数字图书馆个性化服务研究与实践

——基于新型决策支持系统

熊拥军 张建中 袁小一 黄湘林 编著

国防工业出版社
National Defense Industry Press

数字图书馆个性化服务研究与实践
—基于新型决策支持系统

熊拥军　张建中　袁小一　黄湘林　编著

国防工业出版社
·北京·

据仓库基础上，对文献资源进行了关联分析与挖掘，并采用数据挖掘技术对资源访问行为相似读者群进行了划分。第8章介绍数字图书馆个性化信息推荐相关技术，如基于内容过滤的推荐、基于协同过滤的推荐和混合模式推荐。第9章结合实际应用，对数字图书馆个性化服务系统进行分析和设计，对个性化推荐的部分功能进行了实现。

本书在编写过程中参考了大量文献，已尽可能在正文中标注并列在每章的后面，但其中仍难免有所遗漏，这里特向被遗漏的作者表示歉意，并向所有作者表示诚挚的感谢。

本书主要为进行信息管理研究的专业人员提供研究资料，具有很好的可读性和实用性，也可以作为图书情报专业学生的参考资料。

由于作者学识水平有限，书中的疏漏或错误之处在所难免，恳请广大读者批评指正。

熊拥军

2012.7

目　录

第1章 绪 论

近年来，随着“数字化生存”方式逐渐为人们所接受，数字图书馆因其信息量大、占用空间少、更新速度快、不受时空限制等特点而越来越受到人们的关注。但人们在享受着数字图书馆所带来便捷的同时，也深受其庞大而形式多样的信息资源困扰。同样的信息对于不同个体表现出不同的价值，对单个用户来说，不可能对数字图书馆的所有信息资源都需要，而同样的信息也不一定会满足所有的用户。个性化服务是解决用户“众口难调”问题的关键，它是适应用户多样化需求的重要手段，也是图书馆应对信息资源多样化的一个重要措施。

1.1 技术背景

随着计算机技术和通信技术的不断发展，人们对信息的需求越来越高，已不满足于计算机能帮助他们迅速处理具体业务，而是需要从大量业务数据中探索出业务活动的规律、市场的运作趋势，并为他们参与市场竞争做出重要的决策。但是，常规的数据库管理系统因自身的局限性，无法满足对较大规模决策支持数据的分析。决策支持系统在数据的处理、组织与管理方面与信息管理系统有很大差别，它除了通过对历史与现状信息、系统内与系统外的数据进行加工处理，掌握尽可能多且真实、准确的情报，进而发现问题外，还要产生如预测结果、决策方案的实施条件与产生的后果及影响等增值数据。

在20世纪90年代初，一种适用于决策支持系统的数据组织与管理技术应运而生，这就是数据仓库(Data Warehouse，DW)技术[1]。它不仅被理论界和学术界视为是对数据库技术和人工智能技术的重大发展，而且也被企业界看做是能够为其带来巨大社会效益和经济效益的应用领域。目前，一些主要的数据库厂商(如IBM、Oracle等)业已开发出支持数据仓库产品。

虽然数据仓库技术的发展，为决策支持带来了可喜的变化，但是随着信息总量的不断增加，缺乏强有力的数据分析工具，导致了“数据丰富，但信息贫乏”的现象。迫切需要有效的分析工具，来发现大量数据间隐藏的依赖关系，从大量数据中抽取有用的信息或知识。尽管很早就出现了简单的数据统计技术和随着数据仓库同时发展的联机分析处理(On - Line Analytical Processing，OLAP)技术[1]，但它们不能解决这些问题。

数据挖掘(Data Mining，DM)技术[2]旨在能从大型数据库中提取隐藏的预测性信息，又称知识发现。它能发掘数据间潜在的模式，找出企业经营者可能忽视的信息，为企业做出基于知识的决策参考意见。目前国际上在该领域的研究相当活跃，无论在理论上，还是实用技术上都取得了喜人的成果，同时也开发出了各种专用或通用的商业数据挖掘软件。

目前，数据仓库和数据挖掘以及联机分析处理，三者的有效结合，被认为是一种新型的决策支持系统[3]。虽然很多领域都有产品出现，但遗憾的是在图书馆的应用领域，提供决策支持的相关产品还未出现。

1.2 行业背景

在互联网迅速发展的今天，信息资源呈爆炸性增长，用户信息需求的特定性与信息资源分布的

无限分散性之间的矛盾也日益加剧。用户获得信息的主要障碍已从距离上的障碍变为选择上的障碍。这种用户需求的转变决定了图书馆信息服务工作重心的转移，即从以自我为中心的被动服务发展到以读者为中心的主动服务，即个性化的服务。“个性化”理念在图书馆的提出有其当代依据。

