



信息资源组织

主编 柳晓春 黄碧云 王美兰

世界图书出版公司

信 息 资 源 组 织

主 编 柳晓春 黄碧云 王美兰

编 者 (以姓氏汉语拼音为序)

陈先来	虢 毅	胡德华	黄碧云
李 蓓	李 俊	李忠民	柳晓春
谭宇红	王美兰	许四洋	张江华
左少凝			

世界图书出版公司

西安 北京 广州 上海

图书在版编目 (CIP) 数据

信息资源组织/柳晓春主编.

西安: 世界图书出版西安公司, 2003. 7

ISBN 7-5062-3799-7

I. 信... II. 柳... III. 信息管理
IV. G203

中国版本图书馆 CIP 数据核字 (2003) 第 054491 号

信息资源组织

主 编 柳晓春 黄碧云 王美兰
责任编辑 陈 进

出版发行 世界图书出版西安公司
地 址 西安市南大街 17 号
邮 编 710001
电 话 029-7279676 7233647(发行部)
传 真 029-7279675
E-mail wmcxian@public.xa.sn.cn
经 销 各地新华书店
印 刷 西安建筑科技大学印刷厂
开 本 787×1092 1/16
印 张 17
字 数 447 千字

版 次 2003 年 7 月第 1 版 第 1 次印刷
书 号 ISBN 7-5062-3799-7/R·627
定 价 29.50 元

☆如有印装错误,请寄回本公司更换☆

前 言

在信息时代,信息已成为社会发展的重要战略资源。组织化、系统化的信息在社会生产和生活中所起的作用日益突出。尤其是信息与网络的紧密结合,为人们构筑了一个崭新的虚拟空间,在这个空间,人们需要用知识与信息来解决在生存与发展中所面临的问题。谁掌握了知识与信息,谁就掌握了生存与发展的条件。知识就是力量,信息就是财富。

从学术意义上来说,信息资源就是知识资源,因此,信息资源组织实质上就是知识组织。狭义的信息资源组织指的是对文献的分类与主题标引、编目、文摘、索引等一系列整序活动;而广义的信息资源组织指的是以知识单元为对象的整理、加工、分析、推理、挖掘、控制、表达等一系列组织化过程及其方法。人们通常将知识划分为显性知识与隐性知识。显性知识又称编码知识或客观知识,指的是可以通过正常的语言文字方式传播的知识,典型的显性知识存在于如图书、期刊、数据库和网络等之中;隐性知识又称意会知识或主观知识,指的是存在于人们头脑中的思想、智慧、经验等非文献型的知识。本书所要研究的信息资源组织主要指的是显性知识的组织,即纸质信息资源与网络信息资源的描述与组织。

本书在总结和借鉴有关信息组织、知识组织专家学者丰硕的研究成果的基础上,详细地阐述了信息资源组织的基本理论(如原理、原则、流程、方法);信息资源的描述(如基本原理、一般方法、MARC及元数据);信息资源组织的具体方法(如分类法、主题法)及标引方法;情报语言的一体化及其研究成果(如后控词表、一体化医学语言系统);网络信息资源的组织(如分类与主题方式、存储方式、数据挖掘、搜索引擎、数字图书馆);信息资源组织的自动化(如词语的自动切分、自动标引、自动摘要、机器翻译)等。

本书各章节由下列同志负责编写:第一章,柳晓春、黄碧云、虢毅、张江华、许四洋;第二章,李蓓、许四洋、黄碧云、胡德华、李俊、柳晓春;第三章,谭宇红、柳晓春、黄碧云;第四章,李忠民、王美兰、柳晓春;第五章,王美兰;第六章,黄碧云、王美兰、李忠民、虢毅、柳晓春;第七章,李俊、陈先来、黄碧云、胡德华、左少凝、许四洋、张江华、李蓓、柳晓春;第八章,陈先来。全书由柳晓春、黄碧云、王美兰拟订编写大纲并统稿。

本书力图把握时代脉搏,力争反映最新研究成果,希望能够做到内容新颖丰富、深入浅出,适用于图书情报、信息管理专业学生的学习,也能为有关专家学者和图书情报实际工作者提供参考,但是由于编者水平有限,难免有许多错漏之处,敬请专家学者、读者批评指正。

本书的编写及出版得到了许多领导及专家的大力支持,参考了许多学者的研究成果和论著,在此表示衷心的感谢。特别要感谢的是中南大学医学图书馆颜传湘副研究馆员和中南大学医学技术与情报学院李后卿副院长对本书编写的大力支持,中南大学医学技术与情报学院方平教授对本书的有益指导,世界图书出版西安公司对本书出版的热情帮助与辛勤劳动。

编 者
2003年6月于长沙

目 次

第一章 信息资源组织概论	(1)
第一节 信息、信息资源和信息资源组织	(1)
第二节 信息资源组织原理	(4)
第三节 信息资源组织原则	(17)
第四节 信息资源组织流程	(18)
第五节 信息资源组织方法	(18)
第六节 信息资源组织研究内容	(21)
第二章 信息资源描述	(23)
第一节 信息资源描述的基本原理	(23)
第二节 信息资源描述的一般方法	(29)
第三节 MARC	(46)
第四节 元数据	(71)
第五节 标记语言	(78)
第三章 分类法概述	(81)
第一节 原理与结构	(81)
第二节 《中国图书馆分类法》	(93)
第三节 《杜威十进分类法》	(98)
第四节 《国际十进分类法》	(102)
第五节 《国会图书馆图书分类法》	(105)
第六节 《冒号分类法》	(109)
第七节 《国际专利分类法》	(113)
第四章 分类标引	(119)
第一节 标引	(119)
第二节 分类标引的一般方法	(125)
第三节 各学科分类标引规则	(134)
第五章 主题法概述	(152)
第一节 原理与结构	(152)
第二节 《汉语主题词表》	(166)
第三节 《国会图书馆标题表》	(170)
第四节 《医学主题词表》	(172)
第五节 主题标引的一般方法	(178)

