



# 数字图书馆服务链

——服务模式·体系架构·关键技术

黄梦醒 著



清华大学出版社

# 数字图书馆服务链

——服务模式·体系架构·关键技术

黄梦醒 著



清华大学出版社  
北京

## 内 容 简 介

数字图书馆 (Digital Library) 与服务链管理 (Service Chain Management) 两个领域的交叉融合形成了数字图书馆服务链这一新的研究方向。本书是国家 863 计划、国家“十一五”科技支撑计划和国家自然科学基金资助项目的部分研究成果,也是国内第一本关于数字图书馆服务链模式与关键技术研究的专著。全书从服务模式、体系架构与关键技术等方面介绍作者在该领域所做工作及取得的阶段性成果。全书共分 10 章,第 1~2 章介绍了国内外数字图书馆及其服务模式的发展,以及服务链(供应链)管理的最新进展;第 3 章提出了一种新的数字图书馆内容管理与服务模式——数字图书馆服务链,构建了数字图书馆服务链的体系结构和信息服务模式;第 4~7 章研究了基于 SOA、网格计算、云计算的数字图书馆服务链集成与协同体系架构;第 8~10 章研究了云计算环境下数字图书馆服务链信任机制、任务调度和副本管理等核心关键技术。

本书可供高等学校计算机、图书情报、管理科学与工程等相关专业教师和研究生阅读,也可作为相关领域科技工作者的参考书。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。  
版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

### 图书在版编目(CIP)数据

数字图书馆服务链:服务模式·体系架构·关键技术/黄梦醒著.—北京:清华大学出版社,2013

ISBN 978-7-302-33074-5

I. ①数… II. ①黄… III. ①数字图书馆-图书馆服务-研究 IV. ①G250.76

中国版本图书馆 CIP 数据核字(2013)第 151111 号

责任编辑:李 鹏  
封面设计:傅瑞学  
责任校对:焦丽丽  
责任印制:宋 林

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, [c-service@tup.tsinghua.edu.cn](mailto:c-service@tup.tsinghua.edu.cn)

质 量 反 馈:010-62772015, [zhiliang@tup.tsinghua.edu.cn](mailto:zhiliang@tup.tsinghua.edu.cn)

课 件 下 载: <http://www.tup.com.cn>, 010-62795954

印 刷 者:三河市君旺印装厂

装 订 者:三河市新茂装订有限公司

经 销:全国新华书店

开 本:170mm×240mm 印 张:12.25 字 数:213 千字

版 次:2013 年 9 月第 1 版 印 次:2013 年 9 月第 1 次印刷

印 数:1~1500

定 价:29.00 元

# 前 言

随着信息技术的发展和网络应用的普及，人们主要通过互联网来获取需要的信息和知识。数字图书馆是一种基于网络环境下共建共享的可扩展的知识网络体系，是人们获取信息、知识服务和提高国民文化素养的公共服务机构。所以，数字图书馆自 20 世纪 90 年代初一经提出就受到各个国家的重视，从而得到大力发展。但是，随着搜索引擎，特别是 Google 的发展使图书馆面临着被边缘化的威胁。为了应对这种威胁和挑战，世界上许多国家和地区，如美国、欧洲等开始投入研究和构建先进的数字图书馆管理模式——构建模式、服务和运营模式，这些模式都试图借鉴 Google 等搜索引擎的优点并弥补它们存在的不足，从而实现数字图书馆的文化价值和社会价值。我国作为一个文化大国，随着新兴信息技术的发展，研究数字图书馆服务链运营管理模式及其体系结构，构建一种先进的数字内容管理及服务模式及其体系结构，将图书馆的优势与 Google 等搜索引擎的优势相结合，给用户高效、高质量、个性化的知识和信息服务，将能有效地应对 Google 等搜索引擎带来的挑战和文化入侵，对我国数字图书馆的建设和发展具有战略意义。

