数字图书馆
——原理与技术实现

高文　刘峰　黄铁军　等著

清华大学出版社
http://www.tup.tsinghua.edu.cn

数字图书馆

——原理与技术实现

高文　刘峰　黄铁军　等著

清华大学出版社

(京)新登字 158 号

内 容 简 介

数字图书馆是一个新兴的、涉及到互联网、多媒体、数据仓库、版权保护等诸多技术的计算机应用领域，应用和商业前景非常广阔。本书力图介绍数字图书馆从概念原理到系统实现的各阶段所需要的理论、算法与系统的知识，为从事此领域的科技人员和管理人员提供帮助。本书分为十七章，内容包括数字图书馆设计与建设所需要涉及的标准化问题、置标语言问题、多媒体海量数据库的管理问题、软件系统构造问题、个性化的 Internet 检索问题、智能化界面问题、数字图书与音视频作品的录入问题、多媒体数据的压缩与检索问题、版权保护问题、宽带多媒体网络问题、高性能计算机平台问题、与远程教育的关系、与电子商务的关系，以及传统图书馆的数字化改造问题等等。本书的读者对象为计算机、图书情报、网络文化教育等领域的研究人员和工程技术人员、管理人员、研究生、大学生。本书既可以作为教材或教学参考书，也可以作为工程开发的技术参考书。

书　名：数字图书馆——原理与技术实现
作　者：高文　刘峰　黄铁军等著
出版者：清华大学出版社(北京清华大学学研大厦，邮编 100084)
http://www.tup.tsinghua.edu.cn
印刷者：北京密云胶印厂
发行者：新华书店总店北京发行所
开　本：787×1092 1/16　**印张**：30.25　**字数**：718 千字
版　次：2000 年 10 月第 1 版　2000 年 10 月第 1 次印刷
书　号：ISBN 7-302-04079-6/TP・2405
印　数：0001～8000
定　价：39.50 元

序　言

在20世纪的后50年内，集成电路、计算机和通信技术取得了飞速的发展，由这些技术所开发的成果已经渗透到经济、政治和军事的各个领域，已经开始改变我们的生活和工作方式，人们已经深刻地认识到，高速发展信息技术、应用和产业对提高综合国力的战略意义了。

但是，也许当我们欢欣地参加完对20世纪信息技术辉煌成就的“庆功宴会”后，再回到实验室仔细地审视这些技术的发展现状时，就不难发现，信息技术的发展还远远不是“十全十美”的，还存在许多可挖掘的技术潜力和可开发的应用领域。例如，在集成电路方面，科技界致力于追求高性能处理器的速度，而对于开发光电集成、系统芯片、微机电芯片的潜力重视不够；在计算机方面，在提高数据处理的速度方面的确取得了骄人的进步，但当对计算机的智能化程度进行评估时，就令我们这些研究计算机的人汗颜了，人们在埋怨计算机“不好用、不够傻瓜化”，在指责计算机只会“计算”而不会“算计”；在通信方面，我们在通信带宽和互连技术上取得了历史性的突破，但当感受了一阵在信息高速公路上“兜风”、在信息海洋中“遨游”的喜悦后，随之而来的就是渐渐萌生失望之意了。计算机和网络的确能在“弹指之间”向我们提供“浩如烟海”的信息，但我们真正渴望的是知识，是解决问题的方案，是自学成才的环境。目前，我们常常在一大堆“多而无序、繁而不精、华而不实”的“海量信息”中“迷茫”了。

世纪之交，当我们在计算机、网络和多媒体等技术的最新成果的支持下，为能够获得海量信息而自豪时，也急切地盼望着能早日开发出更为有效的海量信息的处理技术，以及更为实用的海量信息的应用系统。人们正翘首盼望，计算机和网络技术能早日成为我们在认识世界和改造世界时，实现“去粗取精、去伪存真、由此及彼、由表及里”的得力帮手。就是在这样的背景牵动下，一些新的设想、新的概念、新的术语就应运而生了，例如：虚拟现实(virtual reality)、数据挖掘(data mining)、数字地球(digital earth)、数字图书馆(digital library)等等。

