

面向21世纪信息管理与信息系统专业
东南大学出版社 核心课程系列教材



信息组织与存储

史田华 李晓鹏 颜端武 曹海峰/编著

信息组织与存储

史田华 李晓鹏 颜端武 曹海峰 编著

东南大学出版社

·南京·

内容简介

本书系统阐述了信息组织与存储的基本概念、原理和方法等内容,包括信息组织的含义及基本内容;分类法与分类语言;主题法与主题语言;标引与著录;信息组织自动化技术;数字化信息的组织与发现;信息组织的数据库方法;数字图书馆信息组织;信息存储概述;计算机信息存储介质和网络环境下信息存储技术等。

本书可用作高等院校信息管理与信息系统和电子商务各相关专业本科生的教材,也可作成人教育、函授教育、高职高专相关专业的教材;并可供各种与信息资源组织有关的机构或部门(包括各类信息机构、网络公司以及企业信息资源管理部门)的工作人员学习参考。

图书在版编目(CIP)数据

信息组织与存储/史田华等编著. —南京:东南大学出版社,2003.7

ISBN 7-81089-073-5

I. 信... II. 史... III. ①信息管理—高等学校—教材②信息存储—高等学校—教材 IV. ①G203②
TP333

中国版本图书馆 CIP 数据核字(2003)第 056357 号

东南大学出版社出版发行

(南京四牌楼 2 号 邮编 210096)

出版人:宋增民

江苏省新华书店经销 南京京新印刷厂印刷

开本:B5 印张:20 字数:437 千字

2003 年 8 月第 1 版 2003 年 8 月第 1 次印刷

印数:1—4000 定价:28.00 元

(凡因印装质量问题,可直接向发行科调换。电话:025—3795801)

面向 21 世纪信息管理与信息系统专业

核心课程教材建设委员会

孙建军（南京大学信息管理系教授、博导）

吴清烈（东南大学经济管理学院副教授、博士）

武 忠（东南大学经济管理学院副教授、博士）

史田华（南京理工大学信息管理系教授）

王曰芬（南京理工大学信息管理系副教授、博士）

郑会颂（南京邮电学院管理工程系教授）

何有世（江苏大学工商管理学院教授、博士）

刘秋生（江苏大学工商管理学院副教授、博士）

周建屏（苏州大学图书馆副研究馆员）

成 颖（南京大学信息管理系）

前 言

信息是社会发展的的重要前提和主要资源,是人类的宝贵财富。信息资源是现代社会最重要的战略资源之一,它的开发利用水平是国家综合国力的体现。信息组织与存储就是根据人类社会发展的需求,以各类信息源为对象,通过对其内容特征等的分析、选择、标引、处理,使其成为有序化集合的信息增值活动,是信息资源开发利用的重要环节。

《信息组织与存储》课程教材是编著者在长期从事信息管理教学、科研和实际工作的基础上,按教育必须三个面向和与时俱进的要求撰写而成的。在内容取材上,它立足现实,放眼未来,注重可持续发展和理论联系实际,广泛汲取了国内外的最新研究成果。全书共分 11 章和附录,第一章、第六章、第七章和附录由李晓鹏撰写;第二章、第三章和第四章由史田华撰写;第五章、第八章和第九章由颜端武撰写;第十章和第十一章由曹海峰撰写。每章末及附录末均有习题。全书由史田华统稿,颜端武统排。