其一，当代“以人为本”的哲学观为个性化信息服务奠定了思想基础。所谓“以人为本”指的是人是一切活动的出发点和最终归宿。一切为了人，一切着眼于人的发展与人的需要。“以人为本”是当代哲学的重要观念。这一思想有着广泛的社会认同。在今天，以用户为中心的思想几乎渗透于社会的每个经济活动中，在信息服务业提出个性化服务，乃是信息服务向纵深发展的一个重要内容，也是当今信息服务业赢得更大服务空间的重要途径。

图书馆最基本的功能就是收集、储存和传播文献信息。在人类进入信息时代的今天，图书馆已跨过“以藏为主”、“藏用并重”的历史时代而迈入“以用为主”的时代。“以用为主”就渗透了当代“以人为本”的观念。图书馆是一个信息服务行业，图书馆工作就是为读者服务，没有读者，图书馆也就失去了存在的意义。

早在70年前，印度图书馆学家阮冈纳赞就提出了“图书馆五定律”，即“每个读者有其书，每本书有其读者，节省读者的时间，图书馆是一个生长着的有机体”[4]，揭示的就是一种“以读者为中心”和“书为人人”的图书馆文化。毫无疑问，图书馆工作的中心，应该是使读者在尽可能良好的环境中，尽可能快捷、准确、方便地获取信息。所以，图书馆文化建设的核心理念就是：一切以人为本。个性化信息服务的含义之一就是“基于信息用户的信息使用行为、习惯、偏好和特点，来向用户提供满足其各种个性化需求的一种服务”，每一个用户都有着不同的兴趣爱好，有着不同的信息需求。只有很好地利用这一点，信息服务才会取得显著的进步。

其二，个性化信息需求为个性化信息服务提供了不竭的动力。20世纪60年代以来，尤其是近年来，计算机技术、多媒体技术和通信技术的迅猛发展与综合利用，使人类进入信息时代。个人所面临的全新的信息环境为：信息量爆炸式增长、信息资源数字化、载体形式多样化、传播方式网络化、传播速度高速化和信息共享的全球化。但是，人类正遭遇信息尴尬：人类的信息总和在不断地以指数增长，尤其是数字化的信息。信息技术和网络技术的飞速发展给人们的交流、信息的传播带来了革命性的变化。信息技术给人们生活、研究和工作等各方面带来了方便，但同时不能不承认另一个事实：人们逐渐被淹没在形形色色、各式各样的信息海洋中。人类社会面临着信息产生与信息获取之间的矛盾，而且愈演愈烈。

因特网上的信息是极其无序的，而且信息量越大，就越难被利用，虽然用户有利用搜索引擎来查找信息的便利，但实际上，搜索引擎并不能有效解决信息过量带来的尴尬。面对纷繁复杂的信息世界，人们只对其中的小部分感兴趣，因此个性化信息服务开始成为网络信息服务需要考虑的一个关键问题。尤其是在高校，由于教学科研的时间有限，时间显得更为珍贵，广大师生信息需求的指向性更为明确，对信息获取的准确性、快捷性要求更为迫切。高校图书馆的个性化信息服务就显得尤为重要。

其三，图书馆的数字化发展为个性化信息服务提供了更为便利的现实条件。数字图书馆是指在计算机网络的基础上，运用信息技术，对数字化形式的文献资源进行搜索、排序和组织，为读者提供当地的和远程的，文本的和声像的，页面的和动画的文献信息的公共服务机构。为了迎接信息时代的挑战，我国几乎所有高校图书馆都已经配备了计算机集成管理系统，实现了管理自动化，并且采购或自建了大量的数字资源，这为数字图书馆建设的新阶段提供了现实条件。

个性化信息服务作为一种服务方式，虽然早已存在，但是只有在信息时代的今天，它才以理念的方式召唤着信息服务的个性化方向。可以预见，图书馆的数字化建设和个性化服务将是我国高校图书馆建设的发展方向。

1.3 研究目的和意义

数字图书馆个性化服务的研究是IT行业新兴起的一个交叉研究领域，它涉及情报学、信息管理、计算机技术、数学等学科。数字图书馆个性化服务包括个性化和主动两个方面[5]。个性化服务是对不同的用户采用不同的服务策略，提供不同的服务内容。主动服务则是指很少需要或不需要用户做什么，而是由系统自动按照用户的信息需求提供相应的服务。个性化主动服务将使用户能以最小的努力获得尽可能好的服务。