本人于 2003 年开始从事供应链协同机制研究，并完成了我的博士论文；2007 年进入清华大学信息科学技术研究院从事计算机科学与技术专业博士后研究工作，主要从事数字平面内容管理与服务的理论与技术研究。在清华大学的两年间，参与了国家“十一五”科技支撑计划项目（项目编号：2006BAH02A12）和国家 863 计划项目（项目编号：2009AA01Z143）的研究工作，在研究中将供应链管理理论与数字图书馆技术相结合提出了数字图书馆服务链这一新的数字图书馆内容管理与服务模式。2009 年 10 月我博士后出站后进入海南大学信息科学技术学院任教，继续开展基于云计算的数字图书馆服务链理论与关键技术研究，研究工作得到国家自然科学基金的支持（项目编号：71161007）。本书是在这些研究的基础上修改、充实而写成的，同时借鉴吸收国内外同类研究的最新成果。

数字图书馆服务链技术研究是一个新课题，就愚见所及，本书可能是国内该研究领域第一本公开出版的专著。由于论题涉及数字图书馆和服务链管理这一交叉领域，研究内容体现了多学科交叉和技术集成创新的特点，并且涉及计算机技

术研究的一些前沿问题，这些都决定了本书具有明显的探索性，稚嫩在所难免。愿借此书抛砖引玉，与同行进行交流。由于本人学识与能力有限，从事这一新领域研究实践比较短，加之新兴信息技术发展很快，很多新东西我并不是很熟悉，在很多问题上只能是一孔之见，书中不妥和错误之处在所难免，恳请各位专家和读者朋友不吝指教和帮助！

感谢我的老师、朋友、同事和学生的支持和帮助，他们是：我的合作导师邢春晓教授，清华大学信息技术研究院 Web 与软件技术研究中的张勇副研究员、杨吉江副研究员、李超博士、戴桂兰博士、朱义老师、郭涑伟老师等，海南大学冯文龙副教授，我的硕士研究生王磊、叶祥龙、杨倩茹等。

清华大学出版社做了大量艰苦细致的工作，付出了辛勤的劳动，正是其敬业与认真促成了本书的迅速付梓，在此特表示衷心的感谢！

黄梦醒

2013年5月

于海南省海甸岛

# 目 录

第 1 章 绪论	1
1.1 研究背景与意义	1
1.2 国内外数字图书馆发展现状	2
1.2.1 数字图书馆的发展历程	2
1.2.2 Google 网络数字图书馆计划	5
1.2.3 欧洲数字图书馆联盟计划	7
1.2.4 我国数字图书馆建设与运营模式	8
1.3 研究框架	9
第 2 章 服务链与服务链管理	11
2.1 服务链与供应链	11
2.1.1 供应链的定义	12
2.1.2 供应链的主要特征	13
2.1.3 供应链的分类	14
2.2 供应链管理	16
2.2.1 供应链管理的定义	16
2.2.2 供应链管理的关键要素	17
2.2.3 供应链管理的关键业务过程	18
2.3 供应链协调理论	20
2.3.1 供应链协调的概念	20
2.3.2 供应链协调的类型	22
2.3.3 供应链协调模式与机制	24
2.3.4 供应链协调的主要方法	26
2.4 本章小结	31
第 3 章 数字图书馆服务链模式	32
3.1 引言	32
3.2 数字图书馆的服务链架构	33
3.2.1 数字图书馆服务链构成节点分析	34
3.2.2 数字图书馆服务链的节点关系分析	35
3.3 数字图书馆服务链的服务模式	36
3.3.1 大众化信息服务模式	37