作为一名计算机的科技工作者，每当提到以上这些新的术语时，总是因为这些术语中文译名的“词不达意”而感到歉意。这一方面是由于这些新术语所表述的是一个发展中的概念和正在探索中的技术，“仁者见仁、智者见智”。另一方面，也是由于我们离达到“信、达、雅”的翻译准则相差甚远，我们多么钦佩能够将“Operations Research”翻译成“运筹学”的范例。细想一下，“虚拟现实”的本质并非字面意义上的“虚拟”，“数据挖掘”的精髓是从信息中提炼知识的过程，而“数字图书馆”的宗旨是为21世纪的人类创造一个开放、灵活、没有围墙、不设门槛、突破时空限制、拓宽人际联系、享受因材施教和终生受益的最佳教育环境。

就说“数字图书馆”吧，也许人们首先想到的是把现有图书馆内的精品馆藏资料进行数字化处理和加工，以使更多的读者可以通过网络和计算机共享这些文化瑰宝。当然这是一件很重要的工作，但“数字图书馆”的真正潜力和内涵绝不仅仅是传统图书馆馆藏资料的“数字化”，也绝不仅仅是一个数字化了的“图书馆”。Digital Library 一词是美国的科技工作者首先提出来的，我很钦佩他们对新事物的敏感和创新精神，但随着当前对 Digital Library 需求的扩展和支撑技术的发展，以及国际上某些先进的实例的演变，看来 Digital Library 一词也已经无法涵盖其实现机理和所涉及的应用了。

我们必须意识到，一个生活在 21 世纪并决心在竞争中取胜的民族，将面对一个更为开放、复杂、迅变的客观世界，我们在认识和改造这个客观世界的同时也必将重新认识和改造自身。在高新技术的支持下，我们认识问题和处理问题的时间尺度和空间尺度都将发生剧变。教育者未必处处和时时都能比被教育者更快地领悟、掌握和开发出某种新技术，交互的、主动的、启发的、适人的（如同体操教练和京剧教师那样地对学生进行手把手地因材施教）、智能共增的（在教育过程中，教育者、受教育者和教育环境的智能得到同步地增长）教育方法和环境必将成为主流。因此，用传统的方法编写出版的教材将跟不上需求和技术的发展步伐，原来的教育宗旨、目标、手段和方法必须做出相应的变化，“授之以鱼，不如授之以渔”必将成为 21 世纪的教育宗旨。其实，我国早在《礼记·学记》中就已经指出，最好的教与学的途径应该是“教学相长”，而我国宋朝的苏轼所提倡的“博观而约取”就应该是我们开发这样一个学习环境中的智能浏览器和知识挖掘器的“指导思想”，而我国晋代的儒学家陆机在他的《文赋》所提倡的“观古今于须臾，抚四海于一瞬”更应该是我们开发这样一个学习环境中的多媒体应用、可视化计算、远程教育和虚拟现实技术的最高追求了。爱因斯坦曾说过：“人的差异在于业余时间”，在信息时代，根据我自己的体会，是否还可以加上一句话，“人的差异还在于是否善于利用网络和计算机”。

从以上这些粗浅的分析出发，是否可以得到这样一个结论：我们必须适应 21 世纪对人才培训的需求，密切结合中华文化的特性和我国的实际情况，充分利用计算机、网络、多媒体和人工智能等技术的成果，积极开展对我国的“数字图书馆”的规划、开发和应用的研究和实践，致力于创建一个以人为本、能够开展自适应学习（adaptive learning）和终生教育的环境。我国的“数字图书馆”将是聚集我国信息界、教育界、文化界、艺术界的成果，凝聚我国五千年文化精髓，面向 21 世纪教育需求的中华文苑。

