白。论文及其所

属的项目把网格、集群等分布式技术应用到了数字图书馆工程中，使得数字图书馆的计算机系统可以实现伸缩性的部署和多样性的部署，从数字图书馆功能的角证明使用了使用网格技术是可行的。

(2) 应用 Java 的工厂模式实现了信息网格的调度规则，同时把语义技术应用于信息网格调度中，提出并实现了一种新的调度模式。模型提出了在网格平台 Globus 上进行二次开发的技术架构体系，并加以实现。目前国内对于网格调度的研究多集中在计算网格的任务队列调度方面，而对于信息网格的调度研究得很少。

(3) 用 Ganglia 实现基于 Linux 的网格监控，与网格平台 Globus 的 MDS 组件集成，提出并实现了基于 Globus 的信息提供模型。对于基于网格平台的信息网格节点登记方式，以及信息网格的网格引擎的开发与实现都提出了具体的实现方式。

全书共九章，可分为三部分：

第一部分（包括第 1 章）主要介绍本书的研究背景。

第二部分（包括第 2、3 章）介绍并分析了数字图书馆网格应用系统的基本技术。第 2 章介绍了数字图书馆技术及互操作协议的选择，第 3 章介绍了网格技术体系、国外数字图书馆网格应用的情况。

第三部分（包括第 4、5、6、7、8、9 章）设计并

初步实现了数字图书馆网格应用系统。

第4章介绍了本系统的总体设计方案及其基本框架。本章站在系统总体设计的角度,描述了整个项目的设计框架,并且按照从底层到顶层的顺序简要分析了各个模块的工作流程、设计方法。

第5~9章分别介绍了系统的各个部分的设计方案及实现方法。

其中:

第5章介绍了信息网格的信息组织技术,包括信息网格信息资源构建、信息组织模块的整体框架及各个部分的设计实现方法。本章对于系统的底层信息资源、元数据的建设方法进行了描述,设计和实现了元数据的收集、存储、索引、发布模块,并且给出了OAI协议中六组请求的响应的实现方法。

第6章介绍了信息网格的资源调度策略,包括网格注册中心(UDDI)、调度中心、调度算法二次开发的实现方法,并且在6.5重点叙述了如何应用WS-Resource工厂模式实现网格调度的方法,即在globus上进行二次开发的方法和过程。其中调度中心又包括监测调度、语义调度两部分,监测调度模块的信息模块由MDS实现,集群监测模块由ganglia实现,语义调度模块采用WSMO框架实现。

第7章介绍了信息网格的检索方法与策略,应用



了 JAVA EE 5 方案。本章基于 EJB 和 WEB SERVICE 构建了业务层，运用 Struts 框架构建了 Web 层。

第 8 章介绍了信息网格的可视化框架及其实现，包括了普通用户、网格系统管理者、图书管理者三个层面，提供了系统的主要截图。

第 9 章介绍了系统模型压力测试情况分析，进行了系统瓶颈的分析，提出了进一步改进的策略。

## 目 录

前 言	(1)
1 引言	(1)
1.1 研究背景和意义	(1)
1.2 本领域的国内外研究现状	(5)
2 数字图书馆技术	(10)
2.1 数字图书馆及其互操作性	(10)
2.1.1 数字图书馆概述	(10)
2.1.2 数字图书馆信息资源构建与管理机制	(11)
2.1.3 数字图书馆的互操作性	(13)
2.2 数字图书馆互操作协议	(15)
2.2.1 OAI-PMH 协议	(15)
2.2.2 Z39.50 协议	(19)
2.2.3 ZING 协议	(22)

2.3	互操作协议实现的技术选择	(26)
3	网格技术	(31)
3.1	网格概述	(31)
3.1.1	网格概念	(31)
3.1.2	网格应用	(33)
3.2	网格平台 Globus	(35)
3.2.1	GSI	(37)
3.2.2	MDS	(41)
3.2.3	GRAM	(43)
3.2.4	GASS	(46)
3.2.5	GridFTP	(47)
3.2.6	GRM	(49)
3.3	国外数字图书馆网格应用方案	(50)
3.3.1	DELOS	(51)
3.3.2	Digital Library Grid	(52)
3.3.3	FreeLib - Peer To Peer Digital Library	(53)
3.3.4	ARCO	(55)
3.3.5	DILIGENT	(56)
3.3.6	GRACE	(57)
3.3.7	GridIR	(59)
4	数字图书馆网格应用系统框架体系	(61)

---

4.1 数字图书馆网格应用系统的设计思路及总体结构	
.....	(61)
4.1.1 设计思路	(61)
4.1.2 系统总体结构	(62)
4.2 数字图书馆网格应用系统的设计与实现	(62)
4.2.1 数据资源层的设计	(62)
4.2.2 网格 Globus 平台	(65)
4.2.3 业务层和表示层部分	(71)
5 信息网格的信息组织技术及其实现	(73)
5.1 信息网格信息资源构建	(74)
5.1.1 节点信息资源构建	(74)
5.1.2 元数据构建与部署	(77)
5.1.3 Web 服务资源构建	(80)
5.2 信息组织模块的整体框架	(81)
5.3 元数据收集模块的设计	(82)
5.3.1 元数据收集执行模块	(83)
5.3.2 元数据存储模块	(86)
5.3.3 元数据索引模块	(92)
5.3.4 属性设置模块	(95)
5.4 元数据发布模块	(97)
5.5 数据层的管理方案	(98)
5.5.1 数据库部署形式: 分布式被动模式	(98)