第六章 情报语言的一体化	(186)
第一节 概述	(186)
第二节 分类法之间的转换	(188)
第三节 分类法与主题法的转换	(190)
第四节 后控词表	(196)
第五节 一体化医学语言系统	(200)
第七章 网络信息资源组织	(209)
第一节 网络信息资源组织的分类与主题方式	(209)
第二节 网络信息资源组织的存储方式	(214)
第三节 数据挖掘	(227)
第四节 搜索引擎	(234)
第五节 数字图书馆	(239)
第八章 信息资源组织的自动化	(245)
第一节 词语的自动切分	(245)
第二节 自动标引	(248)
第三节 自动摘要	(255)
第四节 机器翻译	(257)
参考文献	(261)

第一章 信息资源组织概论

第一节 信息、信息资源和信息资源组织

一、信息的概念

1. 什么是信息？

在人类社会早期，人们对信息（Information）的认识比较宽泛和模糊，常与消息（Message）一词混同。1928年 L. V. R. Hartley 在《贝尔系统电话杂志》上发表题为“信息传输”一文，将信息理解为选择通信符号的方式，并用选择的自由度来计量。1948年，信息论创始人 C. E. Shannon 也在该刊上发表题为“通信的数学理论”，认为信息是使不确定性消除的某种东西，并推导出信息测度公式。而控制论开创者维纳认为信息是人类在适应、控制外部世界的过程中同外部世界进行交换的内容和名称，是一类独立的研究对象，它既不是物质也不是能量。意大利 G. Longo 认为信息是反映事物的形成、关系和差别的东西，包含在事物的差异之中。

信息也一直是图书情报领域的研究热点。美国 M. Buckland 认为许多事物都可以是信息，如文本、图片、录音磁带、博物馆陈列品、自然物体、实验、事件等，其定义严格说只是描述了信息的范围。美国 Jean Tague - Sutcliff 从认识论角度认为信息是读者通过阅读或其它认知方法处理记录所理解的东西。美国 A. N. Smith 在《Information Resource Management》中认为信息是数据处理的最最终产品，是经过收集、记录、处理、以可检形式储存的事实或数据。我国著名学者严怡民教授将信息定义为生物以及具有自动控制系统的机器，通过感觉器官和相应的设备与外界进行交换的一切内容。《情报与文献工作词汇、基本术语》（GB4894 - 85）给信息下的定义是：物质存在的一种方式、形态或运动状态，也是事物的一种普遍属性，一般指数据、消息中所包含的意义，可以是消息中所描述事件的不定性减少。

如上所述，不同学科领域不同专家学者对信息有不同的理解，这也使信息成为一个具有多维和多层含义的概念。

2. 信息与知识、情报、文献

信息是物质的一种普遍属性，是物质存在的方式和运动的规律与特点，信息不仅包括人与人之间的消息交换，而且还包括人与自然界之间、自然界与自然界之间的消息交换。

知识是人们通过信息对自然界、人类社会以及思维活动规律的认识与掌握，是人的大脑通过思维重新组合的系统化信息的集合。

情报是作为交流对象的有用知识，是激活了的、活化了的的知识。

因此，知识是信息的一部分，是一种特定的信息，它来自于对信息的加工处理，因而比信息更具有系统性和概括性；情报来源于知识，是知识的一部分，是进入人类社会交流系统的运动着的知识。

国家标准《文献著录总则》对“文献”一词的定义是“记录有知识的一切载体”。

二、信息资源

1. 信息资源的概念及特征

信息社会最重要的资源是信息,信息作为一种资源是相对于物质资源和能量资源而言的,并与物质资源和能量资源一起构成现代社会经济发展的3大支柱。美国列维坦在题为“在信息生产生命周期中作为‘商品’的信息资源”一文中,认为信息资源是一系列已经制度化的、为一个或多个用户集团反复使用的信息。1996年的信息资源与社会发展国际学术研讨会则认为,信息资源不仅包括文献,还包括人、系统、经费等。

我国学者对信息资源的理解通常包括两个层次:狭义的信息资源一般仅指信息本身;广义的信息资源则是信息活动中各种要素的集合,不仅包括信息本身,还包括与其紧密相连的信息设备与资金、信息人员与信息技术、信息网络与信息系统等。狭义的信息资源反映了信息资源的核心和实质,而广义的信息资源有助于全面把握信息资源的内涵,有利于充分地发挥信息资源的作用,因而也更符合信息活动的实际。

信息资源具有以下基本特征:①依附性,即与载体的不可分割性;②转换性,即信息资源可以变换、加工和转移到不同载体;③传递性,即易传输性;④共享性;⑤时效性,即时间会影响信息资源的价值,如信息的老化;⑥无限性,即就其时间的延续和信息资源的储量而言是无限的;⑦增值性,即信息资源的投入可使自然资源、人力资源和资本资源等增值。

