

网络信息资源 组织和检索

赵志坚 主编
赵志坚 何平 杜方冬 编著

 人民邮电出版社
POSTS & TELECOM PRESS

网络信息资源组织和检索

赵志坚 主编

赵志坚 何 平 杜方冬 编著

人民邮电出版社

图书在版编目 (CIP) 数据

网络信息资源组织和检索/赵志坚主编; 赵志坚, 何平, 杜方冬编著.

—北京: 人民邮电出版社, 2004.5

ISBN 7-115-12229-6

I. 网... II. ①赵...②赵...③何...④杜... III. 计算机网络—情报检索

IV. G354.4

中国版本图书馆 CIP 数据核字 (2004) 第 039623 号

内 容 提 要

本书系统地介绍网络信息资源组织、存储和检索的基础知识和基本概念, 详细地阐述因特网网上信息资源的分布状况及网络检索工具, 描述国内外部分著名网络数据库的结构、资源及检索方法, 全面地介绍全文电子期刊、电子图书、电子报纸的相关知识及资源, 分析部分学科领域网络信息资源的综合利用, 初步阐述新时期信息服务部门如何利用网络信息检索来进行网络信息咨询、定题服务和科技查新等信息服务。

本书是一本网络信息资源组织和检索的著作, 可作为高等学校信息检索课程的教材, 也可作为广大网络用户与企业的科研人员、管理人员、工程技术人员从事文献信息管理应用的实用参考书。

网络信息资源组织和检索

- ◆ 主 编 赵志坚
编 著 赵志坚 何 平 杜方冬
责任编辑 邹文波
执行编辑 魏红梅
- ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号
邮编 100061 电子函件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
读者热线 010-67129259
北京汉魂图文设计有限公司制作
北京隆昌伟业印刷有限公司印刷
新华书店总店北京发行所经销
- ◆ 开本: 787×1092 1/16
印张: 22.25
字数: 534 千字 2004 年 5 月第 1 版
印数: 1-5 000 册 2004 年 5 月北京第 1 次印刷

ISBN 7-115-12229-6/TP · 3944

定价: 29.00 元

本书如有印装质量问题, 请与本社联系 电话: (010) 67129223

前 言

随着信息技术尤其是网络技术的快速发展,人类进入了一个崭新的网络信息时代。因特网成为了世界上最大的、信息种类繁多、覆盖面广的信息资源库。有人曾概括说网络时代是信息爆炸的时代。但同时,网络信息量大、分散无序、类型庞杂,人们处于浩大的网络信息海洋当中,常常找不到所需的有用信息。因此如何解决网上海量信息与用户检索信息间的矛盾显得尤为重要,从而引出了对网络信息进行有效组织和控制的研究。

网络环境下的信息检索所具备的多样性、灵活性远远超出了传统信息检索。在新的环境下每个用户都会感觉到由信息过载(Information Overloaded)而引发的困惑和茫然。在网络信息这个浩瀚、动荡的信息海洋中,如何能够准确、及时、有效地找到、获取与自身需求相关、切题和适用的信息对所有网络用户来说都是十分重要的,同时对每个用户也颇具挑战意味。网络信息检索代表着在网络信息环境下发展起来的一种新型检索模式,而且已发展成为信息检索领域的一个重要分支。随着网络信息检索实践的日渐丰富的发展,“信息检索”这一学科在新的信息环境下的拓展和应用具有十分重要的意义。

本书针对广大因特网用户利用网络获取有价值的信息资源的迫切需求,着重从网络信息资源组织、网络信息资源存储和网络信息资源检索三个方面阐述了相关概念和理论。本书具有以下特点。

1. 体系结构新。本书采用了检索技术和检索方法相结合,方法介绍和案例操作相结合,综合信息的检索与学科领域的信息检索相结合的体系结构。各章内容循序渐进,又有较强的独立性。

2. 检索方法全。本书介绍了丰富的网络信息资源、网络数据库、国内外联机数据库网络检索等众多非常实用的资源及检索方法和技术。

3. 编写目的实。读者系统地学习本书后,可以在基本层次上强化信息意识,提高信息敏感性,掌握并能熟练应用主要的信息组织、存储及检索技术,使读者成为既是精品信息的提供者,又是敏感、高效的信息接受者,从而更好地为社会信息化服务——成为信息技术最大限度上的应用者和受益者。