3.3.2	专业化信息服务模式 .....	38
3.3.3	个性化信息服务模式 .....	39
3.4	本章小结 .....	40
<b>第 4 章</b>	<b>基于 SOA 的数字图书馆服务链集成 .....</b>	<b>41</b>
4.1	引言 .....	41
4.2	面向服务的体系结构——SOA .....	43
4.3	基于 SOA 的数字图书馆服务链集成框架 .....	48
4.4	基于 SOA 的数字图书馆服务链系统架构 .....	49
4.5	服务链集成架构中服务请求订单驱动的执行流程应用 .....	50
4.6	本章小结 .....	53
<b>第 5 章</b>	<b>基于网格计算的数字图书馆联盟协同服务架构 .....</b>	<b>54</b>
5.1	引言 .....	54
5.2	网格计算及其应用 .....	54
5.3	数字图书馆联盟的内容服务网格体系结构 .....	57
5.3.1	数字图书馆联盟内容服务网格体系结构模型 .....	57
5.3.2	数字图书馆联盟内容服务网格体系结构 .....	59
5.4	网格环境下数字图书馆联盟协同服务框架及其应用 .....	59
5.4.1	基于资源共享的数字图书馆联盟网格的协同服务框架 .....	59
5.4.2	内容检索服务过程示例 .....	61
5.5	本章小结 .....	62
<b>第 6 章</b>	<b>基于云计算的数字图书馆服务链 .....</b>	<b>63</b>
6.1	引言 .....	63
6.2	云计算 .....	63
6.2.1	云计算基本概念 .....	63
6.2.2	云计算的主要特征 .....	64
6.2.3	云计算的服务模式 .....	65
6.2.4	云计算体系架构 .....	66
6.3	基于云计算的数字图书馆服务链体系架构 .....	67
6.4	云环境下基于资源共享的协作服务框架 .....	70
6.5	内容检索过程示例 .....	71
6.6	本章小结 .....	72
<b>第 7 章</b>	<b>数字图书馆数字内容管理与服务体系结构 .....</b>	<b>73</b>
7.1	引言 .....	73

7.2	研究综述.....	74
7.3	数字图书馆数字内容管理与服务体系结构设计与实现.....	81
7.3.1	设计原则.....	82
7.3.2	数字对象分析.....	82
7.3.3	总体功能描述.....	83
7.3.4	数据处理流程.....	86
7.3.5	网络拓扑与部署.....	87
7.3.6	核心服务构件.....	88
7.4	本章小结.....	90
<b>第 8 章</b>	<b>云计算环境下数字图书馆服务链信任机制.....</b>	<b>91</b>
8.1	概述.....	91
8.2	云计算环境下服务链信任模型.....	93
8.2.1	云计算环境下服务链信任模型特点.....	93
8.2.2	信任模型设计.....	93
8.3	基于灰色 AHP 的内容服务提供方信任评估模型.....	94
8.3.1	层次分析法概述及基本步骤.....	95
8.3.2	灰色系统理论.....	96
8.3.3	基于灰色 AHP 的数字图书馆服务链信任模型.....	96
8.3.4	实例分析.....	101
8.4	基于双滑动窗口的数字图书馆服务链信任评估机制.....	104
8.4.1	信任度评估原则.....	104
8.4.2	基于双滑动窗口的信任度评估机制.....	104
8.4.3	实验仿真分析.....	107
8.5	本章小结.....	108
<b>第 9 章</b>	<b>云计算环境下数字图书馆服务链作业调度算法.....</b>	<b>109</b>
9.1	概述.....	109
9.1.1	Hadoop 平台.....	109
9.1.2	作业调度算法.....	121
9.2	Hadoop 作业调度算法分析.....	125
9.2.1	Hadoop 调度算法思想分析.....	125
9.2.2	Hadoop 调度算法性能实验分析.....	130
9.3	一种基于内存平衡的 Hadoop 平台公平调度算法.....	141
9.3.1	FMScheduler 调度算法的设计.....	142
9.3.2	FMScheduler 的实现.....	145



9.3.3	FMScheduler 性能实验及实验结果分析 .....	147
9.4	本章小结 .....	150
<b>第 10 章</b>	<b>云计算环境下数字图书馆服务链数字内容副本管理策略 .....</b>	<b>151</b>
10.1	云存储与副本管理 .....	151
10.1.1	云存储系统结构模型 .....	151
10.1.2	副本管理研究现状 .....	153
10.2	静态副本创建策略 .....	157
10.2.1	HDFS 及其静态副本创建策略 .....	157
10.2.2	数据节点选择策略改进 .....	161
10.2.3	仿真实验与结果分析 .....	163
10.3	动态副本创建策略 .....	165
10.3.1	云存储副本创建模型 .....	165
10.3.2	动态副本创建策略 .....	166
10.3.3	仿真实验与结果分析 .....	169
10.4	本章小结 .....	171
	结束语 .....	173
	参考文献 .....	174