2. 信息资源的分类

可从不同角度对信息资源予以分类,如按介质可分为口头、实物、书写、印刷、视听、缩微和电子信息资源等,其中网络信息资源是一种重要的电子信息资源;按表达方式,可分为文献信息资源和非文献信息资源(包括口语、体语和实物信息资源等);按加工程度可分为零次信息资源、一次信息资源、二次信息资源和三次信息资源;按传播范围的受控程度可分为公开信息资源、半公开信息资源、非公开信息资源;按存在状态可分为潜在的信息资源(储存在人的大脑中尚未表达出来的信息资源)和现实的信息资源(已表达出来的、能够为公众所利用的信息资源);按作用层次,可分为战略信息资源和战术信息资源;按反映面,可分为宏观信息资源和微观信息资源;按内容,可分为政治、法律、军事、经济、管理、科技和医学信息资源等。

本书将信息资源分为显性信息资源与隐性信息资源两类。显性信息资源是指可以通过正常的语言文字方式传播的信息,它主要存在于图书、期刊、数据库和网络中,显性信息资源是可以表达的,有固定存在,可确定的。隐性信息资源是指不易用语言文字表达,难以记录,只能意会的信息。本书研究的内容主要是显性信息资源的组织。

三、信息资源组织

1. 信息资源组织及发展

(1) 信息资源组织的定义

信息资源组织,又称信息组织,是将无序的信息资源变为有序的信息资源的过程,具体来说是按照一定的科学原理、规则和方法,对特定信息资源的外在特征和内容特征进行分析、揭示和描述,并使其成为有序化的信息资源集合的过程。在具体实施过程中有不同的表现形式,如在图书馆可表现为具体文献的组织,即按照一定的原则和方法对图书馆文献资源的外部特征和内容特征进行揭示、区分以有效地分布在各个不同的阅览室或书库,再按索书号、架位号等有序地将文献资源(如图书、期刊)排列在书架上;还可表现为图书馆目录的组织,而图书馆

目录本身又是对馆藏文献资源特征的揭示;此外,在网络环境下还可表现为数字图书馆的组织,图书馆主页的组织,图书馆的馆藏数字化资源、数据库和图书馆收集的网上虚拟资源的组织。

(2) 信息资源组织的发展史

从信息资源组织的发展来看,历经三个时期,即清册职能(Inventory Function)时期、查检职能(Finding List)时期和组织职能(Organizing Function)时期。清册职能时期更多地是从收藏和保管角度对信息资源加以组织,它主要通过记录信息资源来告诉人们存在什么信息资源,因而并不要求科学、完备。查检时期的信息资源组织不仅要告诉人们存在什么信息资源,还要告诉人们怎样查找获取信息资源,因而开始重视对信息资源的外在特征和内容特征的描述和揭示,信息资源组织的理论以及各种传统的信息资源组织技术和方法有了很大发展。组织职能时期的信息资源组织以现代信息技术的发展为先导,它不仅告诉人们存在什么信息资源和怎样查找,而且还要告诉人们其他地方有什么信息资源,哪些信息资源适合于自己,并通过多途径、多角度去查检和获取,因而这个时期的信息资源组织已具备网络化组织信息资源的能力,并融信息描述、揭示、分析、存储和检索于一体。

2. 信息资源组织的意义

(1) 解决信息资源需求与信息资源利用困难之间的矛盾

科学技术的迅猛发展,一方面导致人类对信息资源需求越来越迫切,这种迫切不仅表现在对信息资源需求的数量,而且表现在对信息资源的质量和获取速度;另一方面导致对信息资源利用更为困难,这是因为科学技术的发展使信息资源具有以下特点:①增长迅速。以网络资源为例 1981 年 8 月 Internet 上共有 213 个网站,1991 年 1 月共有 376,000 个网站,2001 年 1 月共有 109,574,429 个网站,目前已将近 1.8 亿个网站。②载体多样化。以文献信息资源为例,其主要载体除传统的印刷型之外,还有录像带、录音带、电影、幻灯等各种视听型,缩微胶片、缩微胶卷等缩微型;磁带、磁盘、光盘、网络等机读型。③重复性大。同一信息可用不同形式(如专利文献、学位论文、期刊论文)、不同文字(如英文、中文)、不同载体(如印刷型、视听型、机读型)多次重复发布或发表,特别是网络信息,由于其易于复制和下载,重复现象更为严重。④质量参差不齐。从信息资源的整体来看,正式出版信息和非正式出版信息交织在一起,学术信息、商业信息和虚假信息混为一体,信息质量良莠不齐。这些均增加了信息资源的无序化程度。而信息资源组织就是解决信息资源的有序化问题,以满足用户对信息资源的需求,这是信息资源组织的根本动力所在。信息资源组织是科学技术发展的结果,同时也能极大地促进科学技术的发展。

(2) 是信息资源管理的重要环节

信息资源管理是以信息资源的有效利用为目的,以现代信息技术为手段,对信息资源实施计划、预算、组织、指导、控制和协调的人类管理活动。信息资源组织不仅是信息资源管理重要研究内容,而且还广泛渗透到信息资源管理的其它环节。

(3) 是信息检索利用的基础

信息资源组织是记录描述信息资源的特征并构建有序的信息资源系统的过程。信息检索则是从构建的信息资源系统中查出所需信息资源集合的过程,实际上是信息资源组织的逆过程。为保证较高的查全率和查准率,信息资源组织须采用一定的规则依据一定的规范如信息资源描述规范、情报语言等对信息资源进行处理、加工,信息检索也须按照系统采用的规范和使用的语言的特点进行操作。