本书共 11 章,由湖南大学信息管理与信息系统系赵志坚老师、湖南大学图书馆何平老师和中南大学医学技术与情报学院的杜方冬老师共同合作完成。第 1 章由何平编写;第 2、11 章由杜方冬编写;第 3~10 章由赵志坚编写并任主编。

本书得到了姜灵敏教授和彭一中研究馆员的大力支持与帮助,在此表示衷心感谢。

由于作者时间匆促且水平有限,书中难免存在不足之处,敬请广大读者不吝批评、指正。

编 者

目 录

第 1 章 网络信息资源组织	1
1.1 信息组织概述	1
1.1.1 信息	1
1.1.2 信息组织	3
1.1.3 网络环境下信息组织的内涵	5
1.2 知识组织概述	5
1.2.1 知识的含义	5
1.2.2 知识组织的定义	7
1.2.3 知识组织的方法	8
1.2.4 知识组织的体系	9
1.3 网络信息资源组织	9
1.3.1 网络信息资源组织的环境	10
1.3.2 网络信息资源组织的模式	10
1.4 网络信息资源组织中存在的问题	17
1.5 网络信息资源组织的标准	18
1.5.1 标准规范相关组织机构	18
1.5.2 网络信息标准化技术委员会	18
1.5.3 元数据所涉及的标准和规范	19
1.5.4 网络信息资源标记格式的标准	22
1.5.5 网络信息资源内容格式的标准	23
1.5.6 网络信息资源描述语言的标准	26
1.5.7 网络信息资源组织评价标准	27
1.5.8 网络信息资源标准化体系的构成	27
1.6 网络信息资源组织与管理的发展趋势	28
1.7 网络信息资源组织的实际运用	29
1.7.1 电子期刊	29
1.7.2 电子图书	30
1.7.3 数字图书馆	31
1.7.4 电子商务网站	35
1.7.5 个人数字图书馆	38
1.8 小结	43
第 2 章 网络信息资源存储	44
2.1 信息资源存储	44
2.1.1 信息资源存储的意义和作用	44

2.1.2	信息资源存储的主要类型	44
2.1.3	信息资源存储的主要技术	45
2.2	网络信息资源存储	48
2.2.1	网络存储的含义及特点	48
2.2.2	网络存储主流技术	48
2.2.3	网络信息存储技术的比较	56
2.2.4	清华同方“NAS 教育资源共享存储系统”	56
2.2.5	信息存储技术的发展趋势	57
2.3	小结	58
第 3 章	信息检索基础	59
3.1	信息检索概念	59
3.2	信息检索的类型	60
3.3	检索语言基本知识	61
3.3.1	检索语言概述	61
3.3.2	检索语言的类型	61
3.4	信息检索系统	62
3.4.1	信息检索系统概述	62
3.4.2	信息检索系统的组成	63
3.4.3	信息检索系统的类型	64
3.4.4	评价检索系统的标准	65
3.5	信息检索的基本程序和方法	66
3.5.1	信息检索的基本程序	66
3.5.2	信息检索的基本方法	68
3.5.3	影响检索效率的因素	68
3.5.4	信息检索的基本策略	69
3.6	小结	70
第 4 章	网络信息检索	71
4.1	因特网信息资源	71
4.1.1	因特网上拥有的信息资源类型	71
4.1.2	网络信息资源的新特点	72
4.1.3	网络信息资源的优越性	73
4.2	因特网信息资源的评价	73
4.2.1	评价网络信息资源的必要性	73
4.2.2	评价网络信息资源的目的	74
4.2.3	网络信息资源评价的对象主体	74
4.2.4	网络信息资源评价指标体系	75
4.2.5	网络信息资源评价指标权重的求得	76
4.2.6	网络信息资源评价的步骤	77
4.2.7	网络信息资源评价的方法	78