本书的出版得到了东南大学出版社的大力支持。我们在编写过程中,还参考了国内外许多专家与学者的研究成果,并借鉴了一些网上资料,在此一并表示最诚挚的感谢。

由于本书涉及知识面宽,内容广泛,尽管编著者作了极大努力,不足之处仍在所难免,恳请专家、读者批评指正。

编著者

2003 年 5 月

目 录

1	信息组织引论	(1)
1.1	信息组织的含义与历史发展	(1)
1.1.1	信息组织的含义	(1)
1.1.2	信息组织在信息资源建设中的地位与作用	(2)
1.1.3	信息组织的历史发展	(2)
1.2	信息组织的理论与方法基础	(4)
1.2.1	信息组织的理论基础	(4)
1.2.2	信息组织的方法基础	(5)
1.3	信息组织的基本方法	(5)
1.3.1	信息序化的基本方法	(5)
1.3.2	信息优化的基本方法	(7)
1.4	信息组织的一般过程和原则	(9)
1.4.1	信息组织的一般过程	(9)
1.4.2	信息组织的基本原则	(9)
	习题 1	(11)
2	分类法与分类语言	(13)
2.1	分类法	(13)
2.1.1	分类法的类型	(13)
2.1.2	类目	(14)
2.1.3	号码(标记符号)	(16)
2.1.4	正表	(24)
2.1.5	附表	(25)
2.1.6	说明与索引	(26)
2.2	国际十进分类法——UDC	(28)
2.2.1	UDC 体系结构	(29)
2.2.2	UDC 号码组配	(33)
2.2.3	UDC 辅助表和辅助类号	(36)
	习题 2	(43)
3	主题法与主题语言	(44)
3.1	标题法	(44)
3.1.1	标题法含义	(44)
3.1.2	标题词实例	(45)
3.1.3	标题词词组语法	(46)
3.1.4	标题词语言的缺点	(47)

3.2	叙词法	(47)
3.2.1	叙词性质	(47)
3.2.2	叙词表	(49)
3.2.3	范畴索引	(51)
3.2.4	词族索引	(55)
3.2.5	英汉对照索引	(57)
3.2.6	附表	(59)
3.3	关键词	(60)
3.3.1	关键词含义	(60)
3.3.2	题内关键词	(62)
3.3.3	题外关键词	(65)
3.3.4	单元词	(67)
3.4	参照系统	(67)
3.4.1	See 和 See Also	(67)
3.4.2	叙词与叙词、非叙词间概念逻辑关系	(70)
3.4.3	注释	(72)
3.5	分类主题一体化词表	(74)
3.5.1	分类主题一体化词表的发展及原理	(74)
3.5.2	分面叙词表	(78)
3.5.3	分类主题词表	(83)
	习题 3	(88)
4	标引与著录	(89)
4.1	分类标引	(89)
4.1.1	分类标引的任务与作用	(89)
4.1.2	分类标引的几个问题	(90)
4.1.3	分类标引的依据	(92)
4.1.4	组配分类标引	(94)
4.1.5	专业信息部门分类标引	(96)
4.2	主题标引	(100)
4.2.1	主题标引规则与类型	(100)
4.2.2	主题标引实例分析	(104)
4.2.3	文摘与全文标引	(106)
4.2.4	分组标引	(108)
4.2.5	上位词标引	(110)
4.2.6	适度标引	(111)
4.2.7	靠词标引	(112)
4.2.8	增词标引	(112)
4.2.9	组配标引技术	(114)

4.3	著录	(122)
4.3.1	著录的定义	(122)
4.3.2	文献信息著录总则	(122)
4.4	MARC	(126)
4.4.1	MARC 的产生和发展	(126)
4.4.2	CNMARC 的格式	(128)
	习题 4	(132)
5	信息组织自动化技术	(133)
5.1	信息组织的自动化概述	(133)
5.2	自动标引技术	(135)
5.2.1	自动标引概述	(135)
5.2.2	自动抽词标引和自动赋词标引	(136)
5.2.3	中文自动标引	(138)
5.3	自动分类技术	(143)
5.3.1	自动分类的基本原理	(144)
5.3.2	自动分类的典型算法和系统原型	(146)
5.3.3	国内外自动分类的研究和发展概况	(148)
5.4	自动文摘技术	(151)
5.4.1	自动摘录	(151)
5.4.2	基于理解的自动文摘	(153)
5.4.3	基于信息抽取的自动文摘	(155)
5.4.4	基于结构的自动文摘	(156)
5.4.5	中文自动文摘的研究状况	(157)
	习题 5	(158)
6	数字信息的组织与发现	(159)
6.1	数字信息资源概述	(159)
6.1.1	数字信息资源的概念	(159)
6.1.2	数字信息资源的特征	(160)
6.1.3	数字信息资源的组织	(162)
6.2	Metadata	(163)
6.2.1	认识 Metadata	(164)
6.2.2	Dublin Core 概述	(166)
6.2.3	RDF——元数据的“容器”	(169)
6.3	搜索引擎(Search Engine)	(172)
6.3.1	认识 Search Engine	(172)
6.3.2	元搜索引擎	(176)
6.3.3	智能搜索引擎	(180)
6.4	Z39.50 协议	(184)

