

第一部研究信息资源门户问题的专著

信息资源门户与利用

XINXI ZIYUAN
MENHU YU LIYONG

李世玲 著

中国大地出版社

信息资源门户与利用

李世玲 著

中国大地出版社

· 北 京 ·

内 容 提 要

本书系统论述信息资源门户的理论和方法、门户构建、技术系统、语义网、特殊文献检索、标准规范及信息服务工作等问题。可供信息产业及服务部门的研究人员参考。

图书在版编目 (CIP) 数据

信息资源门户与利用/李世玲著. —北京: 中国大地出版社, 2007. 1

ISBN 978-7-80097-903-3

I. 信... II. 李... III. 信息管理—研究
IV. G203

中国版本图书馆 CIP 数据核字 (2006) 第 151787 号

责任编辑: 李 颖

出版发行: 中国大地出版社

社址邮编: 北京市海淀区学院路 31 号 100083

电 话: 010—82329127 (发行部) 010—82329008 (编辑部)

传 真: 010—82329024

印 刷: 北京北林印刷厂

开 本: 850mm × 1168mm 1/32

印 张: 8.25

字 数: 210 千字

版 次: 2007 年 1 月第 1 版

印 次: 2007 年 1 月第 1 次印刷

印 数: 1—1500 册

书 号: ISBN 978-7-80097-903-3/F·184

定 价: 25.00 元

(凡购买本社图书, 如发现印装质量问题, 本社发行部负责调换)

前 言

信息资源门户与利用问题是随着信息化建设飞速发展而出现的新事物，是适应网络化、数字化对信息资源管理的理论与实践的挑战而产生的一种新的信息组织与利用形式，目前已成为信息资源管理的核心问题，其理论、方法和技术也成为信息资源提供机构和服务机构的核心竞争力。

“门户”（Portal）来源于中世纪的拉丁语 Portal，意思是“城门”。自 1997 年后，“门户”这个词在网络时代相当流行，主要应用于商业领域，尤其是一些大公司，为雇员提供单一信息查询界面，这样他们可以访问来自多个计算机系统的公司信息。1999 年，“门户”这个概念进入图书情报领域，相当于获取信息资源的“入口”或者“接口”。最早的一个门户是于 1998 年北卡罗莱那州立大学图书馆发布的 My library (my. Lib. ncsu.edu)。早期的图书馆门户更多地是体现在图书馆为读者的个性化服务上。目前较具代表性的门户网站是美国研究图书馆协会在 2000 年建立的 ARL (Association of Research Libraries, 简称 ARL) 学术门户。该门户对检索发现工具的要求是：能够检索各种类型的元数据。并确定了门户的 5 个功能：①发现功能 (Discovery)，为研究人员提供发现、检索到的信息资源；②获取功能 (Capture)，能够得到检索到的信息资源；③加工管理功能 (Manipulation)，即能够对文本进行处理以及对摘要进行管理；④发行功能 (Distribution)，即能够进行投稿及出版；⑤咨询功

能 (Consultation), 能够提供在线参考咨询。另外, 1998 年美国科学基金会启动的美国国家科学数字图书馆门户 (NSDL) 以及 1999 年投入使用的康奈尔大学图书馆开发的网络个性化服务平台 My library 也较具有代表性。

我国信息资源门户建设相对比较落后, 但也有较具代表性的项目, 如 2001 年正式启动的中国科学院知识创新工程重大项目 (CSDL)。CSDL 门户提供的网络服务主要包括: 电子期刊集成目录、数据库跨库检索系统、网络用户参考咨询系统等, 基本上包括了门户的几个功能要素, 但在各个要素的功能方面还不完善, 比如其统一检索 (数据库跨库检索系统) 目前只能提供 7 个数据库资源的集成检索, 还不是真正意义上的统一检索平台; 其学科信息门户目前只完成了 5 个学科; 其虚拟咨询还仅限于提供表单咨询而未开发实时咨询。但总体而言, CSDL 仍是国内数字图书馆门户的代表性项目。