5.5.2	元数据管理措施	(99)
6	信息网格的资源调度策略	(101)
6.1	信息网格资源调度模块整体框架	(101)
6.1.1	设计思路	(101)
6.1.2	网格平台注册中心总体思路	(102)
6.1.3	网格平台调度中心总体思路	(102)
6.2	信息网格注册中心的实现	(105)
6.2.1	UDDI	(105)
6.2.2	UDDI的实现方法	(106)
6.3	信息网格监测调度的实现	(110)
6.3.1	信息模块 MDS	(111)
6.3.2	集群监测工具 Ganglia	(114)
6.3.3	集群监测调度模块的设计和实现	(117)
6.4	信息网格语义调度的实现	(126)
6.4.1	网格语义调度的基本思路	(126)
6.4.2	选择使用 WSMO 框架实现网格的语义调度	(127)
6.4.3	使用 WSMO 框架遇到的问题及解决方案	(129)
6.4.4	使用 WSMO 进行 Web 服务描述的实现方法	(130)
6.5	Web Service 资源框架 (WSRF)	(132)

6.5.1	无状态的服务与有状态的服务的比较 .....	(133)
6.5.2	WS-Resource .....	(133)
6.5.3	WS-Addressing .....	(134)
6.5.4	Web 服务描述文件 (WSDL) .....	(136)
6.6	应用 WS-Resource 工厂模式实现网格调度 ...	(140)
6.6.1	WS-Resource 工厂模式 .....	(141)
6.6.2	应用 WS-Resource 工厂模式创建有状态资源 .....	(142)
6.6.3	Grid Service 的部署 .....	(149)
7	信息网格的检索方法与策略 .....	(151)
7.1	信息网格检索模块描述 .....	(151)
7.1.1	检索模块需求分析 .....	(151)
7.1.2	采用 JAVA EE 5 分布式计算框架 .....	(152)
7.1.3	检索模块的多层体系结构 .....	(153)
7.1.4	系统应用服务器和集成开发环境的选择 ...	(154)
7.2	数据层实现 .....	(155)
7.3	基于 EJB 和 WEB SERVICE 构建业务层 .....	(156)
7.3.1	部署 EJB3.0 和 Web Service .....	(156)
7.3.2	数据连接池的设计 .....	(158)
7.3.3	新的 JAVA 持久性运用 .....	(159)
7.3.4	Web Service 客户端的实现 .....	(162)
7.4	基于 Struts 框架构建 Web 层 .....	(165)

7.4.1	MVC 与 Struts 简介	(165)
7.4.2	构建 STRUTS 框架的模型 (Model) 组件	(165)
7.4.3	构建 STRUTS 框架的视图 (VIEW) 组件	(167)
7.4.4	构建 STRUTS 框架的控制器 (Controller) 组件	(167)
7.5	任务分发 (并行检索) 的设计与实现	(168)
7.6	检索模块的集成与部署	(170)
8	信息网络的可视化框架及其实现	(172)
8.1	数字图书馆网格应用系统 (DLGM) 的可视化 框架结构	(172)
8.1.1	可视化方案的总体思路	(172)
8.1.2	可视化系统的体系结构	(173)
8.2	面向普通用户的可视化	(174)
8.2.1	面向普通用户的可视化思路	(174)
8.2.2	检索调度流程的可视化	(177)
8.2.3	检索结果的可视化	(179)
8.3	面向网格系统管理者的可视化	(181)
8.3.1	基于 Ganglia 的监控可视化流程	(181)
8.3.2	监控可视化的视图层	(181)
8.4	面向图书管理者的可视化	(183)

---

8.4.1 元数据管理可视化 .....	(185)
8.4.2 本地数字图书馆资源信息可视化 .....	(186)
9 系统模型压力测试情况分析 .....	(187)
9.1 测试环境 .....	(187)
9.2 测试情况 .....	(187)
9.3 系统软件瓶颈 .....	(190)
9.4 系统优化建议 .....	(190)
参考文献 .....	(192)
后 记 .....	(205)



# 1 引 言

## 1.1 研究背景和意义

20 世纪 90 年代兴起的“数字图书馆”，起源于 1993 年由美国国家科学基金会（NSF）、美国国际部尖端研究项目机构（DARPA）、国家航空与太空总署（NASA）联合发起的数字图书馆创始工程（DLI）。随着数字图书馆的发展，它已经“不仅仅是一个有着信息管理工具的数字收藏的等价词，更是一个环境，它将收藏、服务和人带到一起以支持数据、信息乃至知识的全部流程，包括从创造、传播、使用到保存的全过程”。特别是在美国国家科学基金会 2003 年的《Knowledge lost in information》报告提出之后，数字图书馆作为未来社会“普遍