(4) 是信息资源增值的过程

信息资源组织实际上是根据特定用户需求将杂乱无章的原始信息变成一个有序的精良的信息系统的过程,经过这一过程,零乱地、质量参差不齐的信息变成了有序的由高质量信息组成的信息资源集合体,信息资源被进一步过滤、浓缩,因而也更为有效。同时信息资源组织还可为信息资源的进一步增值即信息资源的分析研究打下基础。

第二节 信息资源组织原理

一、系统论

系统论 (General System Theory) 是 20 世纪 40 年代由美籍奥地利生物学家贝塔朗菲 (L. V. Bertalanffy) 创立的一门学科,是研究一切综合系统或子系统的一般模式、原则和规律的理论体系。目前,它已发展成为一种新的方法论流派,受到学术界的普遍重视,并在社会经济各领域中产生日益重要的作用。

信息组织是对整个社会信息系统进行信息整理和序化,实现无序信息流向有序信息流的转换,以保证用户对信息的有效获取和利用。社会信息系统是一个典型的人工系统,规模庞大,结构复杂,功能综合,因素众多,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,能使信息系统更加有序化,更好地发挥信息资源的作用。

1. 系统概念

“系统”一词源于古希腊语。贝塔朗菲将它定义为“相互作用的诸要素的复合体”、“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”。我国一般采用著名学者钱学森的定义,他根据系统具有同组成它的各要素迥然不同的特点,将其概括为“系统是由相互作用和相互依赖的若干组成部分结合而成的具有特定功能的整体”。

在系统概念中,构成系统的各个组成部分就是它的要素,系统所从属的更大系统,则是它的生存环境。一切系统都有结构和功能,是结构和功能的统一体。科学地把握系统概念必须注意以下三点:第一,任何一个系统,必须由两个或两个以上的要素构成;第二,系统内各要素之间、系统要素与系统整体之间是相互依存、相互制约、相互联系的有机整体;第三,系统作为一个整体对环境表现出特定的功能。

2. 系统论的原理体系

系统论的研究对象是事物的系统,它是抽去事物的具体形态和属性,从整体上考察和研究事物共同的系统方面的特征和性质。这些共同的特征和性质的客观存在,构成了系统论的原理体系。

(1) 整体性原理 系统的整体性指系统的各个要素按一定的方式构成有机整体,要素是整体的部分,系统的要素与整体、环境以及各要素之间相互联系、相互作用,使系统呈现出构成部分所不具有的功能,即系统客体在整体目标上的功能和属性,不等于其组成要素孤立状态时的功能和属性的迭加,而是在质上和一定量度上都“大于”其组成要素的总和。通过提高系统要素的基质和系统整体的组织化程度,能够提高系统的整体功能。

(2) 相互联系性原理 该原理认为系统的整体功能和属性来源于系统和环境、系统和要素、要素和要素之间的相互联系和相互作用。它们既是系统形成的“终极原因”,又是系统存在和发展的根本条件。在系统的诸联系中,结构联系和功能联系占有重要地位。系统的结构决定功能,

功能变化又可反作用于结构,并引起结构的变化。

(3) 层次性原理 系统的层次性是指系统由若干个子系统有机组成,子系统可由更小的子系统构成,从而形成一种多级别、多层次的有机结构。不同层次具有不同的性质,遵守不同的运动规律,这正是区分层次的重要标准。而各层次之间在一定条件下又可以相互作用和相互转化。

(4) 有序性原理 有序表示系统内部由于有规则的联系所形成的结构的趋势和系统稳定性的程度,无序则表示系统内部由于无规则的联系而丧失结构的趋势和系统不稳定的程度。有序性原理是指系统由于自身的运动和变化能够从无走向有序结构。它是系统通过信息的反馈作用,在同环境交换物质和能量的过程中,不断调整自己行为和活动的结果。一切系统的运动变化都是从有序到无序再从无序到有序的辩证发展过程。

(5) 动态性原理 又称历时性原理,指系统的状态随时间而发生变化。任何系统都是不断演化的,系统存在于过程之中。随着时间的推移,系统及其要素、系统所处的环境都在运动变化。另外,系统都是开放的,不存在绝对封闭的系统,任何系统在一定程度上都是开放的系统或半开放系统,要与外界进行物质、能量和信息的交流,有可能从较低有序走向较高有序。

3. 系统方法

系统方法是系统论的重要组成部分,是用系统性原理解决实际问题的方法论体系。它是从系统观点出发,着重从整体与要素之间、整体与外部环境之间的相互联系、相互作用中综合地、精确地考察对象的性质、特点和运动过程,以达到从整体上把握系统功能、结构和动态变化规律,实现最佳化处理问题的一种方法。在实际应用时,系统方法可分为系统分析、系统综合、功能分析法、结构分析法和系统工程法等具体的方法。

系统方法具有整体性、综合性及最优化的特点。整体性是把对象作为整体对待,从整体与部分相互联系、相互结合、相互制约的关系中揭示系统的特征和运动规律。综合性是把任何整体都看作是以诸要素为特定目的而组成的综合体;要求研究任一对象都必须从它的成分、结构、功能、相互联系方式、历史发展等方面进行综合的考察。最优化是根据需要和可能为系统定量地确定出最佳目标,并运用最新技术手段和处理方法把整个系统划分成不同等级和层次结构,在动态中协调整体与部分的关系,使部分的功能和目标服从系统总体的最佳目标,以达到总体最优。