4.2.8	网络信息资源评价站点	79
4.3	网络信息检索概论	81
4.3.1	计算机信息检索的发展	81
4.3.2	网络信息检索概述	83
4.3.3	网络信息检索的一般方法	84
4.3.4	网络信息检索的特点	85
4.3.5	网络信息检索标准 Z39.50	86
4.4	网络信息检索工具	89
4.4.1	网络信息检索工具的一般构成	89
4.4.2	网络信息检索工具的原理	91
4.4.3	相关网络信息检索技术	91
4.4.4	网络信息检索工具的类型和特点	93
4.4.5	网上检索工具发展趋势	95
4.5	Web 检索工具	97
4.5.1	搜索引擎	97
4.5.2	网络目录	107
4.5.3	元搜索引擎	112
4.5.4	智能搜索引擎	118
4.5.5	专门网络信息检索工具	120
4.6	非 WWW 检索工具	130
4.6.1	FTP 资源检索工具	130
4.6.2	邮件列表检索工具	133
4.6.3	Telnet 检索工具	136
4.6.4	用户网 Usenet 检索工具	137
4.7	小结	143
第 5 章	国内网络数据库	144
5.1	中国知识基础设施工程网站	144
5.1.1	中国知识基础设施工程概述	144
5.1.2	中国知识基础设施工程网站收录内容	144
5.1.3	中国知识基础设施工程检索	149
5.1.4	中国知识基础设施工程检索步骤	149
5.1.5	全文浏览器介绍	151
5.2	维普资讯数据库	152
5.2.1	维普资讯数据库概述	152
5.2.2	维普资讯数据库检索途径	153
5.2.3	VIP 全文浏览器介绍	159
5.3	万方数据资源系统	160
5.3.1	万方数据资源系统	160
5.3.2	万方数据资源系统数据库	161

5.3.3	万方数据资源系统检索	162
5.4	中国人民大学书报资料中心数据库	168
5.4.1	中国人民大学书报资料中心介绍	168
5.4.2	中国人民大学书报资料中心出版物的特点	168
5.4.3	中国人民大学书报资料中心产品介绍	169
5.4.4	中国人民大学书报资料中心检索	170
5.5	专利数据库	172
5.5.1	专利文献概况	172
5.5.2	中国专利文献网上数据库及其检索	178
5.6	中国经济信息网和国务院发展研究中心信息网	186
5.6.1	中国经济信息网	186
5.6.2	国务院发展研究中心信息网	187
5.7	中国高等教育文献保障体系	188
5.7.1	CALIS 介绍	188
5.7.2	CALIS 服务功能	189
5.7.3	CALIS 的子项目建设	190
5.8	中国资讯行	193
5.9	中文社会科学引文索引 (CSSCI)	195
5.10	小结	198
第 6 章	国外网络数据库	199
6.1	国外著名的检索工具	199
6.1.1	美国《工程索引》	199
6.1.2	美国《科技会议录索引》	205
6.1.3	《科学引文索引》	208
6.2	专业检索工具	214
6.2.1	《剑桥科学文摘》CSA 数据库	214
6.2.2	《科学文摘》INSPEC 数据库	217
6.2.3	《化学文摘》CA 数据库	218
6.2.4	EBSCO 学术信息、商业信息网络数据库	219
6.3	综合性大型联机检索系统简介	221
6.3.1	OCLC 的 First Search 联机检索系统	221
6.3.2	DIALOG 国际联机检索系统	227
6.3.3	其他联机系统介绍	243
6.4	特色数据库	248
6.4.1	美国政府科技报告数据库 (NTIS)	248
6.4.2	专利文献数据库	250
6.5	小结	255
第 7 章	全文电子期刊	257
7.1	电子期刊概述	257

7.2	电子期刊特点	257
7.3	电子期刊介绍	258
7.3.1	Elsevier Science 全文电子期刊	258
7.3.2	IEEE/IEE Electronic Library (IEL) 数据库	260
7.3.3	Springer Verlag 全文电子期刊	261
7.3.4	Academic Press 电子期刊	262
7.3.5	Kluwer 电子期刊	262
7.3.6	其他电子期刊	263
7.3.7	其他网站	266
7.3.8	引进数据库	266
7.4	小结	267
第 8 章	电子图书	268
8.1	电子图书概述	268
8.2	中国数字图书馆电子图书	269
8.3	超星中文电子图书	270
8.4	书生之家电子图书	271
8.5	方正 Apabi 电子新书	272
8.6	Ebrary 电子图书数据库	273
8.7	其他电子图书	274
8.8	电子图书存在的问题	274
8.9	小结	275
第 9 章	电子报纸	277
9.1	因特网上查找电子报纸	277
9.2	电子报纸的应用	281
9.2.1	人民日报	281
9.2.2	智能化的电子报纸——My Yahoo!	283
9.2.3	中国财经报刊数据库	284
9.3	报纸全文数据库	287
9.3.1	《参考消息》、《经济日报》、《中华读书报》全文数据库	287
9.3.2	中国重要报纸全文数据库 (CCND)	289
9.3.3	《北京日报》、《北京晚报》全文数据库	289
9.4	小结	290
第 10 章	网络信息资源综合利用	291
10.1	医学信息检索	291
10.1.1	医学信息概况	291
10.1.2	综合性 Web 检索工具	292
10.1.3	医学信息检索工具	295
10.1.4	其他医学专业信息检索工具	297
10.1.5	虚拟图书馆	299

