

Fenbushi shuzi tushuguan ziyuan zhenghe
yu fuwu jicheng de guanli yanjiu

分布式数字图书馆资源整合
与服务集成的管理研究

谢春枝 著



浙江工商大学出版社

Fenbushi shuzi tushuguan ziyuan zhenghe
yu fuwu jicheng de guanli yanjiu

分布式数字图书馆资源整合
与服务集成的管理研究

谢春枝 著



浙江工商大学出版社

图书在版编目(CIP)数据

分布式数字图书馆资源整合与服务集成的管理研究/谢春枝著. —杭州:浙江工商大学出版社, 2009. 3
ISBN 978-7-81140-052-6

I. 分… II. 谢… III. 数字图书馆—信息管理: 资源管理—研究 IV. G250.76

中国版本图书馆 CIP 数据核字(2009)第 035661 号

分布式数字图书馆资源整合与服务集成的管理研究

谢春枝 著

责任编辑 余欣然

责任出版 汪 俊

封面设计 汪 俊

出版发行 浙江工商大学出版社

(杭州市教工路 149 号 邮政编码 310012)

(E-mail: zjgsupress@163.com)

(网址: <http://www.zjgsupress.com>)

电话: 0571-88823703, 88831806(传真)

排 版 杭州大漠照排印刷有限公司

印 刷 杭州广育多莉印刷有限分司

开 本 787mm×960mm 1/16

印 张 11

字 数 203 千

版 次 2009 年 2 月第 1 版 2009 年 2 月第 1 次印刷

书 号 ISBN 978-7-81140-052-6

定 价 24.00 元

版权所有 翻印必究 印装差错 负责调换

浙江工商大学出版社营销部邮购电话 0571-88804227

序

用“数字图书馆”做关键词检索 CNKI 的“中国期刊全文数据库”,可以得知 1994 年国内的期刊中开始出现了“数字图书馆”这个名词,报道美国开始实施数字图书馆计划的消息。1995 年,北京大学图书馆朱强先生在《大学图书馆学报》上发表《数字图书馆:21 世纪图书馆的原型》一文,比较详细地介绍了美国的“数字图书馆创始”计划,这也许是最早的系统介绍数字图书馆的文章了。根据检索结果,这两年一共只有 14 篇文章涉及到数字图书馆。其后,关于数字图书馆的文章迅速增加,1996 年到 2000 年有 1617 篇,2001 年到 2005 年激增到 10917 篇,到目前为止,14 年间相关文章的总数达到 16697 篇,多么惊人的数字!

从文章研究的内容来看,对数字图书馆的探索走过了一个逐步深入的历程。从最初的介绍国外动向开始,到研究数字图书馆的系统结构,到数字图书馆,到技术平台,再到资源建设、服务模式、标准问题、管理问题、知识产权问题,等等,涵盖了数字图书馆的各个方面。

探索并没有停留在理论探讨的层面,实际建设数字图书馆的项目层出不穷。少数图书馆很快把自己的机读目录、订购的全文和索引数据库等做在了网页上,打出了数字图书馆的牌子。不过这个做法并没有形成燎原之势,大家很快认识到,数字图书馆并不只是一个技术产品,而是构建了一个全新的信息环境,这是凭一馆之力难以完成的,于是图书馆联盟成为数字图书馆建设的主力军,信息资源共建共享的观念更加深入人心。

从目前情况来看,打数字图书馆牌子的有三种情况。一是图书馆联盟,特别是得到政府资助的图书馆联盟,在积极推进数字图书馆的建设和发展;二是一些单个图书馆,特别是能够得到充足经费的图书馆,仍然在探索构建满足本馆读者需要的数字图书馆;三是数据库服务商推出的各种类型的数字化文献服务系统,如超星数字图书馆、书生之家数字图书馆等,似成三足鼎立之势。

曾有智者提出全球只要建一个数字图书馆的观点,因为信息技术和网络的

发展使得信息的复制和传播变得如此容易,重复制作同一个信息的行为都会被认为是一种浪费,信息大同世界似乎可望可即。然而,现实告诉我们,我们还不能指望一个放之四海而皆准的标准数字图书馆模式,在数字图书馆建设中百花齐放、多足鼎立的局面还将延续下去,我们现在能做的,是整合分布式的、异构的数字图书馆资源,以形成适合特定读者群的资源和服务系统。谢春枝博士的这本专著,就是在这方面的一个积极的探索。这个探索既不是起点,也不是终点,她只是在数字图书馆发展道路上众多的探路者中的一员。我们希望,谢春枝博士在数字图书馆建设的理论、方法和实践方面所做的工作,能够对数字图书馆的建设者们有所帮助和启迪。