随着网络用户呈几何级数增长, 网站数量不断增加, 人们对网络的了解也不断深入, 要求的标准也越来越高, 信息资源门户中的一些缺陷和瓶颈也都一一暴露出来, 如: ①统一检索问题。目前有的统一检索还不能对结果集进行处理, 检索之后只能按库显示, 能对结果集进行处理的也都只是部分处理, 不能对所有的结果集进行去重、排序。②集成性问题。信息资源门户建设一方面要考虑现有的图书馆信息资源和服务基础, 同时也要考虑与其他机构或与其他系统的整合, 强调合作机制和联合运行模式, 长远考虑与更大的数字化环境进行集成, 共享数字化资源、服务、系统及管理。③导航性问题。门户尤其应注重信息资源整合与导航能力, 注重提供完善的和精选的信息服务, 使用户能借助门户的强大导航功能任意获取所需要的信息。④标准规范问题。标准规范应考虑页面流程规范、页面技术规范、服务标准规范、门户元素规范、数据质量规范以及元数据标准等。信息资源门户要展

示和调度方方面面的资源和服务,其流程和界面都要根据预先制定的标准进行控制、管理和建设,同时其后台的技术与环境、支持平台等也应遵守国际通用的标准。标准规范化建设和管理将有利于数字化资源和服务的共享,并有利于信息资源门户与其他数字化环境的集成。⑤虚拟化问题。门户应创建虚拟的信息资源环境,提供数字化服务,如远程访问、同步咨询等。⑥个性化问题。个性化服务功能是门户的基本功能之一,门户应能提供定制服务,信息推送、互动式信息服务等,帮助用户建立个性化门户。当然,作为门户最根本的问题是要解决为每一位用户提供所需要的信息和为每一条信息找到需要它的用户,如何让两者更好地结合以提升用户的满意度和减少信息的冗余成了迫在眉睫的问题。

本书是我近几年致力于信息资源门户与利用问题所做的一些事情。在网络遍及全球,门户大量兴起的信息时代,有必要对信息资源门户与利用问题作一系统梳理。本书的结构主线是第1章,理论前提和方法。谈了3个问题:①信息资源门户概述;②信息资源门户的研究内容;③信息资源门户的发展基础。第2章,门户构建。谈了5个问题:①数字资源门户的构建;②以用户为中心的图书馆门户的构建;③图书馆门户的界定;④混合图书馆门户研究;⑤城市区域性图书馆门户的构建。第3章,管理系统。谈了5个问题:①大学机构库的构建;②基于信息生命周期的数字资源管理系统;③大学数字图书馆的评价模式;④日本国立大学数字图书馆;⑤英国高校信息资源的提供机构。第4章,检索系统。谈了3个问题:①集成信息检索系统;②基于全景检索模型的多媒体检索系统;③网络信息检索技术的评价。第5章,语义网研究。解读语义 Web。第6章,特殊文献检索。谈了3个问题:①专利信息检索;②国外医学文献检索;③专利信息检索的用户教育。第7章,标准规范。谈了4个问题:①XML 研究;②MODS 研究;③ISO/TC46 的标准化工作;④COUNTER 项目。

第8章,信息服务工作。谈了2个问题:①信息服务工作的发展变革;②信息服务工作的条件。其中尤以门户构建和检索为重点,因为门户的服务宗旨是方便用户,而优质的门户提供的检索是信息使用的必备手段。另外,本书突出了信息的“海量”特性,这是数字化时代的基本特点,是今后技术发展的方向。本书还提供了门户的构建方法、技术及应注意的问题;提供了文献检索的参考方案、方法、途径等,使得枯燥的技术可以和实践结合起来,对读者更是一本指南性的读物。

我之所以选择这些题目来写,是因为我在这些方面有过一些思索,想发表一点看法。有些想法可能是不成熟的,谈出来与专家、读者商榷。

出版这本书的目的是想向学术界的同行们说明,我有志于进行信息资源门户与利用问题的研究,并将自己曾做过的一些事情报告给大家。至于这些事情做得怎么样,敬请各位专家批评指正。

