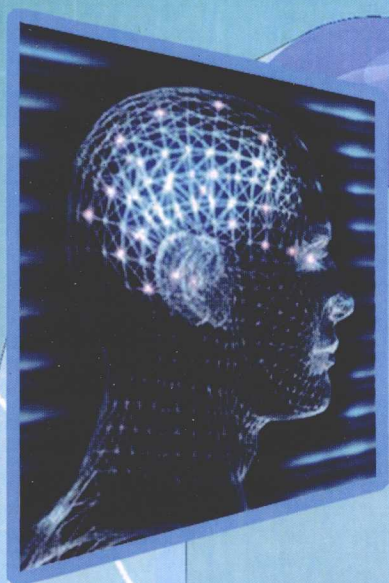


知识组织理论与方法

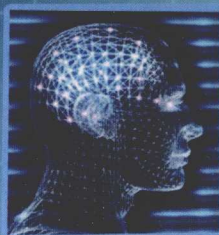
ZHISHI ZUZHI LILUN YU FANGFA

王知津 李培 李颖 等著

知识产权出版社

责任编辑：李德升

封面设计：王芳



知识组织理论与方法

上架建议：知识学类

ISBN 978-7-80247-477-2



9 787802 474772 >

ISBN 978-7-80247-477-2/G·291

(10219) 定价：30.00 元

教育部人文社会科学研究“十五”规划项目
项目批准号：01JA870002

知识组织理论与方法

王知津 李培 李颖 等著

知识产权出版社

内容提要:

本书在已有研究成果的基础上,全面、系统、深入地探讨网络环境下知识组织的规律和特点;知识组织的学科基础、基本原理、方法与技术;知识组织工具的结构、构建、维护及应用;网络环境下知识组织的新概念、新理论和新工具;以及知识组织的法律和经济环境等问题。

责任编辑:李德升

图书在版编目(CIP)数据

知识组织理论与方法/王知津,李培,李颖等著. —北京:知识产权出版社, 2009. 4

ISBN 978-7-80247-477-2

I. 知… II. ①王…②李…③李… III. 知识学—研究 IV. G302

中国版本图书馆 CIP 数据核字 (2009) 第 044682 号

知识组织理论与方法

王知津 李培 李颖 等著

出版发行: 知识产权出版社

社 址: 北京市海淀区马甸南村 1 号

网 址: <http://www.ipph.cn>

发行电话: 010-82000893 82000860 转 8101

责编电话: 010-82000860-8127

印 刷: 知识产权出版社电子印制中心

开 本: 880 mm×1230 mm 1/32

版 次: 2009 年 6 月第 1 版

字 数: 300 千字

ISBN 978-7-80247-477-2/G · 291

邮 编: 100088

邮 箱: bjb@cnipr.com

传 真: 010-82000893

责编邮箱: Ldsnk@126.com

经 销: 新华书店及相关销售网点

印 张: 11.5

印 次: 2009 年 6 月第 1 次印刷

定 价: 30 元

版权所有 侵权必究

如有印装质量问题,本社负责调换。

前 言

“知识组织”这一概念最早是国外于 20 世纪 60 年代中期提出来的。1989 年创立了国际知识组织学会 (International Society for Knowledge Organization, ISKO) 之后, 这一新的研究领域逐渐受到重视, 并开展了一系列学术研究和实践活动, 出版了一批会议录。1993 年, 专门致力于知识组织研究的国际性学术期刊 *Knowledge Organization* 面世, 成为知识组织研究者交流成果的平台。我国学术界对知识组织的研究始于 90 年代初, 近 20 年来, 发表的学术论文数量呈逐年递增的趋势, 近三年达到了高峰, 此外, 还有 10 余篇博士、硕士论文将研究对象聚焦于知识组织。然而, 经文献调查得知, 至今, 无论是国内还是国外, 尚无比较全面、系统地阐述知识组织研究体系的专著出版, 与学术论文的热度相比, 这无疑是一种遗憾。问题在于, 虽然目前所见的学术论文基本上可以覆盖现代知识组织的研究领域, 其研究水平也已经达到了一定的深度和广度, 但大多比较分散和零碎, 缺乏完整性、系统性, 还没有建立起知识组织的理论与方法体系。

为了迎接知识经济的挑战, 在已有研究成果的基础上, 全面、系统、深入地探讨网络环境下知识组织的规律和特点, 知识组织的学科基础、基本原理、方法与技术, 知识组织工具的结构、构建、维护及应用, 网络环境下知识组织的新概念、新理论和新工具, 以及知识组织的法律和经济等环境问题。2001 年, 我们申请的“网络环境下知识组织理论与方法体系构建研究”课题被批准为教育部人文社会科学研究“十五”规划项目 (项目批准号: 01JA870002), 试图以此为契机创立知识组织理论与方法的学科体系, 撰写并出版我国第一部有关知识组织的研究专