图书馆作为信息资源收集、加工和服务的中心，随着信息技术的不断发展，在图书馆积累了丰富的数字信息资源。图书馆的数据库系统可以高效地实现数据的录入、查询、统计等功能，但无法发现数据间存在的关系和规则，无法根据现有的数据预测读者的信息需求，缺乏挖掘数据背后隐藏的知识的手段，以致无法为读者提供更为方便、快捷和高效的服务。

在数字图书馆服务中，本研究考虑采用数据仓库、数据挖掘和联机分析处理技术，对读者、资源以及读者对资源的访问等数据进行分析和挖掘，从中发现读者兴趣和资源的关联，来为读者开展个性化服务，服务内容主要表现在以下几个方面。

采访服务：信息资源的建设是为读者提供服务的基础，在分析历史采购信息、读者信息、资源流通和阅览信息、读者反馈信息以及来自外部的各种学科发展信息的基础上深入了解学科的走势和读者的需求，以读者隐性参与的模式，帮助采购人员确定采购重点，保障图书馆信息资源体系的科学性和合理性，以及采购资金的合理分配，为信息资源的采访提供决策服务。

咨询服务：图书馆的咨询服务由原来读者与馆员面对面的咨询发展到数字图书馆的网络虚拟咨询。通过网络，读者除了可以从咨询馆员或专家那里获取信息之外，还可以进入知识库，获取自助式的服务。其中知识库的建立需要行业专家的组织，以及运用数据分析与挖掘的方法从历史数据中分析，挖掘出隐藏其中的规律信息，形成满足用户需求的深层次信息产品。另一方面，还可以根据读者的历史咨询及访问信息，分析出他们的研究方向和兴趣所在，变被动咨询为个性化主动咨询。

资源分析：通过对数字图书馆的各种数字信息资源进行分类后，从不同的视角观察资源的配置及利用情况。分类的角度可以根据学科专业（如中图法图书分类）、馆藏分布、文献类型、语种、年代等进行划分，从而分析资源的配置、利用率、使用价值以及不同学科之间的关联，为采访、咨询和服务等的决策提供科学依据。

读者服务：主要是对读者群体进行分类后，从不同的视角分析读者对资源的利用情况。划分的角度可以按照读者本身的自然属性，如读者性别、年龄段、年级、学历、院系、专业等，从而分析读者的资源访问情况，识别读者的兴趣、发现潜在的访问规则、读者群等，例如，如果发现有很多读者访问了A文献也会访问B文献，则对访问A文献的读者，可以将B文献推荐给他。又如，根据同一学科的读者对不同学科资源的访问或同一学科的资源被不同类型读者的访问，来发现学科之间的关联和读者的关联。

另外，通过对历史访问数据和读者兴趣的分析和挖掘，预测读者的资源需求，在读者下一次访问时，使用所挖掘的信息，动态地提供个性化推荐服务。

1.4 国内外研究现状

1995年3月，卡内基·梅隆大学的Robert Armstrong等人在美国人工智能协会上提出了个性化导航系统WebWatcher，斯坦福大学的Marko Balabanovic等人在同一次会议上推出了个性化推荐系统LIRA。同年8月，麻省理工学院的Henry Lieberman在国际人工智能联合大会（IJCAI）上提出了

图 1-1　北卡罗莱纳州立大学图书馆界面

在默认界面上，系统给出了百科全书、词典、电子图书与电子文档、目录、索引与文摘，共五大默认定制板块，并且在每一类的下面都预先设置了一些最常用的定制资源。如在词典里包括了牛津词典、韦伯斯特词典等。

北卡罗莱纳州立大学图书馆的 MyLibrary 的工作原理为：首先由学科图书馆员将图书馆的数字资源按学科主题或资源类型为用户创建一个资源列表；其次系统给用户提供一个登录账号，用户通过账号登录后，可以在图书馆所提供的资源列表中选择自己的所需资源及其他的 Web 资源加入 MyLibrary，也可以直接选择某一个专题模板，各类资源一般以文件夹的形式进行组织；另外 MyLibrary 还提供如最新快报服务、书签等服务，用户也可以根据需要选择服务项目，而系统则为每一个用户建立策略文件，内容包含用户的账号、密码和代表用户选择数字资源清单的参数。这个文件以 Cookies 的形式被保存在用户使用的计算机中，或者保存于服务器端的数据库中。当用户以后访问 MyLibrary 的 Web 页面时，策略文件中的参数被提取出来，通过 Web 服务器向用户返回定制的页面内容[15]。