本书在撰写过程中得到河北师范大学的支持,本研究为河北省哲学社会科学规划立项课题成果(编号:200605010)。

李世玲
2006年10月

目 录

前 言	(1)
第1章 理论前提和方法	(1)
1.1 信息资源门户概述	(2)
1.2 信息资源门户的研究内容	(13)
1.3 信息资源门户的发展基础	(14)
第2章 门户构建	(18)
2.1 数字资源门户的构建	(18)
2.2 以用户为中心的数字图书馆门户的构建	(25)
2.3 图书馆门户的界定	(32)
2.4 混合图书馆门户研究	(41)
2.5 城市区域性图书馆门户的构建	(46)
第3章 管理系统	(58)
3.1 大学机构库的构建	(58)
3.2 基于信息生命周期的数字资源管理系统	(66)
3.3 大学数字图书馆的评价模式	(75)
3.4 日本国立大学数字图书馆	(87)
3.5 英国高校信息资源的提供机构	(91)
第4章 检索系统	(98)
4.1 集成信息检索系统	(98)
4.2 基于全景检索模型的多媒体检索系统	(106)
4.3 网络信息检索技术的评价	(113)

第 5 章 语义网研究	(120)
5.1 语义 Web 概述	(120)
5.2 以书目数据为例解读语义 Web	(123)
第 6 章 特殊文献检索	(135)
6.1 专利信息检索	(135)
6.2 国外医学文献检索	(149)
6.3 专利信息检索的用户教育	(156)
第 7 章 标准规范	(163)
7.1 XML 研究	(163)
7.2 MODS 研究	(172)
7.3 ISO/TC46 标准化工作	(183)
7.4 COUNTER 项目	(190)
第 8 章 信息服务工作	(199)
8.1 信息服务工作的发展变革	(199)
8.2 信息服务工作的条件	(204)
参考文献	(212)
附录	(223)
1. 国际图书馆联盟的最新动向	(223)
2. 荷兰发展图书馆事业的启示	(231)
3. 参考电子文献的现状分析	(240)
4. 纸载体和电子载体的比较	(249)

第 1 章 理论前提和方法

近年来,由于 Web 技术的发展,带有信息资源门户的词语在自然科学、社会科学以及人们的日常生活中频繁出现,如带有信息的词语:信息服务、信息管理、信息技术、信息检索、网络信息、多媒体信息、学术信息、文献信息、信息工程、信息分析、信息存储、信息传播、信息处理、信息知识、信息文化、信息产业、信息世界、经济信息、商业信息等等,用信息开头或者结尾的词语广泛出现。带有资源的词语也同样,如土地资源、矿产资源、森林资源、海洋资源、石油资源、能源资源、电力资源、人力资源、资源共享、资源共建等。带有门户的词语也很多,如数字图书馆门户、图书馆门户、专业图书馆门户、知识门户、学科信息门户、专业信息门户、信息门户、生命科学门户、门户网站、企业信息门户、政府门户、网格门户、My Library 等。

使用上述词语的意图作者分析大概有两种可能:一种是象征性地表现特定的社会现象或内容,以引起人们对该社会现象的关注。第二种可能也没有什么实质性的内容,就是为了赶时髦,追新潮,反正上述词语大多数都没有严格的定义,跟着感觉使用好了。无论出于何种使用目的,信息资源门户如此被人们看重、使用,说明信息资源门户实实在在就存在于我们周围,与社会的种种现象密切相关。那么,什么叫信息资源门户?信息资源门户的构造如何?应该承担起怎样的任务?需要有一个明确的认识。