燕今伟

2008年5月于珞珈山

随着因特网和数字图书馆建设的飞速发展,有关数字图书馆建设的理论研究和项目实践层出不穷,数字信息资源和数字信息服务也极大地丰富。用户在享受数字信息快餐的同时,也面临着如何选择、如何过滤、如何简单有效地获取数字信息的困惑,越来越多的用户希望能够得到关于数字资源的空间范围和基于知识内容的相互联系的整体揭示。解决分布式异构环境下的数字资源整合和服务集成问题日益成为数字图书馆界最重要的研究领域之一。

因此,尽管众多已经建成和正在建设的数字图书馆是在分布、异构的系统条件下进行的探索和实践,但在网络带给我们的一体化的数字信息环境中,个体的数字图书馆不再仅仅是基于资源的孤立的服务系统,而是一种以用户和需求为中心来聚合资源、服务、信息利用活动的动态机制,其目标和功能都着力于支持用户发现信息、利用信息、提炼知识、解决问题,通过资源与工具的链接、集成和嵌入,动态地构建符合用户需要的服务机制。

正是由于这种分布异构环境下资源整合和服务集成需求的日益强烈,促使各种探索资源整合和服务集成的技术方案应运而生,数字图书馆也纷纷开展了内敛型的或向外扩张型的资源整合和服务集成活动,呈现出多级、多边、多种形式的合作和运行模式。

论题“分布式数字图书馆资源整合与服务集成的管理研究”是从管理的角度来研究资源整合与服务集成的相关问题。人类的各种领域的实践活动表明,技术和管理之间是一种相辅相成的关系,先进的技术只有辅以相应的科学管理才能达到最佳的绩效,特别是在技术和需求发展到一定程度的时候,滞后的管理甚至会制约技术潜能的发挥。分布式数字图书馆领域业已开展的资源整合与服务集成实践证明,资源整合与服务集成不仅仅是技术问题,更是一个社会问题。

本书试从资源整合与服务集成的管理角度入手,围绕资源整合与服务集成过程中与管理相关的因素,分析了分布式数字图书馆资源整合与服务集成的动因、原则和方法、目标和功能、模式和环境;在挑战与对策部分,着重分析了资源整合与服务集成的策略、标准、质量控制、版权、绩效评估及可持续发展等管理方面的问题;最后,以 Google 数字图书馆项目为例,分析了其在资源整合与服务集

成过程中所采取的策略、技术手段、合作模式、服务功能、赢利模式以及存在的风险等。

全书由绪论、正文和结语 3 部分组成,共 8 章,主要内容如下:

0 绪论

本章介绍了论文的研究背景及选题意义,指出,正是由于未来的数字图书馆是合作互联的知识环境以及分布式环境将是未来数字图书馆系统的主流环境等背景,从而使资源整合与服务集成的研究有较强的现实意义和迫切性;从理论研究、技术和实现方法、管理研究及实例研究四个方面对本选题相关的中外文献进行了回顾分析;简单地介绍了本书的研究内容、研究方法和特色之处。

1 概述

本章探讨了分布式数字图书馆资源整合与服务集成的三个动因,认为其是数字图书馆发展到第二代后的必然结果,是应对用户信息使用行为变化后的必然选择,是缓解分布信息环境下“数据孤岛”现象的迫切要求;分析了资源整合与服务集成的含义,阐述了两之间互动与相辅相成的关系,从而说明了本书将两者进行并题研究的基础和意义。

2 分布式数字图书馆资源整合与服务集成的目标和功能

本章阐述了分布式数字图书馆资源整合与服务集成的目标,着重分析了资源整合与服务集成后系统所具有的功能,如门户、统一认证、统一检索、权限管理、个性化服务及数字参考咨询等;介绍了相关案例,对实现某些功能带来的利弊进行了探讨。

3 分布式数字图书馆资源整合与服务集成的原则和方法

本章从不同层面分析了资源整合与服务集成的原则,包括整体层面上的 7 个原则、操作层面上的 5 个原则;从资源整合的广度和深度两个方面讨论了资源整合的具体方法,分析了这些方法在具体的数字图书馆资源整合与服务集成活动中的运用;对目前资源整合与服务集成活动中所使用的主要技术进行了介绍,如摄取、OAI、收割、RSS、网页采集和元搜索等。