二、自组织理论

自组织理论(Self-organization Theory)是由一系列理论构建起来的理论群。其中,耗散结构理论解决了自组织出现的条件环境问题,协同学回答了自组织的动力学问题,突变论从数学抽象的角度研究了自组织的途径问题,超循环论解决了自组织的结合形式问题,而混沌和分形理论则是从时序与空间序的角度研究了自组织的复杂性和图景问题,这些理论既相互独立又相互关联,在整个自组织理论体系中发挥着不同的地位和作用。

1. 耗散结构理论

耗散结构理论(Dissipative Structure Theory)是由比利时普里高津(I. Prigogine)于1969年的理论物理与生物学国际会议上首次提出。该理论认为,一个远离平衡态的开放系统,通过不断与外界交换物质、能量,在外界条件的变化达到一定阈值时,从原来的无序状态转变为在时间、空间或功能上的有序状态。由于这种稳定而有序的结构需要不断向外界耗散物质与能量才能维持,所以称之为耗散结构。

耗散结构理论提出了耗散结构即自组织系统形成的先决条件:

(1) 系统必须是开放的 信息系统与外界环境始终存在着物质、能量、信息的输入输出,与社会、经济和科学等大系统始终相互作用,互为存在的前提,因此,就其本身而言,它是一种开放系统。

(2) 系统处于远离平衡的状态 根据“远离平衡是有序之源”的原理,只有系统远离平衡状态时,才有可能形成有序结构。判断系统是否远离平衡态主要是看体系的各个组成部分是否均匀一致,体系的各个部分之间的差异越大,体系离开平衡态就越远。信息系统从诞生开始就在不断地更新,存在着各种不平衡。随着社会的进步,又会出现各种子系统,各个子系统也处于远离平衡的“活”状态。

(3) 系统存在非线性作用机制 判断一个体系是否非线性,首先要研究体系的构成是要素还是元素,即组成部分不仅在数量上而且在性质上要相互独立且有相当的差异;其次是组成体系的独立要素数量必须大于等于三。信息系统特别是网络信息系统,信息交流的程度大大提高,其内部复杂程度大为增加,各要素存在着非线性作用。系统内影响信息搜集、加工、整序和传递诸要素之间,不是线性累加的关系,而是相互结合而协同作用的。同时各要素受相互协同作用的影响程度不同,每个要素的变化对系统整体的作用效果也不一致。

(4) 系统通过涨落形成有序 耗散结构在同外界进行物质、能量和信息交换时,输出和输入不可能总是恰好平衡的,因此,耗散结构的存在必然伴随着不同程度的涨落起伏现象。涨落是耗散结构出现的触发器,但只有当它出现在体系远离平衡态的区域时,才能导致系统结构向有序性发展。信息系统本身也是以非稳定态为前提的,受政策、用户需求、管理体制和人员资金方面的影响,因而最终也会导致有序。

综合以上分析,信息系统本身已具备了产生耗散结构的条件,它远离平衡态,具有输入、输出,多次循环及反馈等开放性的基本特征,是一种具有耗散结构的自组织系统,因此可用自组织理论中的耗散结构理论来指导信息系统信息的有序组织。

2. 协同学

协同学(Synergetics)是由联邦德国物理学家哈肯(Hermann. Haken)于1970年提出,它主要研究由完全不同性质的大量子系统(如电子、原子、分子、细胞、神经元、器官、动物乃至人类社会等)是如何通过协作从无序到有序的演化的规律,是自组织动力学的方法论。

协同学认为,自组织系统演化的动力来自系统内部的两种相互作用:竞争和协同。子系统的竞争使系统趋于非平衡,而这正是系统自组织的首要条件;子系统的协同则在非平衡条件下使子系统内的某些运动趋势联合起来并加以放大,使之占据优势地位,支配系统整体的演化。

为了更好地描述系统宏观的有序性,哈肯引入了“序参量”的概念。序参量是在一定程度上能够测度系统协同程度的参量,它是系统通过各要素的协同作用而形成,是系统从无序到有序变化发展的主导因素,决定着系统的自组织行为。当系统处于无序混沌状态时,其序参量为零;当系统开始有序时,序参量为非零值,并且随着外界条件的改善和系统的有序程度的提高而逐渐增大。当序参量达到引起系统发生根本性转变的临界点时,一个很小数量的变化就会使系统出现宏观的新的有序结构。

信息系统是由许多子系统构成,各子系统之间必然存在相互竞争和相互协同作用,特别是对于网络信息系统,信息自组织现象产生的动力就是来自于产生信息的用户之间的竞争与协同。因特网上信息的自组织与用户的直觉思维过程息息相关,是用户创造性思维的一个衍生过程,如虚拟社区、搜索引擎、个人主页和数据库等。

利用协同学的支配原理和协同作用原理,系统的自组织行为是由其序参量决定的。研究信息系统从无序到有序的行为机制,关键在于发掘系统的序参量。信息系统的序参量包括三个,即信息累积变量、信息需求变量和信息序化变量。这三个变量的联合作用及其对其他系统变量的支配作用共同决定着信息系统的行为。因此,我们在进行信息组织工作时,重点在于处理好三者的关系,使它们协同合作,支配信息系统从无序走向有序,形成运行良好的自组织结构。

3. 突变论

突变论(Morphogenesis)是法国数学家勒内·托姆(R. Thom)于20世纪60年代提出的一种拓扑数学理论,它通过数学模型揭示和预测事物的连续性中断的质变过程。