我国 863 计划信息领域智能计算机专家组的同志们意识到创建我国的“数字图书馆”的伟大意义和历史责任，在国家科技部的领导和支持下，及时地部署了对“数字图书馆”的某些支撑技术的研究，并对国内的有关研究小组进行选优和落实。现在，又把其中的部分阶段成果汇编成本书。在此，请允许我衷心地祝贺本书的出版，并向撰写本书的各位作者表示敬意，感谢他们为我国的信息资源建设、知识基础设施建设和我国“数字图书馆”的建设写了一本有关支撑技术的入门书，起到了“立足当前、面向未来”的作用。虽然本书只是着重介绍了某些构建“数字图书馆”的开发平台的技术实现手段，还未能深入地研讨在这个平台上构建各类应用系统的开发方法，也尚未全面论述面向未来需求的“数字图书馆”、

远程教育和自适应学习的教材编写和教学方法等问题，但我相信本书将为构建我国“数字图书馆”和其他“知识系统”添加了一块重要的基石。

让我们共同参与这件“功在当代、利在千秋”的伟大工程吧！

汪成为

2000 年 6 月 14 日

目　录

内容与平台篇

应　用　篇

第 1 章

概　　论

计算机和互联网是20世纪人类最重要的技术发明之一。经过了近二十年从概念到技术的讨论和研究之后，20世纪90年代初互联网开始进入商业领域，并在随后的几年中获得了空前的发展。它先是主导了IT发展的技术市场，并在资本市场的推动下对社会和经济的发展施加了巨大影响。随着技术的发展和应用的普及，它将广泛地深入到人们的社会生活中，使人们进入一个崭新的网络经济的时代。多媒体技术在90年代初开始升温，并由于在电视、电影、新闻图片等方面的技术进步被传媒所重视与接受。互联网和多媒体技术不仅改变了社会的生产力与生产关系，也将改变人们的读书、学习、生活和工作习惯。

1.1　数字图书馆和远程教育是网络与多媒体发展的产物

随着人们对网络带宽需求的增加和宽带网络自身的发展，网络应用的类型在不断扩大，互联网上信息的类型也变得越加丰富，可以列举的数据类型包括文本、图形、图像、视频、音频、动画等等。计算机、网络以及通信的发展使得产生、处理、传播数字信息的能力大大增加，而且数字信息在存储传输和处理时比其他形式存储的信息更方便，加之在过去的几十年中产生了海量的数字信息资源，所以技术上需要一种系统技术来管理数字信息资源。因此，互联网技术领域面临一系列问题：怎样合理和有效地对各类海量数字信息进行组织、检索、访问、利用？怎样有效利用互联网的优势向用户提供海量数字信息服务？针对这些问题，美国科学家在90年代初提出了数字图书馆(digital library)这一概念，力图为高速宽带互联网做好应用准备。数字图书馆是一个驱动多媒体海量数字信息组织与互联网应用问题各方面研究的技术领域。简单地说，数字图书馆是以电子格式去存储海量的多媒体信息并能对这些信息资源进行高效的操作，如插入、删除、修改、检索、提供访问接口和信息保护等。它曾成为克林顿政府倡导的信息高速公路计划NII的重要部分，美国希望通过数字图书馆这一应用推动国家信息基础设施的建设，并最终把传播和利用知识的高速信息公路铺到每个美国人的家里。

传统图书馆是储藏图书资料的仓库，它负责收集、选择和整理图书资料，使其可以被查询利用。保存图书资料和提供便利的利用办法与环境是图书馆的主要任务。数字图书馆所面对的领域远远超出了目前传统图书馆的范围，它不仅需要存储数字化的图书、音视频作品、美术作品图像、电影卡通作品、电子出版物、互联网新闻、互联网上各种需要保存的数据、卫星数据、气象数据、地理数据等各种各样的人文与科学数据，还要提供互联网上