Cookies 是为了弥补 HTML 的一个缺陷而产生的。HTML 是一种无记忆的协议，也就是说用户目前正在浏览的主页对在此之前浏览过的主页没有丝毫记忆和了解。而实际的需要是希望浏览器能够记住一些信息，这个需求 HTML 本身无法解决，于是引入了 Cookies 的概念，也就是由 WebServer 网页服务器向浏览器写入一些信息，这些信息用户无法看到，当浏览器向此网址的其他主页发出 GET 请求时把此 Cookies 信息也会同时发送过去，供该主页使用，这样就实现了一定程度上的 HTML 的记忆能力。

3. 弗吉尼亚公共健康大学的 MyLibrary@VGphu 系统[16]

弗吉尼亚公共健康大学图书馆个性化服务方面的功能有：

（1）快速查找：可选择在图书馆目录、百科全书、词典，以及 Google、AltaVasta 等搜索引擎，共计 13 个数据源中进行快速查询。

(2) 定制数据库:包括8个综合数据库、21个专题数据库以及图书馆推荐的4个数据库,此外还可以由用户自己添加其他数据库。

(3) 定制页面:用户可以通过修改相关参数,来选择符合自己浏览习惯和爱好的色彩、字体大小及页面布局。

(4) 表格与服务:提供问答、续借、影印、预约、到期通知、馆际互借、最新消息等服务项目。

(5) 我的书签:该功能类似于浏览器提供的书签(Bookmark)/收藏夹(Favorite)功能,允许读者挑选自己常用的Web页面URL地址放入书签。与浏览器的书签相比,MyLibrary书签的内容可以让用户在任何机器上访问。

除了最早的康奈尔大学图书馆和北卡罗莱纳州立大学图书馆,美国华盛顿大学、加州数字图书馆、新加坡国立图书馆等都相继采用信息定制和推送等方式开发了自己的网络个性化服务系统,并收到了良好的应用效果。

1.4.2 国内个性化服务系统研究现状

我国数字图书馆的研究与建设真正始于20世纪90年代后期[17]。最早在1999年底,国家科技部支持的"中国数字图书馆示范系统"项目中就提到了数字图书馆的个性化服务问题。2000年初,社会科学基金资助的"基于Web的数字图书馆定制服务系统"项目[18]开始研究开发实用的数字图书馆个性化定制系统,该项目是由北京大学信息管理系余锦风教授负责承担的。2002年10月开始的《中国数字图书馆标准规范建设》[19]科技基础性工作专项资金重点项目在近年也着手开始制定数字图书馆服务标准规范。国家863计划"中国数字图书馆工程"[20]提出了把数据仓库技术应用到数字图书馆建设中,工程的一个重要部分就包括建立分布式存储、集中式管理的大型数据仓库,并对其进行智能化的管理与挖掘,再通过个性化和智能化的人机交互界面实现网络信息服务。

目前,国内数字图书馆个性化服务的应用仍处于探索阶段,一些相关的项目、课题仍在进展中,但个性化服务也得到了一些初步的应用。

1. 深图朗思 ILAS

此系统于2000年初推出,国内很多图书馆都采用此系统,如华南师范大学、汕头大学图书馆等。该系统中有一个为用户提供个性化服务的模块[21],由个人书架、借入书架、预约/预借书架、我的咨询、荐购书架、注册档案、财经档案、服务档案等部分组,涉及注册管理、读者证管理、读者服务查询、预约服务、检索服务、读者荐购、参考咨询等方面。具体功能包括:

(1) 邮件催还服务:当用户借阅的书籍将要在五日之内到期时,"邮件催还服务"就会通过用户提供的E-mail地址发送提醒催还单;当读者已经超期了,"邮件催还服务"还会每五天催还一次。

(2) 入藏新书邮件推送服务:入藏新书邮件推送服务是以E-mail的方式定期向本校读者推送其感兴趣的本馆入藏新书。系统登录的方法和"读者信息查询"系统一样。用户可以选择哪类入藏的新书或哪位作者的入藏新书。

(3) 预约:当用户想借阅的书籍已经全部借出时,用户可以进行预约,这样就可以在书籍还回来的时候去索取。

(4) 续借:用户可根据自身需要,自行在网上续借图书。所借图书在无人预约的情况下方可续借。

除了用户可以得到书目查询、新书通报、联合目录等普通的服务外,系统还能够根据用户的兴趣爱好列出其感兴趣的新书。

2. 浙江大学的 MyLibrary@ZJU 系统

浙江大学图书馆开发研制的个性化服务系统采用目前主流的服务模式,用户通过支持的浏览器登录,设置其账号和密码,并根据自己的知识结构、信息需求及对馆藏数字资源和其他网络资源