6.4.1	Z39.50 协议简介	(184)
6.4.2	Z39.50 协议的原理	(187)
习题 6		(189)
7	信息组织的数据库方法	(190)
7.1	信息组织模型与数据库发展	(190)
7.1.1	层次网状模型	(191)
7.1.2	关系数据库	(192)
7.1.3	面向对象数据库	(193)
7.2	数据库系统的信息组织方法	(194)
7.2.1	单用户数据库系统的信息组织方法	(194)
7.2.2	主从式数据库系统的信息组织方法	(196)
7.2.3	分布式数据库系统的信息组织方法	(197)
7.2.4	C/S 数据库系统的信息组织方法	(198)
7.3	Web 数据信息与半结构化数据库系统	(199)
7.3.1	数据集成子系统	(200)
7.3.2	数据发布子系统	(200)
7.4	多媒体数据库的信息组织方法	(201)
7.4.1	多媒体数据库关键技术	(201)
7.4.2	多媒体数据库的主要实现方法	(202)
7.4.3	基于内容检索与多媒体信息查询	(205)
习题 7		(206)
8	数字图书馆的信息组织	(207)
8.1	数字图书馆简介	(207)
8.1.1	数字图书馆概念与特征	(207)
8.1.2	国外数字图书馆发展概况	(209)
8.1.3	我国数字图书馆工程的进展	(211)
8.2	数字图书馆的体系结构	(215)
8.2.1	数字图书馆系统结构	(215)
8.2.2	数字对象的组织结构	(216)
8.2.3	信息节点的组织结构	(217)
8.3	数字图书馆信息组织方法	(221)
8.3.1	数字图书馆信息组织原则	(221)
8.3.2	数字图书馆信息资源库建设	(222)
8.3.3	数字图书馆信息共享机制	(224)
习题 8		(225)
9	信息存储概述	(226)
9.1	信息存储技术及其发展	(226)
9.2	信息存储结构	(229)

9.2.1	信息的逻辑结构	(229)
9.2.2	信息的存储结构	(229)
9.2.3	计算机存储系统	(230)
9.3	存储器的类型和主要技术指标	(233)
9.3.1	存储器的类型	(233)
9.3.2	存储器的主要技术指标	(235)
习题 9		(237)
10	计算机信息存储器	(238)
10.1	磁存储器	(238)
10.1.1	概述	(238)
10.1.2	磁存储器的特点	(239)
10.1.3	磁存储原理	(239)
10.1.4	磁存储器的基本结构	(240)
10.1.5	磁存储器类型	(242)
10.2	光存储器	(247)
10.2.1	概述	(247)
10.2.2	光存储器的特点	(248)
10.2.3	光盘存储系统基本结构	(248)
10.2.4	光盘的组成结构	(249)
10.2.5	光存储器类型	(252)
10.3	半导体存储器	(254)
10.3.1	概述	(254)
10.3.2	半导体存储器的基本结构	(256)
10.3.3	随机存取存储器	(256)
10.3.4	只读存储器	(258)
10.3.5	快闪存储器(Flash Memory)	(259)
10.4	铁电存储器	(261)
10.4.1	概述	(261)
10.4.2	铁电存储器的存储原理	(262)
10.4.3	铁电存储器的结构	(262)
10.4.4	非易失性铁电存储器	(263)
10.4.5	动态随机读取存储器(FDRAM)	(265)
10.4.6	铁电存储器的应用	(265)
习题 10		(266)
11	网络环境下的信息存储技术	(267)
11.1	概述	(267)
11.2	直接连接存储 DAS	(268)
11.2.1	DAS 的概念与结构描述	(268)