10.1.6	相关医学网站	299
10.2	Internet 教育信息检索	300
10.2.1	通用检索工具	301
10.2.2	ERIC 数据库	301
10.2.3	专业教育信息检索工具	303
10.2.4	基础教育信息检索工具	304
10.2.5	其他网络学历教育资源网站	305
10.3	图书信息资源利用	305
10.3.1	出版机构及其产品信息	305
10.3.2	网上书店和书刊批销机构	306
10.3.3	读书俱乐部、有关图书的论坛和书评网站	307
10.3.4	出版行业组织	308
10.3.5	出版信息服务商及其信息产品	308
10.3.6	数字图书馆	308
10.3.7	OPAC	308
10.4	会议信息及会议文献信息的获取	309
10.4.1	会议信息数据库	309
10.4.2	会议信息网站	310
10.5	学位论文信息的获取	312
10.5.1	学位论文的种类与特点	312
10.5.2	开发利用学位论文信息的意义	312
10.5.3	网上获取学位论文资源	313
10.6	专利信息的获取	315
10.6.1	国内专利信息	315
10.6.2	国外专利信息	316
10.6.3	利用联机数据库获取专利信息	316
10.7	化工化学信息的获取	318
10.7.1	通用搜索引擎	318
10.7.2	Internet 的部分化学化工资源	318
10.8	经济信息的获取	321
10.8.1	Internet 网络资源指南	321
10.8.2	虚拟图书馆	321
10.8.3	经济类报刊	322
10.8.4	经济类论文资料的查询	322
10.8.5	利用 Internet 获取经济信息	322
10.9	如何在网上获取考研信息	323
10.9.1	考研信息包含的内容	323
10.9.2	考研信息的传统获取途径	324
10.9.3	网上考研信息	324

10.10 标准信息的获取	326
10.10.1 标准文献的种类与特点	326
10.10.2 国内标准信息的检索	327
10.10.3 国外标准信息的检索	327
10.11 小结	329
第 11 章 信息检索与信息服务	331
11.1 网络信息咨询	331
11.1.1 网络信息咨询的特点	331
11.1.2 网络信息咨询主体	331
11.1.3 网络信息资源	332
11.1.4 网络信息咨询客户	333
11.1.5 网络信息咨询的资金、设备与场所	333
11.2 定题服务	335
11.3 科技查新	336
11.3.1 科技查新的步骤	336
11.3.2 科技查新检索方式	337
11.3.3 科技查新报告实例	337
11.4 网络信息资源查询辅导服务	339
11.5 小结	340
参考文献	341

第 1 章 网络信息资源组织

近年来,随着 Internet 技术的迅速发展和普及应用,网络信息资源组织正越来越受到人们的关注。那么什么是网络信息资源组织?其原理如何?当前网络信息资源组织的现状及存在的问题有哪些?标准有哪些?以及发展趋势如何?在实际工作中的运用如何?下面做详细介绍。

1.1 信息组织概述

1.1.1 信息

1. 信息的定义

当今世界“信息爆炸”、“信息革命”、“信息时代”、“信息社会”已成为时髦的辞令,在社会的各个角落,信息无处不在,无时不有,信息存在于自然界,也存在于人类社会,信息来自于物质世界,也来自于精神领域。

那么,到底什么是信息呢?它的定义是什么呢?对于这个问题,古今中外的许多学者都对信息的定义进行了探讨。信息论和控制论的创始人之一维纳对信息下的定义是:信息是人们在适应外部世界并且使这种适应反作用于世界的过程中,同外部世界进行交换的内容的名称。目前,有关信息的定义已有上百种,其中比较具有代表性的定义有如下几种。

- (1) 信息是事物运动的状态与方式,是物质的一种属性。
- (2) 信息是负熵。
- (3) 信息是与控制论系统相联系的一种功能现象。
- (4) 信息是用以消除随机不确定性的东西。
- (5) 信息是一种可供开发和利用的并具有其自身价值的一种资源。