著,使之对于指导网络环境下的知识组织实践活动具有指导意义,对于创建知识组织的学科体系具有理论价值。然而,遗憾的是,由于种种客观原因,该项目的最终研究成果迟迟未能出版。今天,我们欣慰地看到,在各方的积极努力下,这一研究成果终于问世了。

本书共四个部分十二章,第一部分包括第一章“知识与知识组织”和第二章“知识组织的学科基础”,论述了知识组织的理论问题;第二部分包括第三章“分类系统与叙词系统的知识组织”和第四章“分类与标引的知识组织”,阐述了传统的知识组织工具与方法,同时,它们也构成了现代知识组织的基础;第三部分包括第五章“自然语言处理的知识组织”,第六章“知识表示”,第七章“知识获取”,第八章“元数据”,第九章“知识挖掘”,第十章“知识本体”,以及第十一章“主题地图”,着重探讨了现代知识组织工具与方法;第四部分即第十二章“知识组织的环境”,重点讨论了知识组织的法律和经济问题。

该研究成果的完成和出版是课题组全体成员共同努力的结果。项目主持人王知津提出了本课题的研究思路,进行了整体设计和具体策划,课题组成员发扬了团队精神,互助合作,协调一致,圆满地完成了这项研究任务。撰写本书的具体分工如下:第一章(王知津)、第二章(李颖)、第三章(郑红军)、第四章(郑红军、蒋伟伟)、第五章(孙鑫)、第六章(蒋伟伟)、第七章(李培)、第八章(路彩妹)、第九章(李明珍)、第十章(赵洪)、第十一章(赵洪)、第十二章(孙立武)。谢瑶为部分章节整理了大量资料。李培和赵洪分别初审了部分章节,王知津负责全书的终审、统稿和定稿,并翻译了两个附录。

在本书的撰写过程中,参考和借鉴了大量的中外文献资料,我们对本书所列主要参考文献的作者表示诚挚的谢意。由于篇幅所限,未能一一列出所有参考文献,因此,我们对未能列出的参

考文献作者表示深深的歉意和诚挚的谢意。正是这些参考文献作者的前期工作为本课题和本书的完成奠定了基础，并为我们提供了强大的研究动力和丰富的创新素材。最后，真诚地感谢教育部的评审专家给了我们知识组织进行深度研究的宝贵机会。

我们尽力把本课题做好，但知识组织毕竟是一个涉及面广又快速发展的新兴领域，限于课题组成员的学识、水平和能力，有些问题的研究还不够透彻，对于缺点、错误和疏漏之处，恳请各位专家、学者和广大读者不吝赐教，以便我们在后续研究中参考和借鉴。

王知津

2009 年 2 月于南开大学

目 录

第一章 知识与知识组织	(1)
第一节 从知识到知识组织	(1)
第二节 知识组织中的知识	(6)
第三节 知识组织的基本理论	(18)
第二章 知识组织的学科基础	(31)
第一节 知识组织与逻辑学	(31)
第二节 知识组织与数学	(40)
第三节 知识组织与系统论	(42)
第四节 知识组织与心理学	(48)
第五节 知识组织与知识工程	(53)
第三章 分类系统与叙词系统的知识组织	(61)
第一节 分类系统与叙词系统基本理论	(61)
第二节 分类系统与叙词系统的结构及要素	(67)
第三节 分类系统与叙词系统的构建	(76)
第四节 关系	(79)
第五节 标记符号	(83)
第六节 分类系统与叙词系统的兼容性和一致性	(86)
第四章 分类与标引的知识组织	(92)
第一节 分类与标引理论	(92)
第二节 分类和标引技术	(97)
第三节 自动分类与标引	(99)
第四节 网络环境下的分类与标引	(112)
第五章 自然语言处理的知识组织	(121)
第一节 自然语言处理与知识组织	(121)