1.1 信息资源门户概述

1.1.1 信息资源门户的概念

要论及信息资源门户,则不得不先界定信息、资源以及门户三者的相互关系。

1 信息

人类意图对信息(information)作科学准确的表述始于20世纪20年代,据不完全统计,目前关于信息的说法已超过100种。^[1]如果在众多的说法中追根溯源,1922年卡森(J. R. Carson)以及1928年哈特利(V. L. Hartley)的工作为信息概念的诞生作出了重大贡献。哈特利把信息设想为代码的序列,而不考虑代码所代表的语义,从而把信息量(H)定义为 $H = N \log_2 S$,即由S代码序列,取N个码,构成 2^N 个可能的信息。^[2]1948年克劳德·仙农在哈特利的信息量只考虑“有多少个可能性”,即等概率值的基础上加以发展,以一篇《通讯的数学理论》宣告了信息论的诞生^[3],克劳德·仙农在该文中提出了更为广泛的信息量概念,其信息量的出发点是“这样或那样事件有多少程度的可能性”。显然克劳德·仙农的信息量概念比哈特利的信息量概念要广,它既包括信源符号是等概率的情况,也包括信源符号是不等概率的情况。但是,克劳德·仙农只考虑了信息的数量方面,而忽视了信息的质的方面。美国控制论创始人N·维纳认为:“信息就是我们在适应外部世界和控制外部世界过程中,同外部世界进行交换的内容的名称。”“信息就是信息,不是物质也不是能量。”^[4]这个定义虽然笼统,但概括性大,N·维纳是从信息的数量、含义、价值以及机器、思维的角度来讨论信息的本质,显然不是狭义的而是广义的信息概念。

下面几部有影响的辞书都是在克劳德·仙农和 N·维纳信息定义的基础上解释信息的。日本《广辞苑》对信息的解释是：“信息是所观察事物的知识。”英国牛津字典的定义是：“信息就是谈论的事情、新闻和知识。”美国韦伯字典中的信息定义是：“用来通信的事实，在观察中得到的数据、新闻和知识。”《词海》中的信息是指：“对信息接受者来说预先所不知道的报道。”《人工智能辞典》则区别了信息和传递信息的载体，并从信息论的角度作出定义：“信息是数据或消息中所蕴含的定义。它与信息载体无关；在信息论中，信息使消息中所描述事件出现的不确定性减少，若不提供信息，则不确定性就大一些。”这是狭义信息的解释。关于信息还有从本质属性或人对物质世界反映的角度广义的认识，如《中国大百科全书》对信息的定义就有狭义和广义两方面的解释：“按照狭义的理解，信息是用来消除不定性的东西。按照广义的理解，又有两种认识：从本体论意义上说，信息泛指一切事物（物质的、精神的）运动的状态和运动的方式，包括事物内部结构的状态和方式以及外部联系的状态和方式；从认识论意义上说，信息是关于事物运动状态和运动方式的反映。”由于信息定义不断向各个学科内渗透，其说法还有：

(1) 信息是物质的一种属性，是物质运动的形态和物质存在的特征；^[5]

(2) 信息是物质的一种普遍属性，是物质存在的方式和运动规律与特点；^[6]

(3) 信息是物质和意识的综合，是中间物；^[7]

(4) 信息离不开物质作用载体而存在、传递，同时，离开反映机能对它的接受，它也不是什么信息、信号，而只是一种物质作用；^[8]

(5) 信息是事物运动的状态和（状态改变的）方式；^[9]

(6) 信息是对客观世界中各种事物运动变化和特征的反映；

是客观事物之间相互作用和联系的表征；是客观事物经过感知或认识后的再现^[10]等。

显然，上述定义的解释大部分偏重于信息对人类社会的作用和观察事物的属性方面。由于各自的研究领域具有不同的性质、内容、目的和方法，各学科、各学派所理解的信息概念也不尽相同，这里，我们不强求对信息的统一理解、统一认识，只求用统一的、联系的观点来看待某些现象，从新的角度探索新的真理。综上所述，由于信息是客观事物的反映，这个反映是通过人类感知或认识后用手势、口语、声音、图形、图像、文字、数据等形式表现出来的东西，所以，作者认为下面的定义更能恰如其分地揭示信息的本质含义：信息是认识主体（人或计算机）感知或认识客观世界中一切事物（物质的、精神的）的反映。