11.2.2	DAS 所采用的技术	(269)
11.2.3	DAS 的优缺点	(272)
11.3	附网存储 NAS	(273)
11.3.1	NAS 概念及结构描述	(273)
11.3.2	NAS 的特性	(274)
11.3.3	NAS 所采用的协议	(275)
11.4	存储区域网络 SAN	(277)
11.4.1	概念及其结构描述	(277)
11.4.2	SAN 的特性	(278)
11.4.3	SAN 所采用的技术	(278)
11.5	NAS 和 SAN 的比较与融合	(282)
11.5.1	NAS 与 SAN 的比较	(282)
11.5.2	NAS 与 SAN 融合	(283)
11.5.3	几种 NAS 与 SAN 融合方案	(284)
11.6	网络存储的管理	(285)
11.6.1	网络存储管理的目的与方式	(285)
11.6.2	网络存储管理的层次	(287)
11.6.3	网络存储的管理阶段	(289)
11.6.4	网络存储的分级管理	(289)
习题 11		(292)
附录:标记语言简介		(293)
1	认识标记语言	(293)
2	标准通用标记语言简介	(294)
2.1	什么是标准通用标记语言	(294)
2.2	SGML 程序的组成	(294)
2.3	SGML 标准体系简介	(295)
2.4	SGML 的不足之处	(296)
3	超文本标记语言简介	(297)
3.1	什么是超文本标记语言	(297)
3.2	超文本标记语言的缺陷	(297)
4	可扩展标记语言简介	(299)
4.1	可扩展标记语言的产生	(299)
4.2	认识可扩展标记语言	(300)
4.3	可扩展标记语言的特点	(301)
4.4	XML 与 HTML 比较	(303)
5	标记语言家族一览	(304)
习题(附录部分)		(306)
参考文献		(307)

1 信息组织引论

- 信息组织的含义与历史发展
- 信息组织的理论与方法基础
- 信息组织的基本方法
- 信息组织的一般过程和原则

1.1 信息组织的含义与历史发展

信息组织属于信息资源管理的基本范畴,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。因此,探讨信息组织的理论和方法是构架和完善信息资源管理体系的迫切要求。

1.1.1 信息组织的含义

信息组织是随着社会信息化而出现的一个趋于规范的术语。人类的生存与发展离不开信息,早期的信息获取主要靠人与人之间的直接交往。有了文字记载以后,主要靠文献获取信息。随着文献的大量出现和急剧增长,便有了文献的整理加工——国内习惯称作情报组织,它为人们索取特定文献信息提供了方便。

信息组织,即信息序化或信息整序,也就是利用一定的科学规则和方法,通过对信息外在特征和内容特征的表征和序化,实现无序信息流向有序信息流的转换,从而保证用户对信息的有效获取和利用及信息的有效流通和组合。信息组织是信息管理的重要环节和基本工作,是信息资源开发利用的主要手段,是信息传播的前期准备。通过信息组织可以达到甄别信息、精化信息、重组信息的目的。信息组织是信息活动的必然要求,其起源在于信息本身的自然无序状态。序是事物的一种结构形式,是指事物或系统的各个结构要素之间的相互关系以及这种关系在时间和空间中的表现,即事物发展中的时间序列及排列组合、聚类状态,结构层次等空间序列。当事物结构要素具有某种约束性,且在时间序列和空间序列呈现某种规律性时,这一事物就处于有序状态。反之,则处于无序状态。

信息的无序状态可以从两个角度来认识:

(1) 从信息个体单元来看,组成信息的各个语言要素,如字、词、句、段落、章节等处于一种杂乱无章的状态;或组成信息的各个内容要素,如命题、观点、认识、推理等处于一种错误矛盾的状态;或组成信息的各个载体要素处于一种零散错位的状态,这样,就无法准确、科学地表述信息内涵,形成信息实体,也就无法被信息用户理解和应用。

(2) 从信息集合即信息流来看,不同信息个体处于彼此毫无关联的自然状态,缺乏科学稳定的框架体系,没有信息的深层次加工,无法融合成一个可以科学地把握其信息内容、有效地查检其信息内容、完善地维护其信息形态、良好地排列其信息形态、充分地实现其信息价值的有序的信息集合体。

显然,对于我们来说,信息组织所要解决的信息无序问题主要是第二种情形,

第一种情形主要是通过信息生产和信息搜集工作来予以解决的。也就是说,信息组织是在信息搜集基础之上进行的信息系统的信息整理和序化工作。一般地讲,信息处于有序状态或是无序状态是就一定参考系而言的。第一种情形的信息无序在信息生产过程中得以解决了,信息生产过程中没有完成的信息序化实际上提供给我们的是一种信息“次品”或“废品”。我们可以通过搜集工作中的选择性评价作为一种社会栅栏或滤化装置将这些信息“次品”或“废品”排除在信息系统之外。