基于内容的多媒体检索，包括对文本、音频、视频、图像、图形、地理、遥感等数据的检索与索引，使得合法用户可以通过互联网利用这些数据。用户也可以用新的媒体工具把多个信息组合在一起生成新的媒体内容。数字图书馆将逐步实现智能化、个性化和自动化服务，使得用户可以使用个人电脑通过网络进行一些基于内容的检索，用户可以用各种形式提出查询请求，甚至是用口语的形式。

数字图书馆是一项非常有意义的研究内容，特别是对于教育领域，数字图书馆将成为非常重要的教育设施。在未来，数字图书馆将无处不在，包括远程教育、电子商务和娱乐。在今后的十年中数字图书馆将大大影响教育的质量和生活的质量。

1.1.1 数字图书馆的真实含义

我们现在所使用的“数字图书馆”一词，是从英文 Digital Library 翻译过来的，因为美国人先使用了这个术语。这个词翻译的对不对，到底应该怎样理解这个词？Digital 的意思很明确：数字化的，但 Library 这个词翻译成图书馆对不对呢？我们来看看 Library 这个词的本意。在英文中 Library 有两个基本解释，一个是“图书馆”，另一个是“库”。Digital Library 的英文本意应该更强调的是“库”，而不是“图书馆”。因此，Digital Library 最准确的翻译应该是数字资料库。数字图书馆和数字资料库的区别有多大呢？很大！因为一说数字图书馆人们想到的是一个图书馆，有一座大楼，有宽敞的阅览室和大型书库，当你查到书要求提出来后有一条自动化传送带把你要的书从书库的深处送到你的面前。但是数字资料库就是另外一回事了，数字资料库是一群计算机系统，它们可能放在一起，也可能北京一台，上海一台，南京一台，然后用互联网连接到一起的分布式系统，用户只要能上网，就可以查数字资料库，所以它和传统的图书馆几乎没有什么共同的地方。当然，由于数字图书馆一词已被广泛使用和接受，我们不妨仍沿用该词，但必要时须强调它的真实含义，以免初涉者望文生义。实际上，现在已经有不少人认为数字图书馆就是将现有的图书馆图书资料数字化后上网，这是一种误解，是把一个复杂的分布海量多媒体计算机管理系统看成一个简单的图书馆信息管理系统，需要我们向从事该领域工作的科学家和工程师们不断说明，让新进入者少走弯路。

1.1.2 数字图书馆的功能与作用

首先，数字图书馆应该是一个国家数字文化平台。其中包含的内容很多，它既可能有网上图书馆，又可能是网上书店、音像店、文物店，还可能是网上文化中心等等。

第二，数字图书馆还应该是一个国家数字教育平台。这并不是说数字图书馆可以取代大学教育。通常人们在图书馆里是进行自学和继续学习，包括文化的学习、休闲爱好的学习以及各种各样知识的学习。因此，数字图书馆也应该提供这样的功能，即成为网上业余教育中心、在职教育中心甚至趣味教育中心等等，很多家庭教育都可以在这上面进行。

第三，数字图书馆也是一个国家数字资源中心，这一点很少有人提到，希望能够引起大家的重视。我们现在面对的数据越来越多，如遥感数据、卫星数据、网上数据，这些数据应该收集保存起来，否则若干年后谁想要研究我们现在的社会，连数据都很难找到。那么，到底这些数据应该由谁负责收集，应该放到哪里？目前这些网上资料大多存储在各用户网

站或者信息中心，并无宏观规划，基本上是市场行为（如互联网公司）和行业需求驱动的信息收集。这并不合理，这样下去我们将对不起后人。从全局统一规划考虑，国家一定要有一个数字资料集中管理的地方，把卫星图像资料、网上资料及其他数字资源的资料，甚至一些产品资料保存起来，成为数字资料存储中心和数字资料处理中心。这个中心，就是国家数字图书馆，它应该是公益性的，而不应该是商业性的。