# 第1章 绪论

## 1.1 研究背景与意义

当前，传统的中华文化遗产的长期保存和利用方式已远不能满足当今发展的需要。中国文化产业的发展，要走以数字化、技术化带动的跨越式发展道路。利用信息技术与因特网等先进手段，实现中华文化遗产的长期保存和高效利用，已成为急需研究的课题。

近几年来，数字图书馆相关技术的发展为中华文化的数字化利用奠定了良好的基础。中华文化经过数字化加工和处理，提供给使用者的内容不仅包括原有的文献，也包括知识的组织和体现。通过数字图书馆的方式，包括数学资源在内的各门类的文献资源和学科知识将更便于交流和利用，更有助于学习和研究。研究成果将会根本改变中国文化信息资源保存、管理、传播、使用的方式和手段，克服中国文化信息资源得不到有效利用和共享的弊病，为知识创新和两个文明建设营造一个汲取文化信息的良好环境。特别对中小学教育以及因信息不畅和文化落后而导致经济不发达的老、少、边、穷地区的发展，将产生巨大的推动作用。

随着信息技术的发展和网络应用的普及，人们主要通过互联网来获取需要的信息和知识。数字图书馆是一个基于网络环境下共建共享的可扩展的知识网络体系，是人们获取信息、知识服务和提高国民文化素养的公共服务机构。所以，数字图书馆自上世纪90年代初一经提出就受到各个国家的重视，从而得到大力发展。但是，随着搜索引擎的发展，特别是Google的发展，使图书馆面临着被边缘化的威胁。为了应对这种威胁和挑战，世界上许多国家和地区，如美国、欧洲等开始投入研究和构建先进的数字图书馆管理模式——构建模式、服务和运营模式，这些模式都试图借鉴 Google 等搜索引擎的优点并弥补它们所存在的不足，从而实现数字图书馆的文化和社会价值。

数字媒体内容运营是下一代网络（NGN）中最重要的业务，也是现代服务业的典型代表。世界各国都非常重视数字媒体的建设。而行业要发展，必须要有政

府的指引和扶持。韩国正是利用IT转型的机遇，大力发展网络游戏，成为网络游戏后起之秀，其收入占世界软件总产值的2.69%，超过中国软件业总产值。在数字媒体行业，中国是高瞻远瞩。要使数字媒体行业迅速、健康发展，首先要建立起规范和标准，这样才能不受制于外国技术和标准，企业间才得以公平竞争、互惠互利、共同发展。研究我国数字图书馆的运营模式及数字媒体内容支撑平台，对于打破国外数字资源管理软件平台的技术壁垒和商业垄断，推动我国海量数字媒体内容管理在长期保存中华文化方面向深层次迈进具有重要意义和深远影响。

另外，现代服务业分工明细，在数字媒体产业链上，内容集成商、内容分发商、内容运营商、技术服务商各司其职。但是，由于链条的各个环节没有协调的运作模式和统一的接口规范，使得这条产业链成为“单链”，下游厂家对上游厂家具有相当的依赖性。这种现状大大增加了厂家的风险，一旦上游厂家出现问题，下游厂家将受到很大冲击，并必须花费相当的投入重建链条；反之，下游厂家也一样会影响上游厂家。因此，研究数字图书馆服务链运营管理模式及其体系结构，构建一种先进的数字内容管理及服务模式及其体系结构将图书馆的优势与Google等搜索引擎的优势相结合，给用户高效、高质量、个性化的知识和信息服务，将能有效的应对Google等搜索引擎带来的挑战和文化入侵，对我国数字图书馆的建设和发展具有战略意义。

## 1.2 国内外数字图书馆发展现状

### 1.2.1 数字图书馆的发展历程<sup>[1, 2]</sup>

在1993年9月~2005年4月的十多年间，数字图书馆的建设经历了研究、规划、试验到起步，各馆建立各自具有馆藏特色的数据库、网站，再到全球数字图书馆的筹建发展过程，大致可划分为三个阶段。

第一阶段是研究、试验、规划阶段。1993年9月，美国国家科学基金会（NSF）、美国国家宇航局（NASA）、美国国防部高级研究计划署（ARPA）联合提出“数字图书馆倡议”，这是数字图书馆的起源。1994年美国就相继出现众多的有关数字图书馆专家研究小组、小组会及正式研究会等。1994年6月19~21日，在美国召开了“第一届数字图书馆理论与实践研究第一次国际会议”。1994年9月，美国国家科学基金会、美国国家宇航局、美国国防部高级研究计划署正式决定投资2400万美元给伊里