2 资源

资源是一个动态的概念，是随着人们认识能力和生产力水平的提高而不断发展的。西方学者对“资源”论述，最具代表性的是美国的阿兰·兰德尔（Allan Landall）。他在《资源经济学》一书中提出，资源可分为两个范畴，一是自然界赋予的自然资源；二是人类社会中的劳动创造的各种资源。自然资源是自然界进入生产过程的各种物质，但没有被人们认识，没有与知识结合的自然物质不能成为资源，只有在被人们认识以后才能成为资源和财富。

根据我国的传统观念，资源又称为“财源”，即资财之源，财富之源。《辞海》对“资源”释义为：“资财的来源，一般指天然的财源。”《现代汉语词典》对“资源”的解释是：“生产资料或生活资料的天然来源。”孟广均在《信息资源管理导论》一书中将资源定义为：“通过人类的参与而获取的（或可获取的）可利用的物质、能量与信息的总合。”^[11]随着科技的不断进步和生产力水平的不断提高，自然资源的种类不断扩大，可以是

地球上一切有生命或无生命的资源。不过通常只是指一定的技术、经济环境中对人类有益的资源。社会资源是人类自身通过劳动创造的资源,除人力、物质资源外,科学、技术、信息和管理等,日益成为重要的社会资源。

综上所述,我国学者一般认为,所谓资源,是指在自然界和人类社会中一切可以用来创造物质财富和精神财富的客观存在形态。

3 信息资源

信息资源是信息与资源两个概念整合在一起衍生出来的复合概念,它具有信息和资源的两面性,他属于一种资源,但其本质是信息。信息可成为资源,但究竟什么是信息资源?国内外学者有不同的见解。

国外最具代表性的观点是霍顿(F. W. Horton)1979年对单数形式和复数形式的“信息资源”的解释:当“资源”为单数时,信息资源是指某种内容的来源,即包含在文件和公文中的信息内容;而当“资源”为复数时,信息资源是指支持工具,包括供给、设备、环境、人员、资金等。^[12]1986年,霍顿与其合作者马尔香(D. A. Marchand)又进一步区分了“信息资源”和“信息财产”两个概念,认为信息资源的含义是:拥有信息技能的个人;信息技术及其硬件与软件;诸如图书馆、计算机中心、传播中心、信息中心等信息设施;信息操作和处理人员。^[13]

国内的学者也对信息资源的含义进行了广泛和深入的研究。1991年,孟广均在《知识工程》杂志中明确指出:“信息资源包括所有的记录、文件、设施、设备、人员、供给、系统和搜集、存储、处理、传递信息所需要的其他机器。”^[14]

乌家培认为:“对信息资源有两种理解。一种是狭义的理解,即指信息内容本身。另一种是广义的理解,指的是除信息内容本身外,还包括与其紧密相连的信息设备、信息人员、信息系统、

信息网络等。”^[15]

马大川认为：“广义的信息资源是指信息和它的生产者及信息技术的集合。即广义的信息资源由3部分组成：第一，人类社会经济活动中的各类有用信息。第二，为某种目的而生产有用信息的信息生产者。第三，加工、处理和传递有用信息的技术。狭义的信息资源则仅仅指人类社会经济活动中经过加工处理有序化并大量积累后的有用信息的集合，它包括科学技术信息、政策法规信息、社会发展信息、经济信息、市场信息、金融信息等多方面内容。”^[16]

朱立文认为：“信息资源是反映客观事物的各种信息的总称，也是各种信息的集合体。”^[17]

汪华明和杨绍武认为：“信息资源是将信息通过在生产、流通、加工、存储、转换、分配等过程中，作用于信宿（用户）进行开发利用，为人类社会创造出一定财富而成的一种社会资源。”^[18]

马费成认为：“信息资源是指人类社会活动中积累起来的以信息为核心的各类信息活动要素（信息技术、设备、设施、信息生产者等）的集合。这里的信息活动包括围绕信息的搜集、整理、提供和利用而开展的一系列社会经济活动。”^[19]

霍国庆认为：“信息资源也就是可以利用的信息的集合……换言之，信息资源是经过人类开发与组织的信息的集合。”^[20]