由此可见，数字图书馆绝不仅仅是数字化的图书馆，它应该是中华文化的传播媒体，是文化产品的网络商务平台，是国家数字资源组织、开发和利用的基础，是网络文化中心和网络文化的聚集地。

数字图书馆是构架在21世纪的计算环境之上的，是21世纪的宽带多媒体信息利用方式。这个环境构造好了，对中华文化及我们国家都是非常重要的。数字图书馆建好了，可使我们跨越式发展达到与发达国家同样的科学与文化环境，使中华民族在下一轮知识经济的挑战中站在更高的起点上参与竞争。

数字化时代的到来给我们的信息技术带来了巨大的机遇，我们应该脚踏实地地做好技术储备，才能跟上形势，抓住出现在我们面前的千载难逢的机遇。

1.2 数字图书馆建设面临的技术挑战

数字图书馆作为一个海量、宽带多媒体网络系统，还有很多需要进一步研究开发的技术问题。这些问题解决的好坏，将会直接影响数字图书馆建设的速度。以下，我们从十个方面讨论这些问题。

1.2.1 信息资源建设

数字图书馆作为一个数字资料库，应把包括历史资料在内的所有资料数字化后放进去，即原有资料的录入；另外还有其他资料的整理入库，包括在线网上资料、广播及媒体资料、数字资源等的整理入库。

已有图书的数字化工作，需要有效的管理机制。我国目前有各种图书馆数千个，分别隶属国家部委（例如中央党校、文化部、中国科学院、中国社会科学院）、各省市、各大学等等。显然，直接从行政入手进行规划难度很大，很难统一调动。我们现在的思路是，通过国家科技发展计划，采用统一规划联合实施的办法进行信息资源建设的协调管理。例如，对于一本书的录入，应该保证如果某个图书馆已经将其数字化录入进去了，其他的图书馆就不要再重复做了，除非在质量上有新的突破。这项工作如果配合不好，很可能同一本书的录入做许多遍，重复劳动，给社会造成很大的浪费。

音视频作品和图像的数字化工作，需要资源持有单位和技术单位的共同努力。目前音频作品的数字化在技术上需要考虑满足MP3、AC3、MPEG等格式需求。视频作品和视频广播题材需要考虑满足MPEG-1、MPEG-2广播级（MP@ML）、MPEG-2高清级（HP@HL）等格式的需求。图像作品需要考虑JPEG、JPEG2000等格式的需求。

版权的使用将是另一个需要花大气力才可能突破的问题。一本书的版权涉及到作者和出版社，怎样才能有效地保护作者、出版社、读者、数字图书馆几个方面的合法权益，涉

及到网络经济的立法问题，尚需进一步探讨。

1.2.2 数据存储与压缩

数字图书馆所涉及的数据类型有文本、图像、语音、图形等，而且所面临的数据是海量的，可能会有10^{12}～10^{15}字节，这么大的数据量是迄今为止其他任何系统都没有遇到过的，需要大规模数据库存储和处理这些数据。目前的数据库能否应付如此海量数据的存储与管理？即使现在的数据库在管理上能满足要求，能否承受得了其系统成本？因此，如何保存和管理海量数据是系统设计的核心任务之一。

在数字图书馆的数据中，文本数据的存储量不是很大，真正大的是多媒体数据。因此，对多媒体数据必须进行压缩，然后保存在数据库中，以降低库的成本，使库的规模保持在可管理的范围内。基于模型的编码就是压缩方法之一，如对一段电视“新闻联播”中主持人讲话的录像，如果将其原封不动地保存下来放到数据库中，可能要占几兆字节的空间，而压缩后，可能只占几百K(千)到几十兆字节。