突变论认为,系统结构的演化,即由一种稳定态转变为另一种不同质的稳定态,可以通过非连续的突变,也可以通过连续的渐变来实现,演化的方式依赖于演化条件。如果演化的中间过渡态是不稳定态,演化过程就是突变;反之则为渐变。突变时,系统的熵增或熵减过程可以在宏观无限小的时间内突然发生,而这种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”。原则上可以通过控制条件的变化来控制系统的演化方式,即渐变和突变在一定条件下可以相互转化。在实践中,人们常通过改变外界条件来使渐变和突变相互转化,以选择对系统最为有利的演变方式。

突变论对信息系统管理提供了有益的启示。信息系统内部各子系统之间存在着非线性的相互作用,这些作用导致系统序参量的变化,我们可以通过控制条件来控制序参量变化的方式。当我们需要系统稳定时,就要采取一切可能的方法,使系统的演化以渐变的方式进行;当系统的功能严重衰退,远远不能满足社会和经济发展的需要时,我们要创造条件,不失时机地促使系统的突变。

4. 其它自组织理论

自组织理论群中还包括超循环论(Hypercycle Theory)、混沌理论(Chaotic Theory)和分形理论(Fractal Theory)。

超循环理论强调自组织形式是一种通过竞争选择而逐渐综合的过程,也就是将存在着差异、相互矛盾的部分联合统一为一种优势兼容的循环系统。超循环不仅是一种形式上的循环系统的整合,而且是一种功能性的综合。由于信息具有一定意义上的共享性,因而其循环利用具有重要意义。实际上,今天的因特网是具有最大超循环结构和功能的“超循环圈”,信息的超循环利用是创新的基本要素。超循环组织是保持信息的稳定性,并促使其继续进化的一个必要前提。

混沌理论阐述了非线性系统中混乱信息走向有序的可能性和现实性,指出混沌既不是简单的无序,也不是简单的有序,而是无序中的有序。因特网上的信息基本上是以自组织的混沌方式运行,其中的信息混沌有两类:一类即平衡态热力学性质的信息垃圾,另一类就是有结构,形成序列的信息规则下的网络信息。面对铺天盖地的混沌性信息,信息组织工作者一方面须制定一些通用标准、协议,另一方面也要保持网络的非中心性。

5. 自组织理论与信息组织

近几十年来,由于信息总量的持续增长,信息技术特别是网络技术的飞速发展,人类信息观念的转变,使信息系统显著地具备了自组织的条件,而自组织理论和方法的成熟与发展拓展了信息组织的内容,开辟了信息组织的新方法和新思路。

根据自组织理论,信息组织应分为信息被组织和信息自组织。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、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。信息自组织的结果有序但不唯一,是信息运动中的一个可能的发展方向。信息被组织是指信息的组织是在外界特定干预下进行的。信息自组织与信息被组织相比,虽然两者目的与结果相一致,但人在其中的地位和作用不同,信息被组织需要人的能动性和目的性去干预和指挥。信息自组织过程中,个人或许带有私人目的,但作为总体的人是没有外部指令或者说处于受动状态的。

以自组织理论为指导,在进行信息组织时我们不能对信息的自组织现象视而不见,而应充分利用。因此,要妥善解决信息的无序、冗余、不稳定的现状,不能依靠单一的信息被组织和信息自组织,必须在两者有机结合的基础上对信息予以有效组织,即信息微观组织-信息自组织-信息的整体协调与宏观组织,使信息组织真正发挥整序信息、科学分流、促进选择、保证利用的功能和作用,形成健全完善的信息组织体系和顺畅通达的信息组织流程。

三、语言学基础

语言是社会约定俗成的表达观念的符号系统。有了语言,科学研究的成果得以广泛交流,前人宝贵的文化遗产得以代代相传,人类社会得以不断进步。语言不但是思维工具,而且是交流工具,是信息的载体。

要把庞杂、分散的信息组织成有序、优化的整体,要实现信息的有效传递与利用,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。于是,在自然语言的基础上,各种与信息有序组织有关的人工语言如情报检索语言和计算机语言便应运而生。尽管这些信息组织符号系统的形式不同,但都和自然语言一样,有着共同的特征:有语词,有词汇,有语法,这就是信息资源组织的语言学基础。

1. 语义

(1) 词义、义位与义素

语言学中的词义是指词的声音形式所负载的信息内容。它包括概念意义、感情意义、含蓄意义、语体意义和搭配意义等多个方面,其中,概念意义是词义的核心内容。概念意义是对词所指对象的区别性特征的概括认识,它反映词与所指对象的对应关系。因此,信息组织符号系统在优选和运用语词时应准确地把握语词的概念意义。一般实词都有概念意义,而虚词没有,因此虚词被排除在选词范围之外。而且由于语词要能直接、精确地表达知识单元所涉及的事物或主题,因此语词往往又限于能表达事物、主题的名词术语或名词性词组。

一个词的意义往往是一个复杂的整体,包括若干相对独立而又互相联系的语义单位——义位。义位是能独立运用的、概括的、固定的语义单位,一个义位一般表示一个概念。一个词如果有多个相互关联的义位就是多义词。例如,“待遇”一词有三个义位:①对待人的情形、态度、方式;②指权利、社会地位等;③物质报酬、工资福利。因此,在对词汇进行控制和规范时,就需要对语词的义位进行分析,决定取舍,或对其范围加以限定,即加限义词,例如:词(语言学);词(文学)。这样才能保证语词与概念一一对应的关系。