第二节	自然语言处理中的语法分析	(126)
第三节	自然语言处理中的语义分析	(135)
第四节	自然语言处理中的语用分析	(144)
第六章	知识表示	(150)
第一节	知识表示的概述	(150)
第二节	规则表示方法	(153)
第三节	结构化表示	(157)
第四节	面向对象表示法	(168)
第五节	其他知识表示法	(174)
第七章	知识获取	(178)
第一节	知识获取的过程	(179)
第二节	知识获取方法	(183)
第三节	知识获取中的机器学习	(188)
第四节	知识获取工具	(197)
第八章	元数据	(207)
第一节	元数据的概念与特征	(207)
第二节	元数据格式	(217)
第三节	元数据的相关问题	(224)
第四节	元数据在知识组织中的应用	(233)
第九章	知识挖掘	(237)
第一节	数据挖掘与知识发现	(237)
第二节	知识挖掘技术	(242)
第三节	Web 知识挖掘	(251)
第四节	文本知识挖掘	(259)
第十章	知识本体	(267)
第一节	知识本体概述	(267)
第二节	知识本体的描述语言与编辑工具	(275)
第三节	知识本体的构建	(279)

第四节	知识本体的知识组织应用	(287)
第十一章	主题地图	(296)
第一节	主题地图	(296)
第二节	主题地图的 TAO 架构	(300)
第三节	主题地图的描述语言 XTM	(305)
第四节	主题地图的创建与在知识组织中的应用	(310)
第十二章	知识组织的环境	(322)
第一节	知识组织的方法论问题	(322)
第二节	知识组织的专业问题	(326)
第三节	知识组织的经济问题	(331)
第四节	知识组织的法律问题	(334)
第五节	知识组织工作标准化	(338)
第六节	知识组织的相关机构	(342)
第七节	知识组织的教育与培训	(345)
主要参考文献		(351)

第一章 知识与知识组织

第一节 从知识到知识组织

1.1.1 知识：力量的象征

第一个提出“知识就是力量”这个口号的人是17世纪英国哲学家法兰西斯·培根（Francis Bacon, 1561 ~ 1626）。面对封建神权对知识的鄙薄和贬损，面对新柏拉图主义怀疑论者对知识能否真正获得表示怀疑，面对社会上对知识的各种偏见，培根首次提出“知识就是力量”。这反映了人类在这个知识匮乏的时期渴望利用科学知识，认识自然、征服自然、发展生产的思想。培根不再将知识仅仅看作一种纯粹服从于心灵需要的智力活动，他强调知识的实用目的，这反映了近代以来科学革命与工业革命的特征。前苏联曾出版过一本相当有影响力的科普杂志，名字就叫《知识就是力量》。几个世纪以来，这个响亮的口号一直回荡在全球。

“知识就是力量”这个口号已被无数事实所证明。假如不懂得地质学，人们就不会知道960万平方公里土地下的宝藏；假如不懂得信息科学，人们就会变成耳聋眼花的现代人；假如不懂得基因科学，人们就不能克服遗传障碍，满足人类生存的需要……知识之所以具有改造世界的神奇力量，是因为知识与自然资源和资本相比具有极大的优越性，自然资源和资本要变成财富，离不开人类的劳动，而劳动又需要有知识的参与。

一般地说，知识是人们在改造世界的实践中所获得的认识和经验的总和；按照西方的解释，知识是通过实践、研究、联系或调查获得的关于事物的事实和状态的认识，是对科学、艺术或技术的理解，是人类获得的关于真理和原理的认识的总和。

知识是力量的一个非常重要的组成因素。促进社会发展的力量有多种，但所有的力量都是由掌握一定知识的人来发挥作用的。无论这些人是掌握着公共权力进行社会管理，还是在社会的其他领域为社会提供服务，只有努力用科学知识把所有的个体、群体武装起来，才能形成发展的力量。

在过去，知识的价值没有得到充分的体现，是因为社会充斥着简单劳动，当自然资源的开发不需要复杂的知识和技术时，知识或智慧的潜能就被掩盖。即使在工业大生产时期，机械化的大生产也并不需要太复杂的知识和技能，劳动者经过简单的培训就可以迅速上岗。当人类的社会劳动变得越来越复杂时，知识的重要性就日益体现出来。知识不只是获取财富的工具，本身也成了开发不尽的资源；知识不仅成了劳动的工具，还成了劳动的对象。

知识的作用是巨大的。热力学的发展把人类带进了蒸汽时代；电磁理论的建立促进了电动机和发电机的发明，人类又从蒸汽时代进入了崭新的电气时代；原子科学的创立，导致激光技术、计算机技术以及自动控制技术的迅速发展，极大地改变了人类生活的面貌，从而显示出知识的巨大威力。

随着知识经济的兴起，知识在现代社会价值的创造中其功效已远远高于人、财、物这些传统的生产要素，成为所有创造价值要素中最基本、最重要的要素，这已成为不争的事实。知识经济是以知识为基础的经济，这充分揭示了知识对现代经济增长的基础性作用。现行的工业经济和农业经济，虽然也离不开知识，但总的说来，经济的增长取决于能源、原材料和劳动力，即以物质