我们认为,信息的概念可以这样来定义:信息是客观世界中各种事物的变化和特征的最新反映以及经过传递后的再现。信息是通过一定的物质载体形式反映出来的,是事物存在的状态、运动形式、运动规律及其相互联系、相互作用的表象。信息来源于世界上的一切事物,它是事物各种运动变化状态的客观显示;信息产生于事物在运动发展变化中的各种差异以及规律;信息是客观事物相互作用、相互联系的反映。一般说来,信号、消息、知识、情报、数据、资料、程序和指令等都可以统称为信息。

另外,对于信息概念的认识,我们还可以从以下几个方面来理解。

- (1) 信息是使人们对事物不确定性的减少或消除的东西。
- (2) 信息是与外界相互交换的内容。
- (3) 信息是事物变化和差异的表现。

(4) 信息是系统的组织程度和有序程度的结合。

信息是我们认识世界,改造世界,取之不尽、用之不竭的宝贵资源。它与物质、能量被视为维系人类社会生存与发展的必不可少的战略性资源,已成为信息社会的关键变量,成为发展科技、经济、文化、教育的重要支柱之一。

2. 信息的分类

分类是人们认识事物的一种最为基本的方法,也是科学研究活动中常用的一种方法。由于信息在信息界和人类社会生产中存在和流动的范围极其广泛,因此对信息的分类也相对较复杂。不同学科领域的研究人员依据不同的分类标准,可对信息进行不同的划分。

(1) 按信息内容划分,可将信息分为人类信息和非人类信息。

(2) 按信息产生的先后和加工深度划分,可将信息分为一次信息、二次信息、三次信息。

(3) 按信息的存在形式划分,可将信息划分为内储信息和外化信息。

(4) 按信息的流通方式和传递范围划分,可将信息分为公开信息、半公开信息和非公开信息。

(5) 按信息载体形式划分,可将信息分为感官载体形式、语言载体形式、文字载体形式、电磁波载体形式、缩微载体形式、光波载体形式、声像载体形式和计算机载体形式。

另外,我们还可以按信息管理组织、按信息的反映面、按信息的时态、按信息的特征、按人对信息的感知方式、按信息运动形态等方式将信息划分成不同的类型。

3. 信息的特性

信息定义所揭示的是信息的本质属性,但信息本身还存在许多由本质属性派生出来的一般特性,包括普遍性、存储性、传递性、共享性、开发性、变换性、转化性和可伪性等,它们大都从某一个侧面体现了信息的基本特点。正是有了这些新颖的特性,才使信息成为既不同于物质又不同于能量的一类新的研究对象。同时,通过研究这些新颖的性质,又会有助于人们加深对信息基本概念的理解。在这里给大家介绍几种信息的常见的特性。

(1) 普遍性:是指信息的存在是普遍的,即信息是无处不在的,也是无时不在的。信息普遍存在于自然界、人类社会之中,也存在于人类的思维或精神领域之中。

(2) 存储性:随着人类社会的进步,信息的存储方式也在不断改进,从古代的结绳记事,直至如今的磁、光、半导体大规模集成电路的广泛应用,大大增加了物质载体存储的信息量,也大大丰富了信息的存储形式。

(3) 传递性:传递性是信息的本质特征之一,是指信息可以通过一定的传输工具和载体进行传递,从而形成信息联系,被人们感知和接受。信息是物质存在方式的直接或间接显示,这种显示靠光、声、磁以及语言、表情、文字符号等得以呈现和传递。信息的可传递性使信息有可能在短时间内广泛扩散、传播。

(4) 共享性:是指信息可被众多主体共享。信息共享是信息的运动规律之一,也是信息的一个重要性质,同时还是与其他资源的一个重要区别。信息在交换和转让过程中,其原有信息一般不会丧失,还可能会同时获得新的信息。正是由于信息可以被共享,所以它是人类社会的宝贵财富。

(5) 开发性:信息是一种可开发的资源,存储与传递信息的目的是为了开发信息资源。信息是点燃智慧的火种,它能给人们以新的知识,引导人们开拓更新的、更高层次的、更广阔的知识领域。人们正是通过对信息资源的开发利用,来促进科学技术的进步和社会的发展。

(6) 价值的不定性: 信息的价值就在于将一人、一事、一时、一地的信息传递给需要者, 从而创造新的物质财富和精神财富。由于人们对信息的需求和理解及判断能力的不同, 信息的价值有很大的差别。同样的信息对于不同的使用者可能有不同的价值, 同样的不同类型的信息, 对于不同的使用者也不一定有相同的价值。