4 分布式数字图书馆资源整合与服务集成的模式和环境

分布式数字图书馆资源整合的模式与环境研究包括如何选择整合对象、采取的合作模式、资源整合的目标定位等。本章结合国内外知名数字图书馆资源

整合与服务集成项目实例,从资源整合与服务集成的对象、功能、范围和主体等方面分析了资源整合与服务集成的模式和环境;最后指出,资源整合和服务集成的模式和环境正呈现出多边合作和提倡资源集合级整合的发展趋势。

5 分布式数字图书馆资源整合与服务集成的挑战 and 对策

此章是本书的重点章节,笔者在前面章节论述的基础上,对管理问题进行了集中研究。主要围绕分布式数字图书馆资源整合与服务集成的策略选择、标准规范、质量控制、版权问题、绩效评估和可持续发展等六个管理方面的要素进行了详细的分析和探讨,结合案例研究,指出每个管理要素的研究和实践现状、面临的问题和解决的办法和措施等。

6 分布式数字图书馆资源整合与服务集成的案例分析

本章以用户熟悉但一直以来备受争议的著名搜索引擎 Google 为例,回顾了其数字图书馆资源整合与服务集成的历程,分析了其功能、合作模式、实现资源整合与服务集成的技术方法、赢利模式和存在的风险等,是运用前文观点对实例进行的一次剖析。

7 结语

本章首先对全书的研究进行了总结和回顾,再次强调了分布式数字图书馆资源整合与服务集成选题的意义,对书中提出的资源整合与服务集成管理方面的几个要素问题进行了梳理和总结,着重分析了几个要素之间的相互关系。此外,对分布式数字图书馆资源整合与服务集成下一步应该研究的命题进行了探讨。

关键词: 分布式数字图书馆 资源整合 服务集成 管理

Along with internet and the digital library construction rapid development, the related digital library construction fundamental research and the project practice come forth one after another incessantly, digital information resources and digital information services also are enormously rich. The user while enjoy the digital information fast-food, also face the confusions that how to choose, how to filter and how to gain the digital information simply and effectively. More and more users hope to obtain the digital resources with the spatial scope and promulgating information based on the knowledge content mutual relations. How to solve the integration of digital resources and services under the distributed, heterogeneous environment has become one of digital library most important research areas day by day.

Therefore, although all the multitudinous digital libraries whether already completed or under construction are carried on under the distributed and heterogeneous environment, in the incorporated digital information environment caused by the network, the individual digital library is no longer merely an isolated service system based on the resources, but is one kind of dynamic mechanism that gets together the resources, services and information utilization which take the user and its demand as the center. Its goals and functions all strengthen to support user to discover the information, use the information, refine the knowledge and solve the problems, and through the resources and links, integration and inserting, to construct the service mechanism which the user needs.

Just because this distributed heterogeneous environment, the demand for the integration of resources and services is intense day by day. Many kinds of technical plans for digital resources and services integration arises at the historic moment, the digital libraries have also carried out digital resources aggregation and the service integration practices by themselves, presenting the

characteristics of multistage, multilateral and various forms.

The thesis studies the related questions of the digital resources aggregation and services integration from the management angle. Mankind's practice indicates that, between the technology and the management, there is a relation which one kind complements another. The advanced technology can achieve the best achievements only with assistance by the corresponding scientific management, especially when the technology and demand have developed in certain degree, the lag management even would restrict the technology from exerting its potentials. The resources aggregation and the services integration which the distributed digital libraries already develop prove, the resources aggregation and the services integration not merely is the technical question, but a social question.

This article tries from the resources aggregation and services integration management angle, to revolve the management related factors in the resources aggregation and the services integration process, and analyze the moves of the distributed digital library resources and services integration, as well as the principles and the methods, the goals and the functions, the pattern and the environment, etc. In the challenge and the counter measure part, the paper emphatically analyze the management aspects, such as the integration strategies, standards, quality-controls, copyright, achievements appraisal and sustainable development questions and so on. Finally, take Google digital library project as the example, the paper analyze its strategy, the technical methods, the cooperation pattern, the service functions which Google adopts in the resources and services integration process and its latent risks.

The paper is composed of 8 chapters, including the introduction, the main text and the conclusion, and the main content is as following:

0 Introduction

In this chapter, the article introduces the research background and the selected topic significance, points out that, just because the future digital libraries will cooperate into an interconnecting knowledge environment as well as the distributed environment will be the digital library system background and the mainstream environment in the future, the resources aggregation and the services integration research have stronger practical significance and impendence. From the fundamental research, the technology and the

realization methods, the management research and the example investigation four aspects, the chapter reviews the related Chinese and foreign literature on this selected topic, simply introduces this article research contents, the research techniques and the characteristic places.