进行筛选、整理。用户完成设置后，即可动态建立页面，显示定制内容。该系统同时也提供面向校外用户个性化定制服务的版本，有“最新资源”、“新书通告”、“数据库”、“图书馆通知”、“我的收藏夹”等主要栏目[22]。

（1）定制资源：包括本馆所有网络数据库、电子期刊、新书书目等。

（2）与搜索引擎衔接：MyLibrary 提供一些著名的搜索引擎列表，用户只要选择其中一种并输入关键词，浏览器便自动跳转至该搜索引擎的搜寻结果页面。

（3）最新消息：在用户访问 MyLibrary 页面时，浏览器会弹出一个窗口，通告图书馆最新动态和最近投入使用的数字资源信息。

（4）定制页面：用户可对界面的布局和风格进行设置。

（5）书签功能：允许用户挑选自己常用的 Web 页面 URL 地址放入书签。

3. 中国人民大学的“我的图书馆”

中国人民大学信息学院与图书馆合作开发的“数字图书馆个性化服务系统”包括数字资源检索、个性化推荐、在线咨询三个子系统。

（1）“数字资源检索系统”为馆藏中、外文图书、网络数据库论文提供了一个统一、集成的用户查询界面，检索点全面、检索方式多样。

（2）“个性化推荐系统”能够根据用户兴趣偏好而主动地向用户推荐图书或论文资料。用户可以浏览资源的基本信息，查询其借阅状况，并可直接阅读全文，同时可以对历史推荐资源进行组织和管理。

（3）“在线咨询系统”为用户提供在线、实时的咨询服务，读者不用到馆，就可以获得即时的服务。

4. 中国科学院文献情报中心的“我的数字图书馆”

中国科学院文献情报中心建立的“我的数字图书馆”[23]，是基于北卡罗莱纳州立大学的 MyLibrary@ NCstate 系统开发。由于其面向社会公众提供服务，已成为基于个性化定制的门户网站。除了个性化定制以外，和北卡罗莱纳州立大学一样，该系统也提供最新消息、我的书架、我的链接等功能。

1.4.3 国内外个性化服务系统比较分析

根据以上列举的国内外的数字图书馆个性化服务方面比较成功实践，各选两个具有代表性的案例进行剖解和分析，分别为北卡罗莱纳州立大学图书馆、康奈尔大学图书馆、中科院文献情报中心数字图书馆、浙江大学图书馆。从以上具有代表性的国内外数字图书馆个性化服务的实践来看，可以发现其在服务的开展上有几个共同之处。

1. 服务方式

“由于开展服务的时间都不长，除了最初的各种创意之外，尚没有太大的突破[24]。”各图书馆一般都能提供资源定制、界面选择、邮件推送等服务，而在理论上很新颖的提法，由于技术上的限制，加之网络安全和用户隐私的考虑，还没有完全投入实用。

2. 服务理念

源于图书馆一直以来的读者至上的优良传统，无论是国外还是国内的图书馆界，对个性化服务的引入都抱着一种积极的态度。从服务的设想到服务的优化，图书馆都能够有计划地开展用户调查，确保服务不流于形式。

3. 服务开发

在服务开发上，各馆均比较注重和专业组织的合作关系。在以上列举的成功案例中，有很多系统都是图书馆联合其他组织机构实现的共同开发。采用这种团队合作的模式，不仅加快了开发的时间，更加提高了系统的专业性、广泛性。

4. 系统体系结构与开发技术

虽然各系统的开发思路基本一致,但在具体实现技术上,都有优劣。

国外的 MyLibray 系统在技术上大多采用 Perl 和 Java 通过 CGI 结合后台数据库的方式。例如北卡罗莱纳州立大学图书馆的 Mylibrary@ NCstate 系统,其本质上是一个数据库应用软件,该系统建立在 UNIX 操作系统上。数据库应用程序是 MySQL。Perl 作为 CGI 脚本语言,将 MySQL 数据库和 HTTP 服务器捆绑起来。

康奈尔大学的 MyLibrary@ Cornell 采用 Java 动态访问 Oracle 数据库的模式,客户端的浏览器内容为 Java 动态创建,与服务器相连的 Oracle 数据库存储用户的各种参数信息。这种系统结构由于具有良好的稳定性、可靠性、安全性和高效的服务能力,并且能够跨多个操作平台,是目前开发企业级产品的常用选择。