1.2.3 分类、索引和检索

在图书馆中，分类与索引是检索的基础，分类方法也有各种学派和门类。目前在数字图书馆领域中，还没有完全统一的分类方法，这就提出了一个如何统一分类标准的问题。如果没有一个统一的索引方法和分类标准，将来开发计算机的搜索工具就非常困难，需要针对不同的分类方法制作不同的搜索工具。

另外，我们所面临的数据类型也不同，如文本信息、地图信息、图像信息及视频、音频、音乐等信息，对不同的内容，需要不同的分类体系和索引机制。而能否制定一个比较好的分类方法、建立一个比较好的索引机制，将直接影响到后续能否开发出一个比较好的检索工具。

对于检索，假如在Internet上检索“数字图书馆”这个词，目前的检索常常是这样实施的：把“数字图书馆”切分成“数字”和“图书馆”，而凡是与“数字”、“图书馆”、“数字图书馆”有关的条目都会检索出来，其实这些并不都是用户需要的，并且由于这样检索出的条目往往非常多，而使用户无从下手，无法找到需要的信息。目前的分类器绝大多数都是尽量多地给出信息，而不管有用没用。

因此，怎样做一个比较好的检索工具，使得提供给用户的信息恰恰是用户最需要的(不需要的一条也没有)，也就是说，海量数据的搜索效率(最优解)与速度是系统面临的最大挑战，其中包括中文搜索、图像搜索、语音搜索和智能搜索。这当中涉及大量人工智能的支持，这也正是数字图书馆与国家863计划所支持项目密切结合的必要性所在。国家863计划前些年在人工智能方面支持了很多研究工作，有很多积累，都可以用到这上面来。

1.2.4 传输与保护

如果你经常上网，你就会发现，访问国内的站点和国外的站点都很慢。据说，访问国内站点慢的原因是网站服务器的处理能力太弱，而访问国外站点慢的原因是国内网络的出口带宽太窄。因此，怎样增强服务器的处理能力或者从服务器端调度协调好，当用户提出

一个服务请求时，用最短的时间对用户的请求进行回答，是网站系统成功的关键，其中有许多问题需要解决，如带宽的有效使用问题。现在所有的搜索中都存在这一问题，搜索工具只管找到用户的解，而不管使用多长时间。另外，当用户提出的关键词(keyword)关联性不很强，其组合方式又很初级时，可以预见到，这种检索会花很长时间，解也会很多，这时应该在检索之前提醒用户，再增加一些约束条件，以加快搜索速度，并使检索出的信息真正是用户所需。这些问题可以在设计工具时预先想到。

在多媒体检索时，将来应该有一种快速图像浏览的服务机制，即多媒体解的分层传输机制：如果用户提交了一个多媒体检索请求，并且搜索引擎一次找到很多照片或图像，系统的做法应该是将找到的照片分成若干层，先将最粗的那层传给用户，再逐渐细化，而当用户认为这张照片不是所要的时，可随时结束传输，再换另一张浏览。这其中有很多工作要做。

保护包括版权保护和系统安全性的保护。版权保护，是数字图书馆作为商业系统运行的前提。没有版权保护的手段，作者就不可能允许数字图书馆经营者把自己的作品放到网上，投资者也不会把钱投到系统的建设上。从美国的经验教训看，虽然数字图书馆在美国的呼声很高，政府已经投入了大量的经费支持，但这些经费都是投给研究部门的，主要是大学。其原因除了说明数字图书馆这个领域目前的技术问题还没有研究透，还需要技术攻关以外，还有一个原因，就是版权保护问题尚未很好地得到解决。如何在网上销售商品，美国已经有了相应的法律，但如何在网上从事商业性借书目前还没有讨论清楚。美国的公共图书馆都是公益性的，有合法身份的居民可以凭居民证件免费办理借书证，可以一次借 5～6 本书。尽管图书馆是公益性的，但作者本人的利益仍可以因为大多数图书馆都购买一定数量的图书而得到保证。数字图书馆如何保证作者利益？一种现成的做法是采用与传统图书类似的方法，即每个数字图书馆要单独购买一定数量的许可证，其他方法需要重新设计，需要有其他法律法规做后盾。