诺伊大学厄巴纳—香槟分校 (University of Illinois at Urbana-Champaign)、密歇根大学安阿伯分校 (University of Michigan at Ann Arbor)、斯坦福大学 (Stanford University)、加利福尼亚大学圣巴巴拉分校 (University of California at Barbara)、卡内基梅隆大学 (Carnegie Mellon University) 等6所大学进行数字图书馆的分项研究。1994年10月, 美国国会图书馆宣布, 它将迈向数字化时代, 到公元2000年美国国会图书馆成立200周年时, 将完成该馆最重要的图书资料的数字转化, 当时已有20多万页图书资料实现了数字化。从此, 数字图书馆的研究、试验、规划在全世界广泛开展, 并不断深入。数字图书馆建设发展中的一些技术难题, 以及有关数字图书馆建设的部署、起步、经费来源、人才培养等问题逐步得到解决。1995年5月1日, 美国国会图书馆、美国图书保护及存取委员会、哥伦比亚大学、卡纳尔大学、埃莫里大学、哈佛大学、国家档案及记录管理委员会、纽约公共图书馆、宾夕法尼亚州立大学、普林斯顿大学、密歇根大学、南加州大学、田纳西大学、耶鲁大学等在哈佛大学共同签订了“全国数字化图书馆联合协议”。这些机构原则上同意将国内外的数字化材料集中管理, 便于用户在任何地方利用它们。对有独特收藏的小图书馆, 虽然没有财力开发数字化工程, 亦不排除在全国数字图书馆之外, 它们亦可加入。按照计划, 美国国会图书馆带头协调所有数字化图书馆在技术和政策方面的行动, 以便把它们连接到同一计算机网络上。尽管这一计划行动缓慢, 没有达到预期的结果, 但毕竟是数字图书馆的发展方向之一。

第二阶段是起步与应用阶段。这一阶段有商业性数字图书馆和公共图书馆开发的数字图书馆。公共图书馆开发的数字图书馆主要是在一馆内, 将本馆具有特色、没有版权问题的珍本、孤本、稀有本, 或有某些特色的文献资源进行扫描, 建成数据库、网站, 通过网络供读者使用。它以物理实体的传统图书馆为基础, 并依附于它, 将馆藏文献的书目、文摘或某些具有特色的馆藏历史文献、档案资料、稀有珍贵图书进行扫描, 建立数据库, 并不断采购新的电子文献进行补充。它是传统图书馆与数字图书馆的有机结合, 人们称它为“复合图书馆”或“混合图书馆”。这也是第二阶段数字图书馆存在的一种主要模式或基本模式。这种模式要求不很高, 只要有一定基础的馆藏、技术条件、经济实力, 就可建立一批规模大小不等的数字图书馆。由于复合图书馆建立的条件比较现实可行, 因此从1995年以来这种模式的数字图书馆发展比较快, 比较普遍。但目前仍限于国家图书馆、大型公共图书馆、高等学校图书馆、科研机构、专业图书馆等。如美国华盛顿大学的数字图书馆, 1995年平均每天就有45000名以上的读者上网查询。美国国会图

图书馆的“国家数字图书馆”项目“美国的记忆”，从1996年到2000年的5年时间，耗资6000万美元，将馆藏500万册（件）文献转换成数字化信息，并通过网络向外界发布。我们知道，美国国会图书馆2001年馆藏已达1.25亿册（件），每年平均增长300万册，而仅将500万册转化为数字信息就耗时5年，可谓工程浩大，耗资巨大。美国纽约公共图书馆的数字图书馆，仅2004年就有来自200个不同国家和地区的人登录该馆网站，一年的点击率达到75000万人次，而且该馆从来没有为网站做任何市场公关或广告活动，它的联机图片库含数万取自图书、报刊与照片、印刷品、明信片的数字化图像，被评为2004年“最佳免费参考网站”。美国密歇根大学图书馆这些年来数字化图书率在美国图书馆界名列前茅。《华盛顿邮报》报道：2001年美国加利福尼亚的公共图书馆和大约1800所遍布美国的其他图书馆，正式试验进行网络图书馆服务。20世纪末，日本国会图书馆的数字图书馆，已完成的数据总量仅图像数据就达1000多万页，存储在3000张光盘中。英国图书馆通过“网络OPAC”可提供600万条数据。法国国家图书馆的资源数字项目已取得阶段性成果，包括1万本图书的全文文献、10万本图书的扫描影像、20万张图片和3000件录音资料。