1.1.2 信息组织在信息资源建设中的地位与作用

信息资源建设的基本内容与环节包括信息的采集、组织、开发和利用。信息组织在其中属于关键因素,处于关节点的地位。它既是一个相对独立的环节,又广泛渗透到其他环节,即信息采集中有信息的优选,信息的开发与利用中有对信息的分析、比较、信息重组等。现代社会信息的基本特征是庞杂性和分散性,人们对信息需求的特点是专业的高度选择性、内容的精良性、时间的紧迫性。两者之间的联系及其矛盾的调解就靠信息组织。

信息组织是一个信息增值过程。在这个过程中,杂乱无章的原始信息变成一个有序精良的信息系统。一个相对“粗放”型的信息贫集转化为一个“集约”型的信息富集,并为信息的进一步增值(如信息的分析研究)打下基础。

信息像其他产品一样,也有自己的生命周期。信息组织是信息生命周期管理思想的最积极、最活跃的因素。做好信息组织工作,会促进信息流通,发挥信息效用,使信息流形成良性循环,使信息在时间上更快速地传播;在空间上更广泛地传播,满足人们尽快、尽早地获得“地球村”各处的信息。

信息组织是信息检索利用的基础。没有它,就不会有信息检索系统,也不会有信息检索,从而也就谈不上信息的有效利用。信息组织为人们获取信息提供了方法和工具。

1.1.3 信息组织的历史发展

信息组织的发展主要是与信息管理整体发展和信息组织理论方法及信息技术发展紧密联系的,因此,在发展阶段上,大致与信息管理发展阶段保持基本的一致性。由于在原始信息管理时期几乎谈不上任何形式的信息组织活动,信息的传递通常是不正规的、多为自发性的。因此,我们仅把信息组织的发展区分为三个阶段:清册职能时期、查检职能时期和组织职能时期,主要与古代信息管理时期、近代信息管理时期和现代信息管理时期相对应。必须明确指出的是,信息组织发展阶段的演化并不是一个阶段对一个阶段的全面否定,而是一种扬弃,一种理论和方法的完善和扩展,即使是最早的清册职能也并未在现代信息管理时期被彻底废止,而是融入了一种更为综合的组织职能之中。

1) 清册职能时期

在古代信息管理时期,信息管理的着眼点不在于“用”而在于“管”,信息管理活动并不要求完备、科学的信息组织给予支持。这一时期的信息组织活动主要体现

为一种清册职能(inventory function),它主要通过信息的记录和登载,如同“信息账房先生”,告诉人们有什么信息而已。当然,这时的信息主要是指图书、档案等文献信息。偶尔人们也使用“catch word”等方式进行最粗浅的主题揭示,分类法的使用本身也多在为信息编排之用,中国古代目录学家们的提要或提要式文字也并未改变总体的清册职能。显而易见,这种财产登记式信息组织活动非常简单而不规范,手工方式和信息交流匮乏也决定了其发展的不成熟性。图书馆目录虽然远远不能与今天的图书馆目录相比,却已成为这一时期的代表成果,围绕图书馆目录展开的初步研究中已初步显露了现代信息组织思想和方法的萌芽,在某种意义上“辨章学术,考镜源流”是现代信息组织中主题揭示和内容分析的滥觞。

2) 查检职能时期

一切信息机构存在的目的都不可能是单纯的收藏和保管,利用才是终极目的。要促进信息的利用就必须有完备、科学的信息组织活动作基础。图书馆等文献信息机构的发展,特别是管理思想、技术和方法的发展也为信息组织的发展创造了有利条件。因此,信息组织活动不仅告诉人们有什么信息,而且告诉人们怎样找到信息,构成这一时期的信息组织活动的主要职能就是查检职能(finding list function)，“信息账房先生”摇身一变成了“信息药房先生”。但仍然可以发现,这里的信息还没有冲破文献信息的巢窠,现代意义上的信息还缺乏有效的、有意识的控制活动。但是,信息组织的理论、技术和方法已经不可同日而语了。信息组织已经不再是一种记簿式处理,对信息的外在特征和内容特征的描述和揭示都得到不同程度的重视,尽管对信息的概念单元分析仍然存在不足。信息组织不仅能够为用户提供更多的检索点(access points),同时,索引和文摘的出现使二次信息体系中书目形式一统天下的局面被彻底打破。信息组织的纯手工操作方式也不复存在,机械化、半机械化手段开始导入信息组织中,出现了穿孔卡片系统和缩微胶片系统等新的信息组织技术手段。不仅信息描述内容丰富化、信息揭示方法多样化,而且信息组织从载体单元到内涵单元的重点倾斜具有里程碑意义。无序信息向有序信息的转换使信息的利用有更多的方法和途径,在近代信息管理时期,信息组织确立了“用户中心论”,并在检索点的形成和排列、整序成果的推广和应用,传统技术的更新和演化等方面得以充分体现。在这一时期,信息组织的发展可以用主题揭示语言对分类揭示语言的冲击、文摘和索引型序化成果对书目型序化成果的冲击、机械化和半机械技术手段对手工技术手段的冲击,给予充分地概括,这些变化特征为信息组织由查检职能为主导向完整意义的组织职能发展奠定了坚实的基础。