1.1.2 信息组织

1. 信息组织的定义

《牛津英语词典》(Oxford English Dictionary) 中, 有关“组织”(Organization) 的定义是“组织就是使之成为一个有机的整体, 为其提供有序化的结构; 设计并使之形成工作顺序; 对其进行排列”。简而言之, 组织的含义是结构化。因而, 信息组织就是将收集的信息予以结构化, 也就是指采用一定的方式, 将某一方面大量的、分散的、杂乱的信息经过加工、整理、提炼、优化后形成一个便于有效利用系统的过程。

通常信息组织主要包括以下 6 个方面的活动。

- (1) 按照信息能够被获取的形式, 识别所有类型承载信息的实体之存在。
- (2) 识别被包含在那些承载信息的实体内或作为它们一部分的作品。
- (3) 把这些承载信息的实体系统化地组织到图书馆、档案馆、博物馆、因特网通信文档和其他类似存放处内的集合中。
- (4) 根据预先编制的标准的引文规则制作这些承载信息的实体之清单。
- (5) 为这些承载信息的实体提供名称、题名、主题和其他有益的检索点。
- (6) 提供定位每个承载信息的实体或其复本的方式。

信息组织是通过揭示信息之间内在的逻辑关系, 采用各种方法和手段使信息有序化、组织化、系统化的过程, 从而达到有效利用信息的目的。它是信息资源管理的基本范畴之一, 是信息资源建设的中心环节, 是建立信息系统的重要条件, 是信息检索与咨询的基础, 是开展用户服务的有力保证。

随着因特网的普及和发展, 网络信息已成为全球最大的信息资源库, 正以惊人的速度增长, 网络信息的组织和传播也以全新的方式进行。在信息网络化过程中, 面对种类繁多且杂乱无序的海量信息, 传统的文献组织理论和方法面临着困境和新的挑战。

2. 信息组织的理论基础

任何学科的发展都离不开理论的支撑, 传统的文献组织主要是以语言学、逻辑学和知识分类作为理论基础, 分类法、主题法就是在此基础上建立起来的。但网络信息具有高度无序性、分布动态性、信息数字化和载体多媒体化等特征。传统的文献组织的基本理论和方法已显得捉襟见肘, 迫切需要拓展信息组织的理论和方法, 以便更好地对网络信息进行有效组织管理。

(1) 关于知识组织的理论。知识组织是在图书馆学、情报学的分类系统和叙词表研究的基础上发展起来的, 是文献组织的延续和升华。与传统的文献组织以文献加工为本位, 揭示文献所载的知识内容不同, 知识组织是以知识单元为加工本位, 它不仅揭示文献的学科、主题内容, 而且更注重揭示文献所载的知识单元。知识组织将文献和信息中所包含的知识内容用语词和概念表达, 采用知识工程中的知识表达方法——OAV 三元组法进行知识标引和组织, 能更有效地对网络信息进行组织和检索。关于知识组织的理论基础还包括整序、概念学、数学方法、系统科学、知识工程、心理学等。

(2) 系统科学理论。系统科学的思想是 20 世纪 20 年代由奥地利学者贝塔朗菲提出来的,他把系统定义为“相互作用的诸多要素的复合体”,认为系统的定义可以确定为“处于一定的相互关系中并与环境发生关系的各组成部分的总体。”系统的基本特性有整体性、内部相关性、环境相关性、层次性、有序性和目的性等。系统的整体性是指各个要素按一定的方式构成的有机整体,其要素与整体、环境以及各要素之间相互联系、相互作用,使系统整体呈现出构成部分所不具有的功能。因此,可以通过提高系统要素的基质和系统整体的组织化程度,使整体的功能大于它的部分功能之总和。系统的层次性是指系统由若干子系统有机组成,子系统可由更小的子系统构成,从而形成一种层次结构。系统的有序性表现为结构的有序性和系统运动的有序性。在信息系统中,可将系统科学理论应用到信息组织、信息开发利用以及信息资源管理中,使信息系统更加组织化、有序化,更好地发挥信息资源的作用。