5. 服务种类

除一些最基本的服务之外,各个 MyLibrary 系统提供的服务各有千秋[25,26]。

以上 MyLibrary 系统都提供了资源定制功能,这也是 MyLibrary 系统应提供的基本功能,也就是让用户根据自己的信息需求建立起"我的图书馆",创建不同的文件夹来组织资源,并且可以根据需要随时随地对文件夹中的资源进行命名、添加和删除等操作。

其次,有些 MyLibrary 系统基于向用户提供全面便捷服务的目的,即用户进入自己的 MyLibrary 系统就如进入图书馆网站一样,可以享受图书馆网站所提供的各项检索服务,因此一般都提供了与图书馆自动化系统 OPAC 的接口,北卡罗莱纳州立大学的 MyLibrary@ NCstate 和康奈尔大学的 MyLibrary@ Cornell 都可以直接查阅图书馆目录,MyLibrary@ NCstate 系统还可以查询用户的借阅记录。

个人链接收藏或叫书签功能也是 MyLibrary 系统所能提供的主要服务之一,它类似于浏览器提供的 Bookmark 功能,即将自己经常访问的 URL 地址收藏到书签中,又有它自身的特点:浏览器的 Bookmark 功能只限于在某一台计算机上使用,但 MyLibrary 所提供的书签功能则在任何一台机器上只要登录到"我的图书馆"中就可以使用。

另外各个 MyLibrary 系统在设计中还提供了一些比较有特色的服务功能,如 MyLibrary@ NCstate 系统中的本站快速检索,可以对本站全部资源或某一学科资源进行站内检索,而且其检索结果通过主题可以链接到其他的相关资源。另外它还可以与著名的搜索引擎连接,用户选择某一搜索引擎并输入关键词,浏览器便会自动跳转至该搜索引擎的搜寻结果页面,此系统还提供了检索结果与图书馆的馆际互借和全文传递系统的链接,检索到的信息如没有全文,则可以链接到全文传递系统并提交全文传递请求。

相对而言,国内的数字图书馆个性化服务系统虽然也提供了基本的服务功能,但深层次的信息服务还没有广泛开展,这就需要广大的图书情报工作者将这项工作重视起来,利用现有的技术对个性化服务进行深入地挖掘与开发。

对未来图书馆服务的发展趋势是资源的数字化和服务的个性化[27],鉴于数据仓库、数据挖掘和联机分析处理技术在数据的组织、分析与知识发现等方面存在的巨大潜力,学术界普遍认为其可为数字图书馆的个性化服务提供关键技术。

1.5 参考文献

[1] 林宇. 数据仓库原理与实践. 北京:人民邮电出版社, 2002.

[2] 邵峰晶,于忠清. 数据挖掘原理与算法. 北京:中国水利水电出版社,2003.

[3] 高洪深. 决策支持系统. 北京. 清华大学出版社,2004.

[4] 刘景宇. 图书馆学五定律、信息资源共享四定理和图书馆学 2.0 五定律. 图书馆,2007(5):14 - 17.

[5] 陈海英. 关于数字图书馆个性化服务现有问题的思考. 图书情报工作,2003(4):65 - 67.
[6] 杨林,茅玉蓉. 个性化:定制你的网络服务. 软件工程师,2003(7).
[7] 徐小林. 程时瑞信息过滤技术和个性化信息服务. 计算机工程与应用,2001(11):545 - 550.
[8] 曹学校,秦莲霞,孟遂珍. 国外数字图书馆发展概况. 中国图书馆学报,2002,28(1):67 - 68,84.
[9] 郑惠伶. Cornell 大学图书馆个性化服务方式——MyLibrary. 图书馆学刊,2003(5):59 - 60.
[10] 安结. 网络中个性化服务及其在国外应用实例. 现代情报,2003(7):74 - 75.
[11] http://mylibrary. cornell. edu.
[12] 刘勇敏. 美国大学图书馆的个性化信息服务简介. 高校图书馆工作. 2004 (5).
[13] 何剑峰. 混合图书馆个性化信息服务模式. 图书馆理论与实践,2003,(2):45.
[14] http://www. lib. ncsu. edu/.
[15] 郭海明,刘昆雄. 数字图书馆个性化服务方式综述. 津图学刊. 2003(1):33 - 36.
[16] http://www. lib. virginia. edu/.
[17] 孙正东. 我国数字图书馆建设现状分析与思考. 情报资料工作,2002(2):42 - 45.
[18] 程家华. 数字图书馆发展动态及其现状分析. 现代图书情报技术,2002(2):11 - 13.
[19] 中国数字图书馆标准与规范建设. http://cdls. nstl. gov. cn/cdls2/w3c/.
[20] 刘峰. 国家 863 计划——构筑中国数字图书馆. 中国计算机用户报,2000(22):10 - 12.
[21] http://www. ilas. com. cn/ilasIIImodule6. asp.
[22] http://210. 32. 137. 206/mylib/mylib/index. asp.
[23] http://mylibrary. csdl. ac. cn/index. htm.
[24] 杨宗建. 对网络上隐私权法律保护的思考. 集美大学学报,2003(6):57 - 61.
[25] 卢共平. 数字图书馆的个性化信息服务. 图书情报工作,2002(8):24 - 27.
[26] 吴慧华,李禄香,刘泉峰. 网络环境下高校图书馆个性化信息服务. 图书馆学刊,2004(2):31 - 32.
[27] 马文峰. 数字图书馆个性化信息服务的探索. 图书馆杂志,2003,22(5):30 - 32.
[28] (加)Jiawei Han,Micheline Kamber 著. 数据挖掘概念与技术. 范明,孟小峰,等译. 北京:机械工业出版社,2003.