1 Outline

In this chapter, the paper discusses the three moves of the resources aggregation and services integration of distributed digital library, including that it's an inevitable result when digital library develops into the second-generation, it's an inevitably choice what the digital library should answer to the user information utilization behavior change, and it's the urgent need if the digital library want to alleviate "the data isolated island" phenomenon in the distributed information environment.

This chapter analyzes the resources aggregation and the services integration meanings, elaborates the relation between the digital resource aggregation and the services integration, points out that they interact with each other in fact, thus explains why this paper studies these two questions together and accounts for the foundation and the significance of this topic.

2 The resource aggregation and service integration goals and functions of distributed digital libraries

In this chapter, the paper separately elaborates the resources aggregation and the services integration goals of the distributed digital libraries, and emphatically analyzes the functions which the system has obtained after the resources and services integration, such as the gateway, the unification authentication, the unification retrieval, the jurisdiction management, the individuality services and the digital reference services and so on. This chapter introduces some correlated cases, realizes some advantages and disadvantages when these functions are carried on.

3 The resource aggregation and service integration principles and methods of distributed digital libraries

In this chapter, the paper analyzes the resources and services integration principles from the different planes, including the 7 principles in overall plane,

the 5 principles in operation plane, etc. The chapter discusses the concrete methods of the resources aggregation from two aspects of the resources integration breadth and depth, and analyzes how these methods are utilized in the concrete digital library resources and service integration process.

The main technologies which are used in the present resources and services integration have be introduced in this chapter, such as ingesting, harvesting, RSS, OAI, the Web crawling and Metasearching and soon.

4 The resource aggregation and service integration patterns and environment of distributed digital libraries

The studies on the patterns and environment of the resources and services of the distributed digital libraries include how to choose the cooperation patterns, what digital resources are target objects that should be aggregated and so on. This chapter unifies the domestic and foreign well-known resources aggregation and the service integration examples, from the resources aggregation and the services integration's objects, the functions, the scopes and the main bodies, as well as its patterns and the environment. Finally, the paper pointed out, the resource aggregation and the service integration patterns and the environment are presenting the multilateral cooperation and the digital collections aggregation have become the chief tendency.

5 The resource aggregation and service integration challenge and countermeasure of distributed digital libraries

This chapter is the key chapter of this article. Based on the above elaborated researches, the paper emphatically probes into the management questions. Mainly revolves the strategy choices, the standards and criterions, the quality-control, the copyright questions, the achievements evaluations and the sustainable developments and so on. Those six resource aggregation and service integration management questions have been carried on the detailed analysis and discussion in this chapter, then points out the present research situations, faced questions, the solutions and measures of each aspect.

6 The case study on the resource aggregation and service integration of Google digital library

In this chapter, this paper takes the famous searchers—Google as an

example, although Google digital library has been familiar with users, it has been bearing disputations and rebukes since its appearance. This chapter reviews Google digital library's resource aggregation and service integration course, and analyzes its functions, the cooperation patterns, the technical methods, the profit patterns and the latent risks and so on. All these researches are carried on the viewpoint foundations that have been discussed in the foregoing article.

7 Conclusions

In this chapter, the paper, on the basis of summarizing the contents of the whole research and pointing out the limitation in the research, emphasizes once more the significance of study on the resources aggregation and services integration of distributed digital libraries. Reviews and concludes several essential factors about the resources aggregation and service integration management aspects which proposed in the foregoing article, emphatically analyzes the correlations between these factors. In addition, puts forwards the future study planning to perfect the research.

Keywords: distributed digital library resource aggregation service integration management

目 录

Contents

0 绪 论	1
0.1 研究背景及选题意义	1
0.1.1 未来的数字图书馆是合作互联的知识环境	1
0.1.2 分布式环境将是未来数字图书馆系统的主流环境	3
0.1.3 分布式数字图书馆亟需资源整合与服务集成	4
0.2 研究现状	4
0.2.1 分布式数字图书馆资源整合与服务集成的理论研究	5
0.2.2 分布式数字图书馆资源整合与服务集成的技术和实现 方法	6
0.2.3 分布式数字图书馆资源整合与服务集成的管理研究	11
0.2.4 国内外一些分布式数字图书馆资源整合与服务集成 实例	16
0.3 研究内容与研究方法	17
0.3.1 研究内容	17
0.3.2 研究方法	19
0.3.3 研究特色	20
1 概 述	22
1.1 分布式数字图书馆资源整合与服务集成的动因	22
1.1.1 数字图书馆发展到第二代后的必然结果	22
1.1.2 数字图书馆应对用户信息使用行为变化后的必然选择	24
1.1.3 数字图书馆缓解分布信息环境下“数据孤岛”现象的迫切 要求	25
1.2 分布式数字图书馆资源整合的定义	27
1.3 分布式数字图书馆服务集成的定义	28
1.4 分布式数字图书馆资源整合与服务集成的相互关系	30