由中国国家图书馆牵头的“中国数字图书馆工程”，于2000年6月经国务院批准启动。中国国家图书馆于1999年3月成立了文献数字化中心，每天可加工20万页全文影像数据。中国国家图书馆网站仅在2000年一年内网上点击达1.4亿人次。据2000年统计，中国已有40多所公共图书馆和171所高校图书馆建立了自己的数据库和网站，开展网络服务。上海交通大学图书馆拥有数字化馆藏300GB以上，已有25万种馆藏文献数字化。清华大学图书馆与IBM合作建立了具有中文全文检索功能的“中国高校学位论文联机服务系统”，收录15所高校的博士硕士学位论文，用户可在国内15所联网高校和互联网上进行检索。2003年10月，经国家新闻出版总署批准，中国学术期刊（光盘）电子杂志社成立了互联网出版分社，并在清华大学启动了《中国知识资源总库》建设工程。到2005年初，授权“中国知网”（www.CNKI.NET）接入“CNKI网络资源共享平台”的国内数据库已有1100多种，其中包括科学数据、题录摘要、全文、专题等各类数据库，文献信息总量已达6300多万条。互联网出版分社已经获得授权，为各类数字图书馆用户接入这些数据库，丰富和扩充用户单位数字图书馆的知识资源。商业性数字图书馆，如超星等，也发展很快。

由以上看来，数字图书馆起步与应用阶段的十多年里，虽然取得一定成绩与

效益,但仍存在以下问题:一是发展不平衡,不仅馆与馆、类型与类型发展不平衡,而且国与国、地区与地区之间的发展也不平衡;二是数字化资源有限,与传统图书馆的资源存在很大距离;三是效益有限,由于数字化资源的限制,影响了使用与传播,影响了用户与读者的上网积极性与点击率。产生上述问题的原因主要有:一是知识版权的制约;二是经费的制约;三是观念的制约,把发展数字图书馆仅限于国家拨款及一馆之内,没有充分发挥社会、市场、企业的作用;四是只想共享,不想共建,谈共享积极性高,谈共建设没有干劲。2005年3月4日,一位在美国的华人学者木木,在网上发表《数字图书馆》一文,他说,美国虽有普遍发达的公共图书馆设施,却没有办法找到一个完善的数字图书馆。原因何在?主要是缺乏利益驱动,这样浩大的工程没有国家政府或一个资金雄厚的组织资助;其次是版权问题,西方为了保护作者利益的版权法,在实际运作中已成为文化交流,乃至科研工作的巨大障碍。事实上,各行各业的大多数作者们很少有人因为任何版权法而变为富人,版权法更多的是在保护出版商的利益。他认为,数字图书馆如能实现,在西方也需要很长时间。

第三阶段是规模扩大阶段。数字图书馆经过传统图书馆与数字图书馆相互结合、共存互补的复合图书馆建设之后,已进入数字图书馆建设的第三个阶段。这一阶段的数字图书馆模式新、规模大、资源丰富、技术更先进、服务面广、受益者多、影响深、竞争激烈。

## 1.2.2 Google 网络数字图书馆计划<sup>[3, 4]</sup>

目前全球最大的网络搜索引擎美国Google公司数据库中的网页已达80亿页,每天有亿万用户检索使用,其图书资料已获得广告收益。Google奠基人Sergey Berin和Larry Page从创办该搜索引擎开始,就有数字图书馆计划的设想,发誓将世界上的所有图书、文献、资料通过网络浏览器向所有人提供服务,并呼吁图书馆积极参与,加入Google网络服务,为Google提供馆藏。