第 2 章　数字图书馆概述

随着信息技术的发展，需要存储和传播的信息量越来越大，信息的种类和形式越来越丰富，传统图书馆的管理机制显然不能满足这些需要。因此，人们提出了数字图书馆的设想。数字图书馆是一个电子化信息的仓储，能够存储大量各种形式的信息，用户可以通过网络方便地访问它，以获得这些信息，并且其信息存储和用户访问不受时间和地域限制。

2.1　数字图书馆的起源与发展

数字图书馆是传统图书馆在信息时代的发展，它不但包含了传统图书馆的功能，向社会公众提供相应的服务。信息化、网络化、数字化，这一连串的名词符号其根本点在于信息数字化。同样电子图书馆、虚拟图书馆、数字图书馆，不管用什么样的名词，数字化也是图书馆的发展方向。

1. 数字图书馆的起源

关于数字图书馆的研究起源，应追朔到电子图书馆的研究起源。因为，数字和电子都只是信息存储的方式而已。数字图书馆的前期，也称为电子图书馆[1]，它包含一些电子模拟信息和资料。上海交通大学杨宗英教授认为："1992 年以前，人们多用电子图书馆；1992 年—1993 年间多数并行使用这两个术语；1994 年以后，使用数字图书馆的逐渐多起来[2]"。美国密安大学的研究者认为，数字图书馆可以定义为电子图书馆[3]。Fox 称[4]，1991 年—1993 年间"电子图书馆"这一术语逐渐向"数字图书馆"的转变，说明人们似乎更愿意使用后者，这是人们对数字网络、数字音频、与电子出版有关的数字视频等的兴趣越来越大的缘故。

纵览历史，许多图书馆学、情报学史上的研究者，包括 Watson Davis，Vannevar Bush 和 Fremont Rider 等一直都在努力用缩微技术创建微型图书馆[5]。这种基于缩微技术的微型图书馆是电子图书馆思想的最早期形态。在电子图书馆思想发展史上，美国学者 Vannevar Bush 占有显赫的地位。他作为美国总统的科学顾问和研究发展局局长，于 1945 年在《大西洋》月刊上发表了日后被广泛引用、转载、重印的著名论文《如我们所能想象的》(As We May Think)，文中提出了用名为 Memex 的桌面机械以类似于人脑的方式将文献加以存储、连接和检索的构想。Bush 被后人誉为"具有非凡想象力与创新精神的技术设计者和管理者"。他所设想的 Memex 成为日后几乎所有信息检索项目的试金石，并被尊为超文本技术(HyperText)的先驱。美国著名图书馆学家情报学家 F. W. lancaster 也曾在 1995 年撰写的一篇书评中指出："将电子图书馆的最早思想来源追溯至 V. Bush 显然是恰当的。"大约半个世纪之后，在美国加州大学 Chico 分校成立了一家用 Memex 命名的机构"Memex Research Institute"，该研究所的宗旨即为"开发电子图书馆，实现 Bush 博士的梦想"。Bush 观点的重要之处在于他的两个构想：首先必须有能及时得到所需信息的设备，其次是读者自己就能检索这些信息。可见，Bush 的"Memex"对个人用户的信息存取来说是一种理想的模型，他点燃了当时和后来许多图书馆员、文献学家、工程技术人员的智慧火花[7-9]。