(3) 耗散结构理论。耗散结构理论是由比利时布鲁塞尔学派领导人普里高津教授于 1969 年在理论物理与生物学国际会议上首次提出来的,这一理论着重从“非平衡”和“开放系统”两个方面,论证了那些存在进化发展的系统有序性增加的原因。所谓耗散结构是指一个远离平衡态的开放系统,通过不断地与外界交换物质、能量,在外界条件的变化达到一定的阈值时,从原来的无序状态转变为在时间上、空间上或功能上的有序状态,这种在远离平衡情况下所形成的新的有序结构,就称为耗散结构。耗散结构理论的基本思想有两点:一是系统内部非平衡,是有序之源;二是开放系统通过与外界交换物质、能量而增加、维持有序性。信息系统正是一种耗散结构系统,它是远离平衡态的开放系统,具有输入、输出、多次循环及反馈等开放性的基本特征,因此耗散结构理论可作为信息系统信息有序组织的理论基础。

(4) 协同论。协同论是由前西德科学家哈肯于 1970 年提出来的,是一门专门研究系统进化普遍规律的科学,它研究由许多子系统构成的系统(如电子、原子、分子、细胞、器官、生物、社会等)是如何通过协作从无序到有序演化的规律,在对系统进化规律研究的具体或普遍程度上,协同学比耗散结构理论更进了一步。而信息系统正是由许多子系统构成的,如何建立各子系统之间的协同作用机制,使信息系统由无序向有序转化是信息组织研究中的一个重要课题,因此协同学也是信息组织的理论基础之一。

(5) 突变理论。突变理论是现代数学的一门新兴分支学科,是 20 世纪 70 年代由法国数学家勒内·托姆首先提出来的,它用形象而精确的数学模型来揭示和预测事物的连续性中断的质变过程。突变理论指出系统的熵可以增加也可以减少,这种熵增、熵减可以在宏观无限小的时间内突然发生。突变熵减少产生的有序性,可以抵消、战胜自然界某些自发熵增趋势,有可能造成有序性的发展,由此得出突变理论的一个重要观点:“突变是产生有序性的重要源泉”。因此突变理论为信息组织理论的发展与完善提供了理论基础。

(6) 信息自组织理论。信息自组织理论是信息组织方法的拓展,是信息组织理论研究中的一项新课题。凡是能够不再借助于外部控制而能实现从无序到有序的转变,并维持稳定有序状态的系统,就称为自组织系统。任何自组织系统都是通过谐振、反馈和放大来完成信息增强,并保持其有序效应的。信息自组织是指作为信息系统组成要素的信息,由于人与人之间、人与系统其他要素之间存在的相关性、协同性或默契性而形成特定结构与功能的过程,也就是信息系统无须外界指令而能自行组织信息,自我走向有序化和优化的过程。近几十年来,由于信息总量的持续增长、信息技术的飞速发展,使信息系统显著地具备了自组织的条件,特别是网络信息已经具有自组织系统的开放性、远离平衡和非线性相干等特征,因此研

究信息自组织理论对于网络信息的有序组织具有非常重要的理论与实践意义。

3. 信息组织的方法基础

(1) 语言学。要把庞杂分散的信息组织成有序优化的整体,就必须建立符号系统。有了这种符号系统,信息系统的有序特征才能体现,信息单元的个体特征才能被揭示出来,各种信息单元才能对号入座,被纳入到这种符号系统的框架之中,形成一个便于检索的序化信息集合。尽管各种信息组织符号系统的形式不同,但都和自然语言一样,有着共同特征:有语词、有词汇、有语法。这就是信息组织的语言学基础。

(2) 逻辑学。形式逻辑是关于思维规律的科学。信息组织其实就是人脑思维活动的一种外化,它既然是一种思维活动,就必须遵循科学的思维方法。也就是说,进行信息组织工作必然用到形式逻辑的一些方法,信息组织的行为只有符合逻辑思维规律,才能保证信息组织的优化,序化质量。

(3) 知识分类。分类是人们认识事物的一种最为基本的方法。知识分类是一门研究知识体系结构的学问。信息的主体是知识,信息组织活动必须建立在人们对知识体系认识的基础之上。