在经过一年多时间与世界5个著名图书馆协商、谈判和签约,并在密歇根大学图书馆做了一段时间的图书无损快速扫描试验之后,Google终于在2004年底推出筹建全球最大数字图书馆的计划,并于12月14日在加州正式宣布:该公司已与英、美两国5个世界著名的图书馆达成协议,共同筹建全球最大的网络图书馆。协议规定:通过公司与图书馆的合作,将这些图书馆的馆藏图书扫描制作成电子版,放

到网上，使全世界用户可以免费在线查询和使用这些文献。预计到2015年工程完工时，Google将建成全球最大的网络图书馆。这可能是向全球网络图书馆迈出的第一步，是向世界的挑战。

与Google合作筹建全球最大网络图书馆的成员是美国哈佛大学图书馆、密歇根大学图书馆、斯坦福大学图书馆、纽约公共图书馆、英国牛津大学图书馆。牛津大学具体是博德利中心图书馆。2005年初，又传出普渡大学图书馆也将加入这一行列。这些图书馆图书资源丰富，5馆所藏图书接近6000万册。其中纽约公共图书馆2000万册、哈佛大学图书馆1500万册、斯坦福大学图书馆800万册、密歇根大学图书馆700万册、牛津大学图书馆750万册。由于版权限制及担心图书受损等原因，只有斯坦福大学图书馆和密歇根大学图书馆同意将全部馆藏图书交给Google扫描，制作成电子版。其他3所图书馆则各有一定保留，如纽约公共图书馆只同意把没有版权限制图书中的一部分制作成电子版，如一些易碎的、珍贵的、学术性的馆藏；哈佛大学图书馆打算第一期计划是让Google扫描其馆藏图书中的4万册，然后根据结果，决定是否提供更多图书。牛津大学图书馆在合约中规定：对1900年之前出版的馆藏图书限量扫描、转换成电子版；在今后三年内数字化100万册以上失去版权保护的旧书，但不包含牛津大学大量独特或特别珍贵的研究资料（手稿、档案、地图或早期印刷图书），如中世纪（14至17世纪）的珍贵书籍，美国内战时期（1861至1865年）的稀有书籍和达尔文等名人的手稿、档案等。

其合作与运转方式为：由上述图书馆提供馆藏图书；Google位于加利福尼亚山景市的总部（the Mountain View, Calif.）负责这些馆藏的扫描、数字化转换工作。斯坦福大学图书馆离Google总部很近，因此该馆馆藏将采用运输方式，在Google复制中心扫描后，再送回图书馆。对密歇根和哈佛大学图书馆馆藏将采用远程扫描方式进行数字化转换。

将上述5馆馆藏的全部或部分制成电子版后，Google允许网上用户阅读所有无版权限制图书的全部内容，但剪贴、复印、打印将被限制，而那些有版权保护的图书，用户仍只能在网上阅读目录及内容摘要，并可依靠这些信息做出决定是否购买印刷文献。这所网络图书馆将是一所公益性的网络图书馆，不具盈利性，全球各地的读者都可免费查询、阅读。

Google此项数字化计划打造的网络图书馆为Google Scholar，起源于Google Print项目，是它的延伸。早在2003年10月，著名网上图书零售商亚马逊推出基于图书全文的“书内搜索”，引起很大反响。Google于12月就推出Google Print试用

版,是一种书摘搜索,与亚马逊的全文搜索差距较大。2004年10月与美国著名出版商兰登书屋合作后,范围内容扩大,Google开始索引图书全文。但在图书馆数字化计划宣布前,一直声明暂无与图书馆合作的计划。由于在版图书所占图书比例有限,因此Google Print的目光早就瞄向了图书馆的丰富馆藏。

Google Print项目分为出版商和图书馆两部分内容。该项目目标是使用户能够通过互联网查询和检索图书。Google Print主页开宗明义,“Google的使命是组织全世界的信息,使之可广泛存取与使用。”当用户输入查询术语后,可以找到与其要求相匹配的图书及其书评。用户可以通过“购买此书”链接进入网上书店,或从最近的图书馆找到此书。对没有版权保护的公共领域作品,用户可以直接在线阅读,版权作品只提供部分内容或书目数据。该项目的图书资源来自出版商和图书馆两方面。出版商可将书名放在Google Print上,通过免费链接扩大图书市场,吸引新的读者。采取该措施可以获得广告收益,用户点击书名后,Google给出版商付点击费。Google在同5个著名图书馆合作后,可将其馆藏数字化,通过Google Print为用户提供查询服务。Google称其目的是维护图书和图书馆在互联网中的地位,使图书馆的“离线”图书成为“在线”资源。通过Google吸引更多用户使用图书馆,使一些绝版图书在数字化后获得重生。Google将Google Print向图书馆的扩展称之为“Google出借图书馆藏书”。