综上所述，信息资源可从狭义和广义两个角度理解，狭义的信息资源是指人类在信息活动以及其他社会实践活动中形成的大量序列化的有用信息的集合。也就是说信息资源包括信息内容；信息生产者、提供者、服务者及其用户；信息技术和其他信息活动要素（如信息设备、设施、经费等）。信息资源门户所说的信息资源是广义范围的。

4 门户及信息资源门户

“门户” (Portal) 来源于中世纪的拉丁语 Portal, 意思是“城门”。一般认为, 门户是呈现给门户使用者的单一的公共入口点, 通过这个入口点, 门户网站的使用者可以根据各自的角色和权限进行各种操作。^[21] 在网络时代, 由于门户这个词流行于各行各业, 也就出现了各种门户定义, 如学科信息门户 (也称主题门户或主题网关) 是按照一定的资源选择和评价标准、规范的资源描述和组织体系, 将特定学科或主题领域的信息资源、工具与服务集成到一个整体中, 对具有学术价值的网络资源进行搜索、选择、描述、组织和整合, 并提供浏览、检索、导航等增值服务的专门性信息门户。^[22] 知识门户是一种智能应用, 它建立了用户和知识之间的联系, 通过提供个性化和自适应的交互手段, 帮助用户方便地找到需要或知识载体 (如专家) 并与之交流, 它集成已有的应用和技术, 帮助用户主动发现知识, 或通过分析用户的使用习惯和兴趣, 向用户推荐相关知识或知识载体。其主要功能是管理和组织相关资源, 自动寻找这些资源的关联关系, 为用户提供快速而准确的知识导航和交互环境。^[23] 学科门户是将特定学科领域的信息资源、工具和服务集成到一个整体中, 为用户提供一个方便的信息检索和服务入口。^[24] 图书馆门户是向读者提供信息与服务的通用入口及个性化接入点, 能对资源与服务做到无缝链接, 能从不同的分布资源获取内容, 并以统一的形式显示给用户。^[25] 数字图书馆门户是一种基于学术性信息资源的网络服务, 它运用分布式检索和整合结果集等技术, 从不同的分布资源获取内容, 并以统一的形式显示给用户。^[26] 企业信息门户是一个应用框架, 它将企业所有应用和数据集成到一个信息管理平台之上, 并以统一用户界面提供给用户, 使企业可以迅速地建立企业与企业和企业对内部员工的信息通道。^[27] 政府门户是将政府服

务以某种逻辑方式整合在一起进行重新组合,通过这种方式,居民只需一个入口,就可以享受整体的政府服务,不必再沿用一个个部门拜访的串行工作方式。^[28]

上述各种门户都是针对专业、行业或部门用户需要的特点,组织信息资源为用户提供网络化服务的接口。所以,信息资源门户实质是一个为用户提供信息资源存取的网关。该网关具备用户身份识别,并根据用户特征提供个性化以及用户定制的网络服务系统。通过门户,用户可以根据自己的兴趣、爱好方便地存取信息资源门户中的内容,使用门户提供的服务。

1.1.2 信息资源门户的类型

(1) 按门户的服务内容可分为两种类型:水平门户和垂直门户。水平门户(Horizontal Portal)是提供综合性信息和服务的门户网站,如提供新闻、搜索引擎、电子邮件、虚拟社区、聊天室等;垂直门户(Vertical Portal)是提供专业化信息和服务的门户网站,如学科信息门户、图书馆信息资源门户等属于垂直门户。^[29]

(2) 按门户的功能、服务对象可分为3种类型:通用门户、特色门户和个性化门户。通用门户(Main Portal)提供门户的核心服务,如检索、浏览、注册、登录和个人信息、帮助,以及可选门户部分,如公告消息和管理端。大型数字图书馆项目或者单个图书馆的数字图书馆门户一般属于通用门户。特色门户(Specialized portal)是为特殊的主题和特殊的读者群配置的资源和服务。如NSDL的科学图片库、中国的国家科学数字图书馆项目(CSDL)下属的学科信息门户、某些图书馆的知识门户、中国高等教育文献保障系统(CALIS)的特色库子项目门户等均属于特色门户。个性化门户(Personalized portal)是特定用户或者用户