义素是构成义位的语义成分,是从一组相关的词中抽象出来的区别性语义特征。义素分析就是把词的义位分解为若干义素的组合。例如,“父亲”这个词的义素分析结构式为[男性+长辈+生育关系+亲属]。义素分析不仅能够比较准确地描述和解释语义,给词所表示的概念下定义,还能反映词与词之间的关系。共同义素越多,词义越接近;义素完全相同,则是同义关

系；若一个义素分析结构式包含了另外一个结构式，则二者之间构成了上下义关系。例如，“女人”与“女子”的义素分析结构式同为 $[-\text{男性} + \text{成年} + \text{人}]$ ，因此二者是同义关系。“亲属”的义素分析结构式为 $[+\text{血缘关系或婚姻关系} + \text{人}]$ ，“父亲”的义素结构分析式中包含了“亲属”的义素结构分析式，那么二者具有上下义关系，“亲属”是“父亲”的上义词。可见，义素分析对优选和规范语词，揭示语词间的各种关系，构建词汇的语义网络系统具有极其重要的意义。

(2) 语义场

语义场是指具有共同义素的词语组成的集合，例如：足球、篮球、排球、乒乓球、保龄球等等。这些词都有一个共同的义素，就是“运动用球”。由共同义素聚合在一起的词语形成一个语汇网络，彼此之间有着纵横的联系。例如：

动物

无脊柱动物

.....

脊柱动物

两栖类动物

鸟类

鱼类

爬行动物

哺乳动物

该例中，“动物”是最大的语义场，它包含两个较小的语义场“无脊柱动物”和“脊柱动物”，其中“脊柱动物”又包含五个更小的语义场。这些语义场还可层层上推或下推，如“动物”上面还有更大的语义场“生物”，“哺乳动物”下还有更小的语义场“人类”。一般把较大语义场称为其下层较小语义场的母场，而较小的语义场则称为其上层较大语义场的子场。这样层层划分就形成了一个巨大的语汇系统。这实际上就为信息资源的分类组织提供了理论依据。分类组织就是以某种属性（义素）为标准来层层划分类（语义场）从而构成一个分类体系（语汇系统）。上例中的纵向母场与子场的关系实际上就分类体系中的上下位类关系，横向的关系就是同位类的关系。上例中的“鱼类”还可与“肉类”、“禽类”和“蛋类”等词语同属另外一个语义场“食品”，这也反映了语词的多向成类的特点。

(3) 语义关系

从语义上分析，词语间的相互关系可分为上下义词、类义词、同义词、反义词。

两个词在概念意义上有包含和被包含的关系，这两个词就是上下义词。例如，“学校”和“大学”互为上下义词。上下义词是相对而言的，一个词对甲词是下义词，对乙词则可能是上义词。如“大学”是“学校”的下义词，它又是“理工大学”的上义词。上下义词又分为直接上下义词和间接上下义词。直接上下义词就是最邻近的两个语义层次所包含的词；间接上下义词是超越了最邻近的两个语义层次而形成的上下义词。例如，“艺术”与“音乐”属于直接上下义词，“艺术”与“古典音乐”则属于间接上下义词。间隔的层次越多，语义的差异就越大。语义层次越往上越泛指，越往下越专指。上下义词的区分对于信息资源的分类和主题组织有着重要意义。

广义的类义词是指属于同一个语义场、表示同种概念的所有词，包括上下义词、同义词和反义词。狭义的类义词则不包括上下义词、同义词和反义词。类义词必须是具有共同义素，而

且必须是表示同类概念,属于同一语义范畴的词。例如,“苹果、梨子、香蕉、桃子、西瓜……”都是属于水果这一语义场的类义词。类义词反映了词语之间横向的联系,它实际上就相当于分类体系中的一组同位类。

同义词是指意义相同或相近的一组词。同义词通常被分为等义词和近义词,等义词是指概念意义完全相同的一组词,例如“马铃薯”和“土豆”。近义词是指概念意义基本相同但又有细微差异的一组词,例如“经济扩张”和“经济侵略”。为了达到语词与概念的一一对应,同义词往往是词汇控制的对象,需要做出取舍。通常优选比较概括的、通用的和比较正式的词。例如前面两组词中,前面的词为叙词,后面的词为非叙词。

反义词是指意义相反或相对的一组词。它是客观事物或现象的矛盾对立在词汇中的反映。反义词又分为正反反义词和两极反义词。正反反义词又叫绝对反义词,两词之间的关系是一种非此即彼的关系,不存在中间状态的概念,肯定A必然否定B,否定A必然肯定B,例如“生”与“死”。两极反义词又叫相对反义词,这类反义词是表示某些事物或现象相互对立的两端,两词之间一般存在着某个中间状态的概念,肯定A必然否定B,但否定A却不能肯定B,例如“黑”与“白”。一组反义词必然有相同的义素,某些反义词所表达的概念虽然处于反对关系中,互相排斥、互相否定,但实际上是从不同的角度来说明同一问题,因此又相互依存。在信息资源组织中人们往往将其作为同义词处理,可避免同类信息的分散。这种情况下一般优选正面词、肯定词。例如“稳定性”和“不稳定性”这一对反义词中,前者为叙词,后者为非叙词。

根据信息的内容特征对信息进行描述、组织的方法就是语义信息组织法。目前最常见的语义信息组织法有两种,一是分类组织法,它是用分类号来表达各种概念,将各种概念按照学科性质进行分类和系统排列。另一种是主题组织法,它直接用语词作为表达主题概念的标识。