3) 组织职能时期

以电子计算机为代表的现代信息技术的导入,把信息管理推上一个新的发展水平。也把信息组织引入一个新境界,使信息组织过去连想也未曾想到的新概念和新设想变成了现实。信息组织在最完整的意义上实现了整序信息、科学分流、促进选择、保证利用的组织职能(organizing function)。这种组织职能已经绝对不满足于告诉人们有什么信息和怎样找到信息,它不仅可以告诉人们我们这里有什么信息,而且还可告诉人们其他地方有什么信息;人们不仅可以找到最需要的信息,

而且还可以知道什么信息适合于自己;人们不仅可以利用一种途径获得信息,而且可以多途径、多角度地查检信息。在这里,信息的概念被高度扩充,它不仅是文献信息,而且是非文献信息;不仅是文字信息,而且是声音、图形或动感信息;不仅是知识信息,而且是满足非学习研究型需求的其他信息;不仅是提供线索的二次信息,而且是直接满足要求的一次信息和三次信息;不仅是被动接收的信息,而且是人们参与生成和选择的信息……所有这一切,反映了现代信息组织的超前性和现实性、精细化和普及化、协作性和专业性、标准化和多样化、体系化和特色化、市场性和调控性、自动化和智能化的有机结合。信息组织融信息描述、信息揭示、信息分析、信息贮存于一体,现代信息管理时期中信息组织活动的联合化、信息组织产品的商品化、信息组织方式的专营化、信息组织成果的媒体化、信息组织基础的需求化成为这一时期最明显的标志。在未来发展的道路上,信息组织的功能不会弱化,而且会进一步强化。信息组织、信息搜集和信息传播甚至于信息生产的连接会更加紧密,一体化趋向会更加明显,但其中的信息组织的功能将依然存在并得到充分体现。

1.2 信息组织的理论与方法基础

1.2.1 信息组织的理论基础

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,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、耗散结构理论和协同论使信息组织有了理论基础。

系统论认为,系统内部各要素是相互关联地有机地结合在一起的,它们之间的关系不是一种简单的结合,正是由于系统要素间这种有机的相互关联关系,系统的整体功能才产生了质的飞跃,远远超出各单个要素的功能之总和。这就是系统论的“整体大于部分之和”的原理。在信息组织中,如果将大量的、分散的、杂乱的信息组织成一个系统,建立起内在的关联性,那么,信息系统的整体功能将大于各个信息单元的功能之总和。也就是说,这将能充分发挥信息资源的价值与作用。基于这一原理,信息组织的目标是要建立信息系统。

20世纪60年代,比利时学者I.普里高津提出了一个著名论断:在开放系统中,系统不断与外界进行物质和能量的交换,熵趋于最小值,能量远离平衡,混乱度最小,从原来无序结构转为一种时间、空间和功能上的有序结构。这就是适于一切开放系统的耗散结构理论。普里高津找到了非平衡、开放系统是有序之源。这一原理给信息组织的启迪是:所要建立的信息系统应该是一个开放系统。现实存在的包括信息组织和信息检索在内的信息系统正是这样一个开放系统,它与外界进行着信息的交换,既采集信息,也输出信息。

一个信息系统是由多个子系统构成的,建立各子系统之间的协同作用机制是非常重要的,这正是来自协同论的指导思想:无论什么系统从无序向有序的演化,

都是大量子系统之间相互作用而又协同一致的结果。

1.2.2 信息组织的方法基础

语言学、逻辑学、知识分类是信息组织的方法基础。

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,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。