1.2.5 交互界面

交互界面(用户界面)是数字图书馆的重要组成部分，是系统与用户交流的窗口。其实，这不仅是数字图书馆所面临的挑战，任何系统都有这个问题，即怎样设计一个理想的用户界面，让用户使用时得心应手，能够友好、直观、方便，并具有人性化、智能化，充分利用图形、语音，将其融为一体等等。

交互界面设计的技术核心是如何吸引用户的注意力和为用户的操作提供最方便的支持。

1.2.6 输出与信息表现

在计算机上用各种可能的技术表现信息是非常具有挑战性的工作。信息的输出和表现是数字图书馆中可能为未来社会带来很大好处的一个方面，除了在经济上、学习上有好处之外，通过对数字图书馆的研究，使得人类对信息的发现、信息的利用更上一个档次。

实际上，有了多媒体后，许多东西都可以直觉化、可视化，用图像、图形、语音等直接表现出来。我们平时在 PC 上所面对的数据量很小，还达不到能从这些数据中得到更重要知

识的程度。有了数字图书馆后，大量的资料数据都很容易到手，这对研究人员是一个大福音，因为利用这些数据可以有许多创新的机会。我们不仅可以从这些数据中挖掘到目前为止人们还不了解的知识和规律，还可以用多媒体技术描述这些知识，使不同知识表现之间可以相互转化，再利用信息可视化技术、虚拟现实技术将各种各样的知识表现出来。

1.2.7 多语言问题

互联网上的信息可能是用英语、汉语等多种语言书写的，书也可有多种语言的问题，为了让更多的人能够方便地阅读各种语言的图书资料，需要提供机器翻译能力。这个问题也是国家863计划一直在支持研究的项目，包括自然语言理解、机器翻译问题、多语言浏览器等等。现在市面上也有一些多语言浏览器，但还不够理想，今后，还需在这方面进一步努力。

1.2.8 工具与平台

工具包括图书录入工具、音像制品录入和编辑工具、浏览器工具、开发工具等，平台包括软件平台、数据库平台等。目前已经有一些商品化的软件平台，但是仍然需要专门为数字图书馆设计的专用工具与软件，这是一个最大的挑战，对象包括总体结构标准、软构件技术、信息录入工具、搜索工具和知识挖掘工具等问题。

1.2.9 标准

在开始建设中国数字图书馆时，面临的一个严峻事实是没有相应的电子图书标准、元数据标准，也没有多媒体信息标准。而现阶段，技术标准在一定程度上已成为国家主权的延伸。因此，在建设中国数字图书馆工程时，相关技术标准工作的研究尤为重要。技术标准的草拟应该由信息产业界、图书情报界以及与标准相关的国内软件开发商共同参与，在标准讨论的同时开发一批建立在这些标准基础上的软件系统。

1.2.10 开放性

数字图书馆是一个集成各种数据资源和工具环境的大规模系统，因此系统的开放性是成功的必要条件。所谓开放性应遵循以下原则：第一，统一性：不论哪种类型的图书馆都必须服从整体协议；第二，分布式：不可能有一个中心，这是现今网络系统的基本点；第三，开放式：内容必须具有开放性；第四，可扩展性；第五，要简单易行；第六，能比较充分地利用现有信息的服务设施。

1.3 国外数字图书馆的发展历史与现状

数字图书馆在美国已经作为“高性能计算和通信计划(HPCC)”的子课题“信息基础技术应用(ITTA)”中的挑战性课题得到政府支持。从1993年开始，美国国家自然科学基金会(NSF)在联合受理“数字图书馆预研(Digital Libraries Initiative，DLI)”联邦项目方面担当领导角色。数字图书馆能够成为一个研究、开发、应用和实践的重要领域，是和