2 分布式数字图书馆资源整合与服务集成的目标和功能	32
2.1 分布式数字图书馆资源整合与服务集成的目标	32
2.2 分布式数字图书馆资源整合与服务集成的功能	33
2.2.1 数字图书馆门户	34
2.2.2 统一认证	35
2.2.3 统一检索	37
2.2.4 数字权限管理	39
2.2.5 个性化服务	40
2.2.6 合作数字参考咨询	42
3 分布式数字图书馆资源整合与服务集成的原则和方法	45
3.1 分布式数字图书馆资源整合与服务集成的原则	45
3.2 分布式数字图书馆资源整合与服务集成的方法	48
3.2.1 资源整合的广度	48
3.2.2 资源整合的深度	52
3.2.3 资源整合与服务集成方法的应用	56
3.3 分布式数字图书馆资源整合与服务集成的技术运用	59
3.3.1 摄取(Ingesting)	59
3.3.2 OAI	60
3.3.3 收割(Harvesting)	60
3.3.4 RSS	61
3.3.5 网页采集(Web Crawling)	62
3.3.6 元搜索(Metasearching)	63
4 分布式数字图书馆资源整合与服务集成的模式和环境	65
4.1 整合与集成的对象	65
4.1.1 围绕某一主题的整合与集成	65
4.1.2 围绕某一资源类型的整合与集成	67
4.1.3 综合性的内容整合与集成	69
4.2 整合与集成的功能	71
4.2.1 基于技术试验型的整合与集成	71
4.2.2 基于资源保存型的整合与集成	72

4.2.3	应用服务型的整合与集成	74
4.2.4	商业利益型的整合与集成	75
4.3	整合与集成的范围	76
4.3.1	馆内整合	76
4.3.2	区域整合	77
4.3.3	国家级整合	78
4.3.4	跨国间的整合与集成	79
4.4	整合与集成的主体	81
4.4.1	以数字图书馆联盟为主体推动的整合与集成	81
4.4.2	以政府部门为主体进行的整合与集成	82
4.4.3	以文献信息等机构为主体开展的整合与集成	83
4.4.4	各种基金会或行业组织发起或开展的整合	84
4.4.5	以数据库商或出版商为主体进行的整合与集成	85
4.5	资源整合与服务集成模式和环境的发展趋势	86
4.5.1	多边合作成为资源整合与服务集成的主要方式	87
4.5.2	资源集合级整合正成为主流	90
5	分布式数字图书馆资源整合与服务集成的挑战 and 对策	95
5.1	策略选择	96
5.1.1	整合什么(what)	96
5.1.2	由谁整合(who)	97
5.1.3	为谁整合(whom)	100
5.1.4	如何整合(how)	101
5.2	标准规范	102
5.2.1	数字图书馆标准规范的重要性	102
5.2.2	数字图书馆标准规范的主要内容	103
5.2.3	数字图书馆标准规范的研究进展	104
5.3	质量控制	105
5.3.1	质量前控方式	106
5.3.2	质量后控方式	107
5.4	版权问题	109
5.4.1	国际图书馆员协会联合会(IFLA)关于数字环境中 版权问题的立场	110

5.4.2	国家版权局网络传播权保护条例	111
5.4.3	分布式数字图书馆资源整合与服务集成的版权过滤	112
5.5	绩效评估	115
5.5.1	建立自我评价机制	115
5.5.2	确立绩效评价的指标体系	116
5.5.3	对整合与集成活动进行全程监控和评估	117
5.6	可持续发展	118
5.6.1	积极的沟通与交流是必要手段	119
5.6.2	牢固的协议和诚信是重要保证	120
5.6.3	持续的经费支持是必要条件	121
5.6.4	合理的利益分配是驱动力	121
6	分布式数字图书馆资源整合与服务集成的案例分析	123
6.1	Google 的影响	123
6.2	Google 的功能分析	125
6.3	Google 的合作模式	128
6.4	Google 的技术方法	129
6.4.1	Google 与图书馆资源整合的实现方式	129
6.4.2	Google 与出版商资源整合的实现方式	130
6.4.3	Google 搜索结果的排列算法	130
6.5	Google 的赢利模式	131
6.6	Google 的风险分析	132
6.6.1	版权纠纷	132
6.6.2	Google 的内容缺陷	134
6.6.3	文化入侵	134
6.6.4	可持续发展	136
7	结 语	138
	参考文献	141
	后 记	159