1962 年，美国在西雅图举办的"21 世纪图书馆"的展览会上提出了"没有图书的图书馆"的观点，可以说这是电子图书馆的最先的舆论准备[10]。1969 年，美国国会图书馆正式发行 MARCII 机读目录，这是图书馆进入自动化的标志。1975 年，美国图书馆学家 R. W. Christian 出版了"Electronic Library：Biblio－graphic Databases：1975—1976"一书，书中首次提到了"Electronic Library"这个名词。在整个 20 世纪 60 年代～70 年代，对电子图书馆思想贡献最大者莫过于 J. C. R. Licklider，他在

1965 年完成的图书馆学史上的经典之作《未来的图书馆》中，不仅展望了 2000 年的图书馆，而且意识到在图书馆馆藏中使用数字存储技术的优越性。他提出的“关联索引”（Associative Indexing）及其他富于创新性的计算机检索概念，成为 60 年代中后期一些试验性示范项目探索的重点之一。70 年代末、80 年代初，F. W. Lancaster 在其专著《通向无纸情报系统》和《电子时代的图书馆与图书馆员》中描绘了电子时代图书馆的面貌和前景，但他本人并未明确提出电子图书馆这一术语并确定其内涵。1983 年，美国人 Hugh. F. Cline 和 Lo - raine. T. Sinnott 在其专著《电子图书馆——自动化对学术图书馆的影响》中使用了“电子图书馆”的术语，但本书也仅仅如其书名副标题所反映的那样，只是一部关于图书馆自动化在美国四所大学图书馆中应用和开展情况的专著。首次对电子图书馆这一概念给出明确定义的是美国人 K. E. Dowlin，他在 1984 年出版的《电子图书馆：前景与进程》一书中写到：“所谓电子图书馆是一个提供存取信息的最大可能性并使用电子技术增加和管理信息资源的机构”[11-13]。

数字图书馆这一名词的出现与美国政府提出兴建国家信息基础设施（NIT）和因特网的迅速普及处于同一时期[12]。数字图书馆的基础根植于整个 20 世纪 80 年代对联机情报检索的追求和探索，以及全文本、多媒体信息处理技术的成熟，其发端可归因于因特网出现后美国政府对信息基础设施的研究和投入。

2. 数字图书馆的发展

数字图书馆的发展，大约经历了三个阶段[14]。

第一阶段是图书馆自动化发展的初级阶段，即图书馆自动化管理集成系统发展阶段。这一阶段从 20 世纪 60 年代末、70 年代初开始，以美国国会图书馆正式发行 MARCII 型的机读目录为标志。

第二阶段为图书馆在网上进行全球性、整体化的电子文献信息服务的新阶段。这一阶段发生在 1985 年左右，以 CD - ROM 光盘和局域网络开始在图书馆得到应用为主要目标。人们可以在图书馆、办公室、实验室甚至家中访问图书馆的目录、机读目录、局域网上的光盘数据库和文献检索系统。到了 20 世纪 90 年代，由于因特网的迅猛发展，更是将图书馆网上的电子文献信息服务推向了全球性服务的新阶段。

第三阶段是图书馆自动化的高级发展阶段，也称数字图书馆阶段，这是发生在 20 世纪末 ~21 世纪初。这一阶段国内外掀起了数字图书馆开发的热潮，一大批数字图书馆如雨后春笋般的出现，使人们实现了“秀才不出门，能知天下事”的梦想。

2.2 数字图书馆的含义和特征

关于何谓数字图书馆，目前国内外尚没有形成统一定论，这是由于研究的出发点不同，有的专门从事理论研究，有的从事应用研究，有的从事技术研究，而且数字图书馆研究处于发展阶段，它涉及多门学科交叉范畴，包括计算机科学、网络技术、通信技术、图书情报学、信息科学等。总之，数字图书馆是网络技术、多媒体技术、计算机技术与图书情报等多门技术发展的产物，进行数字图书馆研究的都有不同定义。

2.2.1 数字图书馆的含义

自从第一台电子计算机诞生以来，有识之士便热衷于讨论这样一个设想：对现有的传统图书馆可以起到补充、加强功能作用的，甚至最终可以取而代之的计算机化的“图书馆”。在长达几十年的讨论中，有许多名字和定义都在历史舞台上扮演过特定角色，“数字图书馆”是这些名字和定义中在最近几年所流行的名称，可以预料它也会在某一时刻被更新的名字所取代。

定义一：1995 年召开的美国联邦信息基础设施技术与应用项目（Information Infrastructure Tech-