形式逻辑是关于思维规律的科学,信息组织是一种思维活动,它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。

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。

1.3 信息组织的基本方法

在谈及信息组织的基本方法前,我们必须先介绍两个基本概念——信息的外在特征和内容特征,它们是信息组织的基本对象和管理依据。所谓信息的外在特征就是指信息的物质载体所直接反映的特征,构成信息载体的外在的、形式的特征,如信息的物理形态、题名和责任者及信息的类型、信息生产和流通状况等方面的特征。所谓信息的内容特征就是信息所包含和承载的具体内容,即通过信息载体传递和交流的具体内容,两者共同统一于信息本身。

信息的序化和优化是信息组织的结果和目标。信息的序化就是按照一定的方法将无序的信息组织成有序信息的过程。经序化过的信息会变得一目了然,便于管理和利用。信息的优化则是在序化的基础上,按照结构功能优化的目的对信息进行再序化的过程。通过序化和优化可以使原始信息转化为一种有序精良的信息系统或集约化信息系统,这为人们快速、准确、方便地获得所需要的信息提供了有利条件。因此,信息组织的基本方法就可以概括为,以信息外在特征和内容特征为基础的信息序化方法和信息优化方法。

1.3.1 信息序化的基本方法

信息序化靠的是符号系统。从符号学的理论来认识各种符号系统所能表达的信息有语法信息、语义信息和语用信息。它们是信息序化的三种基本类型。

1) 语法信息序化法

语言学中的语法学(grammar)是研究语言符号之间的结构规则的,主要包括词的构成和变化、词组和句子的组织,研究的语言内容属于形式的范畴。信息组织

借用了“语法”二字,表征的是从形式特征方面进行组织的信息。即根据信息的外在特征,使用一套形式化的符号系统来组织信息的方法,序化符号不揭示所组织信息的内容特征。这类序化法有号码法、物名法、专门代码法、引证关系法、时序法、地序法等。

(1) 号码法

如用邮政编码组织信件,用身份证号、学号组织人群信息,用地址码(门牌号)组织住户信息等等。文献编号,如专利号、报告号、标准号、档号,在组织各类文献信息中有着重要作用。号码法一般具有明确、简便、惟一、标准的特点。

(2) 物名法

是利用事物名称的字顺特征来序化信息。用这一方法组织的信息系统多种多样,如人名录、机构名录、地名词典、产品目录、货物清单等。在文献工作方面有题名系统、著者名系统、机构名系统等。任何事物总有其名,所以这是一种常用的信息序化组织法。

(3) 专门代码法

是指在某一专业领域采用的有一定专业含义又形式化了的编码系统,如化合物的分子式、环状化合物的环状代码等。利用这类代码组织信息,形成专用的信息系统,如分子式索引、环系索引。这类序化法在特定的专业领域是专业信息组织的重要方法。

(4) 引证关系法

是利用信息之间的引证与被引证关系来组织信息。例如,利用法院判案之间的引证关系所组织的《谢泼德引文》系统;利用文献之间的引证关系所组织的各种引文索引系统。引证关系序化法不仅是组织信息的一种独特方法,而且还是进行各种评价研究的基础。

(5) 时序法

以信息的形成时间为序化符号,按时间顺序组织信息。大事记、传记资料、生平事迹多以时间为线索组织信息。这种方法能揭示信息内容的发展变化过程。

(6) 地序法

以信息的形成地区或信息内容所反映的地区为序化符号,按行政区划排列法来组织信息。这种方法能反映有隶属关系和横向联系地区的信息。

2) 语义信息序化法

任何信息除了外在特征之外,更重要的是其内在特征。语义信息中的“语义”取之于语义学(semantics),研究语言符号与它代表的对象之间结构关系之义,是为研究信息的内容特性而对信息进行表征的。即根据信息的内容特征,使用一套含有语义的符号系统来组织信息。信息的内在特征表现为信息的知识范畴、主题内容、概念属性。所以,语义信息序化法有分类法和主题法两种基本类型(概念属性归于主题法)。

人们的分类对象可概括为三种:一是实物,如商品;二是概念,如知识;三是概念与实物的结合体,如文献。于是有三种不同的分类方法,在信息组织的实践中。