2. 语法

语法是语言的结构规则,是把各种语言成份组织成词句、篇章的各种规则的总和,它既是正确表达的基础,也是正确理解的基础。信息组织中的语法是指如何创造和运用标识来正确表达信息内容,并将这些标识按照一定的结构结合成一个有机的统一体的一套规则。与语言学中的语法一样,信息组织中的语法也包括词法、句法和篇章语法。信息组织用的各种标识(语词)都有其特定的构造规则和使用规则,如分类号的编制规则,主题词优选规则,这就是它的词法。使用多个标识进行组合来正确表达主题概念的规则就是它的句法,例如同级词组配、方面词组配。将相关标识结合成一个有机的统一体的规则就是它的篇章语法,例如分类表、叙词表的编制规则。

3. 语用

语用是指人们在一定的语境中对于语言的实际运用活动,特别是话语表达和理解等活动。语用学和语义学都是研究语言的意义,但语义学研究的是不受语境影响的词语、句子等本身的意义,而语用学研究的是在语境中才能确定的意义,即语言使用上的意义。信息同样如此,随着环境与使用者的不同,它的意义和重要性也有所不同。例如,在我国的《中国图书馆分类法》中,马克思主义、列宁主义、毛泽东思想位列二十二大类之首,而在《杜威十进分类法》中则无此大类,这就是它们各自的使用环境不同的结果。分类表中交替类目的设置正是基于语用学的考虑。例如,《精神障碍与犯罪》一书在法律图书馆中可归入D917,在医学图书馆中则可归入R749。

4. 信息组织语言的标准化

在信息组织过程中,世界许多国家和地区从方便本国和民族应用的角度出发,制定了相应

的编码标准和内码体系,如日本的 JIS X 0208 和 JIS X 0212,韩国的 KSC5601 和 KSC 5651 等。为了让各国文字透过不同语言文字平台,在计算机屏幕正确显示而不会出现乱码,真正实现信息的自由传递和共享,国际标准化组织(ISO)正在展开有关 Unicode 单一字码体系的研究。汉字部分,ISO 下设立汉字工作组(IRG),成员包括中国大陆、台湾、香港、日本、南韩、北韩、新加坡、越南及美国九个国家和地区,共同草拟汉字标准。我国信息产业部和国家质量技术监督局于 2000 年联合发布了 GB18030-2000《信息技术信息交换用汉字编码字符集基本集的扩充》。该标准作为国家强制性标准对我国中文信息技术和民族软件产业的发展具有深远影响。它是我国继 GB2312-1980 和 GB13000-1993 之后最重要的汉字编码标准,是未来我国计算机系统必须遵循的基础性标准之一。它收录了 27484 个汉字,采用单/双/四字节混合编码,总编码空间超过 150 万个码位,而且繁简字处于同一平台。在大字符级的国际标准得到采用后,它将解决两岸三地间 GB 码与 BIG5 码字码转换不便的状况,届时大陆、香港和台湾将处于统一的文字平台。这将是我国电脑文字标准走向世界标准的重要步骤。它还收录藏文、蒙文、维吾尔文等主要少数民族文字,为推进少数民族地区的信息化奠定了坚实的基础。它全面兼容 GB2312,在字汇上兼容 GB13000.1,可以充分利用已有资源,保证不同系统间的兼容性,最大限度地共享资源,并使我国的中文信息技术再上一个台阶。

四、逻辑学基础

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,它是对信息的各个元素(包括外部特征和内容特征),在事物——概念——语言符号的基础上,进行科学思维和合理排序的过程。因此,信息组织工作必然要研究、运用思维形式及其规律,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序化质量。

思维具有形象思维和抽象思维两种形式,人们的认识发展过程是由形象思维到抽象思维,由感性认识到理性认识的过程。感觉、知觉和表象是形象思维形式,属于感性认识的阶段;概念、判断与推理是抽象思维形式,属于理性阶段。作为信息资源的主体的知识情报是人的大脑通过思维重新组合的、系统化的信息集合,是人类对自然界、人类社会以及思维方式与运动规律的认识与掌握,因此,信息的有序组织也就属于抽象思维的范畴,而抽象思维是在各种概念的基础上进行的。事实上,信息组织的符号系统,无论是语词的还是符号的,都是表达一系列概括文献情报内容的概念及其相互关系的概念标识系统,因此,它们都是建立在概念逻辑或词项逻辑的基础之上的。

1. 概念

概念是人们对事物的认识上升到理性阶段的产物,是对事物本质属性的概括。科学认识的成果都是通过各种概念来加以总结和概括的,所以知识信息的内容也只有用概念才能科学地加以揭示。在日常思想交流中,概念是用各种自然语言的语词来表达的。但这些自然语言的语词除了具有理性意义,表达一定的概念外,还往往具有情感、语气方面的意义,而且这些语词往往是多义的,可以指称不同的事物,表达不同的概念。而在信息组织过程中,往往需要保证语词(标识)和概念的一一对应,这就需要抽象掉这些语词的非逻辑意义,并使其意义确定下来,使其一方面指称特定的事物(对象),另一方面表达特定的含义。这两个方面就构成了概念的两个逻辑特征:概念的内涵和概念的外延。

概念的内涵就是它所反映的事物的本质或特有的属性,即概念的含义或涵义。例如,“由立法机关制定的,国家政权保证执行的行为规则”这一本质属性就是“法律”这一概念的内涵。