1.1.3 网络环境下信息组织的内涵

随着社会信息化进程的加速,网络信息资源呈几何级数增长,信息组织作为网络信息资源开发和利用的一个重要环节,其内涵得到了延伸和拓展。

(1) 任意层次的信息内容元素、信息单元和信息集合体系正逐步以计算机可识别和理解的方式被定义、描述、指向、链接、传递和动态组织。

(2) 信息资源、信息组织工具、信息系统日益聚合在同一数字介质层面,它们间的动态链接、转换、互操作和集成等正逐步实现。

(3) 信息资源系统、信息服务系统和用户及其信息系统,日益连接在同一网络空间,它们之间的链接、交换、互操作和协作也日益成为可能。

(4) 各种基于网络、基于知识、基于协作的信息组织工具和系统机制也日益成熟,并可按照开放的、计算机可识别处理的方式进行描述、交换和互操作。

信息组织是由来已久的一种人类社会实践活动,在其发展过程中,通过不断从相关学科的理论和方法中汲取营养,使自身逐渐得到充实和完善。系统论、耗散结构理论、协同论和突变理论等理论使信息组织有了理论基础。而计算机技术、网络技术、多媒体技术等信息技术的发展使信息组织有了技术支持。因此,只有将大量的、分散的、杂乱的信息组织成一个科学的信息系统,建立起内在的关联性,才能使信息系统的总体功能大于各个信息单元功能的总和,这样才能充分发挥信息资源的价值与作用。

1.2 知识组织概述

1.2.1 知识的含义

1. 知识的定义

自有人类文明以来,知识作为人类认识自然和社会的阶段性的成果,就一直受到了古今中

外学者的重视。如英国著名学者培根有句流传得非常广泛的名言：“知识就是力量”。

希腊时代的诡辩学认为，所有知识都是来自于个人的经验，而每个人对事物的感受不同，因此知识仅仅是相对于个人而言的，没有绝对意义的知识。

著名哲学家罗素在《人类的知识》一书中经分析得出结论，知识是一个意义模糊的概念。

诺贝尔物理奖获得者彭亚斯认为，知识不是一堆杂乱无章的信息，而是有关世界怎样进行的信念。

彼得·德鲁克认为，知识是一种能够改变某些人或某些事物的信息，这既包括使信息成为行动基础的方式，也包括通过对信息的运用使某个个体有能力进行改变或者进行更有效行为的方式。

美国学者贝尔认为，知识是对事实或者思想的一套有系统的阐述，提出合理的判断或经验性的结果。它通过某种交流手段，以某种系统的方式传播给其他人。

在以上观点中，“知识”不再是一个简单的、多元素的无序集合，而是被纳入一个动态的、与人或组织交互的系统。

1983年3月，国家科技领导小组在“关于知识与国家知识基础设施的研究报告”中，对“知识经济”中的“知识”定义为：经过人的思维整理过的信息、数据、形象、意向、价值标准以及社会的其他符号化产物，不仅包括科学技术知识，还包括人文社会科学的知识、商业活动、日常生活以及工作中的经验和知识、人们获取并运用和创造知识的信息，以及面临问题作出判断和提出解决方法的知识。这一定义基本上概括了国内专家和学者对知识概念的理解和认识。

另外，对知识概念通常有两种理解，即广义知识和狭义知识。广义知识是指人们通过学习、积累、发现、发明各种知识的总和，包括普通知识和专业知识。狭义知识是指知识经济研究的知识，通常是专业知识。

2. 知识的分类

古往今来，人类积累了无数的知识，如果不对其进行合理的分类，人们就难以找到知识创新的规律和科研工作的主攻方向，也难以提出和发展新的边缘学科或者交叉学科。我们可按照不同的标准对知识进行分类。

- (1) 按领域划分，可将知识分为哲学知识、社会科学知识和自然科学知识；
- (2) 按用途划分，可将知识分为科学知识、技术知识和文化知识；
- (3) 按状态划分，可将知识分为存量知识和流量知识；
- (4) 按水平划分，可将知识分为低级知识、中级知识和高级知识；
- (5) 按性质划分，可将知识分为普通知识和专业知识；
- (6) 按载体划分，可将知识分为隐性知识和显性知识。

另外，1996年，世界经济合作与发展组织在题为《以知识为基础的经济》报告中，将对经济有重要作用的知识分为四种类型，分别是知事类知识（Know-what）、知因类知识（Know-why）、知能类知识（Know-how）、知人类知识（Know-who）。也有学者扩充了知识的类型，增加了知时类知识（Know-when）、知地类知识（Know-where）和知数类知识（Know-quantity）。这就是所谓的具有中国特色的知识分类，可以用公式记为 $K=6W+Q$ 。

3. 知识的特性

知识有许多特性，主要包括智慧性、客观性、依附性、不磨损性、非遗传性、增值性和