为基础。“知识经济”这个术语源于对知识，特别是科学技术知识在经济增长中的巨大作用的充分理解，把科学、技术和经济很好地紧密联系起来。归根到底，知识经济的中心含义是：知识已成为经济的核心。

1.1.2 知识组织：提供有序的知识

信息管理经过漫长的文献管理阶段、技术管理阶段、资源管理阶段进入知识管理阶段，由传统的文献存储与整序到信息流的控制，依托信息技术的支撑，逐渐向以用户信息需求为中心的知识流控制靠拢。知识经济的发展主要依赖知识创新，知识创新是知识经济的引擎，也是信息产业发展的生命力所在。知识经济的产生推动了知识组织的研究与发展，而知识组织则为知识管理和知识创新提供有序的知识。

在当今社会，每天都有大量信息产生，相应地，作为信息载体的出版物，也像潮水一样排山倒海地向人们涌来，人们常常把这一趋势称为“信息爆炸”、“信息泛滥”、“信息过剩”、“信息污染”等。但是，必须清醒地认识到，如今泛滥成灾并将我们淹没的不是知识，而是信息。

众所周知，信息在各个学科领域以指数增长。然而，当我们翻阅旧词典或旧手册时，往往感到惊讶的是，人们很早以前就已经知道某件事了。这表明，一方面，每一代人都为知识的创造和存储作出了贡献，并使知识在不断地增长着；另一方面，知识的增长速度远不如信息，也就是说，两者不是在同步增长。不过，尽管知识的增长速度较慢，但也毕竟给人们对知识的获取和利用带来了麻烦。

对我们来说，知识还是缺少的，其原因不在于很难获取必要而基本的知识和大量而准确的数据，而在于知识的存储过于庞大和无序，造成这种状况的归根结底还是人类自己。由于信息污染的不断加剧，造成知识存储无序化，知识存储无序化又推动信息

污染。这种恶性循环表明，科学本身越来越难于利用自己的成果。由于知识广泛分散在大量出版物之中，因此，导致知识存储的无序化。而人的阅读、吸收和消化能力是有限的，但出版物的增长却是无限的。面对信息污染，人们只能吸收某个或某些方面的信息，而信息又在不停地增长，为了掌握这个或这些方面的全面情况，人们不得不连续缩小自己的专业领域，致使人们的视野越来越狭窄，属于自己的领域越来越少。与此同时，人们疏远了更多的领域，加上新领域层出不穷，在这些领域，人们变得越来越肤浅、无知和无能。

因此，在一个人身上往往体现出双重性：一方面，他是某个或某些领域的专家和能手，另一方面，他在另一个或另一些领域却是外行。事实上，人们面临的就是这样一种选择。当今，人们对知识越来越感到难以把握，因此，如何有效地组织知识，提高知识的利用率，已成为亟待解决的问题。

现在，人们普遍认识到知识整序的必要性。但在古代，图书馆员和哲学家被排斥在知识整序领域之外。后来，百科全书的作者和教育家开始从事这项活动。接着，文献学家和情报学家加入了这个行列。再后来，术语学家也开始参与这件事。现在，人工智能、专家系统、超媒体以及教育专家都对知识组织产生了浓厚兴趣。

1.1.3 知识组织的实质和意义

从发生学角度看，人类的知识组织活动源于对客观知识无序化状态加以控制并使其有序化的愿望。也就是说，如果没有客观知识无序化状态的存在，便不会产生知识组织活动。从知识组织活动的目的看，满足人类客观知识主观化的需要是知识组织活动的最终目的。因此，我们可以说，知识组织的实质是以满足人类的客观知识主观化的需要为目的、针对客观知识的无序化状态所实施的一系列有序化组织活动。

据发达国家一些科技组织的统计,科技人员在一个科研项目中,用于查找和阅读相关资料的时间已占整个项目完成时间的一半以上。以至于越来越多的研究人员认为,与其花费大量的精力与时间到文献中去查找有关成果的报道,还不如以很高的代价去重新发现这些成果。知识的增长是堆积的,同时,虚假知识、外围知识与核心知识之别使知识存储无序,给用户利用知识带来困难,该过程呈现恶性循环。西方一位哲学家有一则名言:“我们淹没在信息中,却有着知识的饥饿。”这正是知识组织产生的原因所在,知识组织正是抑制这种恶性循环的措施。

知识组织的意义表现在以下几个方面:

(1) 认识论意义。客观知识对人类的认识发展具有极其重要的意义。人类可以通过掌握客观知识,间接认识客观世界,拓宽认识视野,摆脱“事必躬亲”式的实践方式的束缚,从而加速人类认识客观世界的进程。不过,客观知识必须经过主观化过程转化为主观知识(这一过程即为客观知识的主观化),才能成为人们进一步认识和实践的现实推动力。然而,客观知识作为脱离人脑而存在的客观性存在,虽不受个体的直接控制,但要使个体获取客观知识,并实现客观知识的主观化,却必须借助一定的组织手段。这就需要对客观知识加以组织,使其保持个体能够控制和使用的状态。因此,知识组织是个体实现客观知识主观化的必要社会条件。

(2) 控制论意义。知识对于人类及其社会发展所具有的重要意义,使得人类的知识创造活动连绵不断且加速进行,由此形成了庞大的客观知识系统。与此同时,不可避免地造成了客观知识系统内部的熵增长。知识熵的存在及其增大趋势,为人们有效利用客观知识造成了极大障碍。因此,对客观知识系统采取一系列的组织化措施,控制其熵增长,使其处于有序化状态,从而保证人类对客观知识的有效利用,成为知识组织活动

的根本任务。

(3) 动力学意义。知识运动的生产、交流和利用三个环节,都离不开一定的组织化手段作为其动力,而且各环节之间的衔接和循环运动,更离不开一定的组织化手段作保障。因此,知识组织是保证或加速知识运动顺利进行的必要动力。

第二节 知识组织中的知识

1.2.1 知识单元:知识组织的对象

知识组织中的“知识”不是指哲学意义上的“知识”,即“人类对社会实践经验的总结,是人的主观世界对客观世界的概括和如实反映”;也不是人工智能中所定义的“知识”,即“知识=客观事实+主观信念+主观信念与客观事实之间的一致性关系”,而是除去内涵的更加抽象化的知识,即着重从微观角度对知识进行认识。

在这里,可将知识(知识单元,即文献中的数据、公式、事实、结论等最小的独立单位)视为一种网状结构,这种特定意义上的知识单元是由众多结点(即知识因子)和结点间的联系(即知识关联)两个要素组成的。知识因子是组成知识单元的最细微的成分,一个概念、一种事物(如“企业”、“计算机”等)都是组成知识单元的一个知识因子,也就是说,知识单元由一个或多个知识因子组成。知识关联则为若干个知识因子之间建立起的联系,因为知识单元是有机联系的网状结构,而不是各个知识因子的散乱分布。知识关联在产生新知识、形成新文献中起重要作用,是使知识有序化的必要条件。由知识因子和知识关联的网状结构所表示的知识单元是知识组织的基本对象。针对知识单元的两个构成要素进行的序化,就是知识因子的有序化和知识关联的网络化。

在知识组织领域，知识有两种存在形态：一是寓于个人意识之中的主观存在，即存在于人的大脑之中的个人知识或主观知识，其载体是人的大脑细胞；二是依附和记录于各种载体（纸张、胶片、磁盘、光盘等）的客观形态，这种知识称为公共知识、社会知识或客观知识。

狭义的知识组织等同于文献的分类、标引、编目、文摘、索引等一系列整序。广义的知识组织则是针对知识的两个构成要素进行的，即知识因子的有序化和知识关联的网络化。

知识组织的任何方法都要建立在知识表示的基础上，知识表示是知识组织的基础和前提。知识表示是把知识客体表示出来，以便人们对知识的识别和理解。在知识表示过程中，人们已经建立了以文献单元为基础和以数据单元为基础的知识组织体系。目前正在研究以智能为基础的知识组织系统。不同的知识组织所体现的知识表示的方法是不同的。比如，对文献单元的知识表示，目前还停留在分类法和主题法阶段，两者对文献知识内容的表示不能反映文献中的具体知识单元。一旦实现了知识单元（包括知识因子和知识关联）的表示，将产生极大的知识增值，进而大大推进人类对知识的利用和创新。

国外有代表性的知识组织研究立足于从人类创造过程中利用知识的特点出发来组织知识，建立知识组织系统。比如，英国学者布鲁克斯（B. C. Brookes）所说的知识组织是对文献中所含的内容进行分析，找到人们创造与思考的相互影响及联系的结点，像地图一样把它们标示出来，以展示知识的有机结构，为人们直接提供创造所需要的知识和信息。其实质是期望找到知识生产创造过程中的关键数据（知识单元），然后用图来标示其联系与结构，实现知识的有序化组织。这样的系统虽然是高效的，但目前还不成熟，可行性也值得推敲。