Google Scholar和Google Print是两个不同的产品,作用不同。Scholar是一个查询工具,主要用于学术文献查询。Google的主索引能够发现文献,但不能获取在线文献。有些图书没有集成到Google Print中,因此将通过这次的Google数字化扫描计划解决这一问题。

Google的数字图书馆计划引起了世界图书馆界的强烈反响,甚至引起现有数字图书馆的恐慌,其原因是Google数字图书馆计划实现的可能性极大,一旦成功,将强有力地冲击现有数字图书馆,因为它的资源、经费、技术、免费使用都是现有数字图书馆无法相比的。

### 1.2.3 欧洲数字图书馆联盟计划<sup>[5]</sup>

自Google筹建全球最大网络图书馆的消息宣布后,引起世界的广泛关注,特别在欧洲首先引起法国国家图书馆的不安。为了与美国Google搜索引擎公司全力打造的全球最大网络图书馆抗衡,为了对Google的资源垄断进行“欧洲人的反击”,



并让人们了解欧洲人的智慧、历史及文化遗产，2005年3月16日，法国总统府宣布：法国总统希拉克将向欧盟所有成员国提议，加速数字化欧洲图书馆的建设。法国总统府在新闻公报中表示，加速欧洲图书馆数字化建设是维护文化多样性的一项艰巨任务。希拉克总统已提出，包括法国在内的欧洲各国图书馆资源必须在最大范围内，以最快的速度在网上被人们共享。

2005年4月27日，欧洲的19所国家图书馆已经联合起来，抵制拥有全球最大搜索引擎的美国Google公司发动的创建全球网络图书馆的计划。这19所图书馆支持另一项耗资数百万欧元的反攻计划：由欧洲国家建立自己的欧洲网络图书馆。此项计划由法国国家图书馆牵头，参加者有奥地利、比利时、捷克、丹麦、爱沙尼亚、芬兰、法国、德国、希腊、匈牙利、意大利、立陶宛、卢森堡、荷兰、波兰、斯洛文尼亚、斯洛伐克、西班牙和瑞典。

欧洲数字图书馆计划的主要目的：一是为了与Google筹建的全球最大网络图书馆抗衡，对Google的信息资源垄断进行“欧洲人的反击”；二是为了维护世界文化的多样性，在互联网上不能只有美国人的声音、只有英语、只有经美国人选择的历史、文化和美国人的行为准则影响人；三是为了阻止或削弱互联网上英语统治地位的进一步加强；四是不能只由美国来选择世界各国的历史文献，用他们的观点加以解释，强加给人；五是在法国和欧洲人看来，Google全球网络图书馆不只是具有文化意义或信息传播功能，它还具有一定的政治意义，Google选择的标准是以英美思想为主导的；六是为了让欧洲人的智慧、历史及文化遗产，以最快的速度在互联网上供全世界共享。

#### 1.2.4 我国数字图书馆建设与运营模式

我国数字图书馆的建设与运营模式主要有以下三种形式<sup>[6]</sup>。

(1) 国家投资，完全公益性运作。主要有中国实验型数字图书馆和中国高等教育文献保障体系(CALIS)。中国公益性数字图书馆相比于国外，主要是以国家投资为主，没有来自于各社会经济组织、基金会的资助，这就给它的运作带来了一定程度上的资金障碍。以中国实验型数字图书馆为代表的公益性数字图书馆主要运作不涉及知识产权的信息资源，在互联网完全公开免费使用；而以中国高等教育文献保障体系(CALIS)为代表的数字图书馆则主要运作学术、科研类信息资源(包括涉及知识产权和不涉及知识产权两部分)，服务对象主要为