本体与知识组织

胡兆芹◎著

BenTi Yu ZhiShi ZuZhi

中国文史出版社

本体与知识组织

胡兆芹◎著

BenTi Yu ZhiShi ZuZhi

中国文史出版社

图书在版编目（CIP）数据

本体与知识组织 / 胡兆芹著. —北京：中国文史出版社，2013.11

ISBN 978-7-5034-4411-1

Ⅰ.①本… Ⅱ.①胡… Ⅲ.①医学哲学—研究 Ⅳ.①R-02

中国版本图书馆 CIP 数据核字（2013）第 263941 号

责任编辑：李晓薇

出版发行：中国文史出版社

网　　址：www.wenshipress.com

社　　址：北京市西城区太平桥大街 23 号　邮编：100811

电　　话：010-66173572　66168268　66192736（发行部）

传　　真：010-66192703

印　　装：北京天正元印务有限公司

经　　销：全国新华书店

开　　本：170mm×240mm　1/16

印　　张：13.5

字　　数：202 千字

版　　次：2014 年 1 月北京第 1 版

印　　次：2014 年 1 月第 1 次印刷

定　　价：39.00 元

前　言

本体（ontology）最初是哲学领域的概念，是对世界任何领域内的真实存在所做出的客观描述。20世纪90年代以来，人们将本体的概念引入人工智能、知识工程和图书情报领域，在这些领域中，本体主要是一种知识组织体系，通过对领域中概念及概念间关系的描述来进行知识的表示和组织。近年来，关于本体的研究与应用发展迅速，但国内关于本体的研究无论是在理论研究、实证研究，还是在技术手段的实现和应用方面都相对落后，与国外研究相比存在很大差距。目前，国内图书情报领域关于本体的研究尚处于起步阶段，尚未见到有关国内构建具有推理功能的学科或领域本体系统的报道。

本书从知识组织的定义、内容和模式分析了知识组织的内涵；从本体的概念演变、组成要素、类型和作用等方面分析了本体的定义和其理论基础；从本体表示语言、本体构建工具、本体构建方法、本体的映射和演化，分析了本体应用所使用的关键技术；从基因本体、UMLS、TCMLS和其他应用分析了本体在医学领域中的应用；以医学知识组织为例，构建一个完整的知识本体框架—医学知识本体模型，探索并揭示构建医学领域本体及医学知识组织的步骤、框架和关键性问题。

本书是2012年湖北省教育厅人文社会科学研究项目“Ontology在医学学科知识组织中的应用研究”（编号：2012G272）及湖北省高校图工委

研究基金项目“基于本体的学科知识组织模式研究”（编号：2011YB10）研究成果，也是多年来作者研究本体理论及其应用的总结。由于时间仓促，水平有限，加上新技术的不断涌现，书中难免存在错误和不妥之处，欢迎广大同仁批评指正。

作者

2013 年 9 月

目　录

CONTENTS

图表目录

第一章

知识组织理论概述

第一节　知识组织的定义

知识是人类对客观世界的认知过程的产物。从根本上来说，知识是人类在认知客观世界的过程中所形成的关于客观世界及人类自身的主观印象，它超越了个人感性经验的限制而形成的一个关于认知对象理性的结构模型。知识是具有一定逻辑结构的复杂体系，能够物化在一定的符号体系中，从而可以通过学习和推理而获得，不一定都经实践加以证明。知识是人类特有的信息，是信息的一部分。后来，人们普遍认识到知识整序的必要性，于是出现了知识组织。

英国著名分类学家 H. B. Bliss 最早于 1929 年提出“知识组织”这个概念，他在《知识组织和科学系统》、《图书馆的知识组织》两部著作中，阐述了以文献分类为基础的知识组织思想。此后，在美国、澳大利亚等国家，通常把文献分类、标引、编目等课程称为“知识组织”课程。在我国，最早使用“知识组织”一词的是著名文献情报学家袁翰青教授，他于 1964 在一篇文章中指出：“文献工作是组织知识的工作……通常所谓文献工作实际上有两个方面：知识组织工作的一方面和情报检索工作的一方

面。”此后国内外不同领域的学者结合领域特点对“知识组织”进行了解释，而对知识组织进行专门的深入研究并开展各种学术活动是近 20 年的事情。到了 20 世纪 90 年代，知识组织这一概念在学者们的研究下逐渐清晰。国际知识组织学会德国分部“知识组织与网际网络”工作小组 Alexander Sigel 认为“知识组织是将含有知识的集合物加入信息价值的一种跨学科领域的文化活动，以便为用户群提供最好的相关信息体系”[1]；贾同兴认为，“所谓知识组织，是指对事物的本质及事物间的关系进行提示的有序结构，即知识的序化”[2]；王知津将其定义为“对知识进行整序和提供”[3]；蒋永福认为，“知识组织是指为促进或实现主观知识客观化和客观知识主观化而对知识客体所进行的诸如整理、加工、引导、揭示、控制等一系列组织化过程及其方法”[4]；储节旺等认为，“知识组织是按照知识的内在逻辑联系，运用一定的组织工具、方法和标准对知识对象进行诸如整理、加工、表示、控制等一系列的序化、系统化的活动”[5]。纵观知识组织理论研究的历程，不同学者对“知识组织”从不同角度进行了描述，尽管目前还没有统一权威的定义，但通过归纳、总结上述定义后认为，知识组织是按照知识的内在逻辑联系，运用一定的组织工具、方法和标准，通过整理、加工、表示、控制等一系列的序化、系统化的活动，揭示知识的本质，实现知识的关联的过程和方法，主要包括知识的重组和优化，知识结构、关系和语义的描述，知识的提取、挖掘和智能化表述。知识组织的本质特征即用一定的方法和手段对知识的各种要素加以组织，以便知识传播、提供和利用。同时，知识组织概念随着技术进步不断发展完善，呈现与时俱进的特点。知识组织的目标是对知识存贮进行整序和提供知识，此外，知识组织也与未来关系密切，它必须有效地制止已经暴露出来的错误倾向，以便防止无序化再次发生。这意味着，知识组织不但要处理大量的现有知识，而且还要对减少知识的增长有所作为。

在国际专业组织方面，首次正式使用“知识组织”一词的是德国法兰克福国际知识组织学会。国际知识组织学会是国际上主要的研究分类法

和主题法的学术团体，它的全称是 International Society for Knowledge Organization（ISKO），成立于 1989 年 7 月，其宗旨是推动对知识概念组织的各种方法的研究、发展和利用。这里所说的“知识组织”主要是指人类记录下来的知识的组织，包括文献分类、主题标引及知识表示（含相关的语言学问题和术语学等），所以知识组织是在分类法的基础上发展起来的。在第二届国际 ISKO 大会上，提出了将国际性学术刊物《国际分类法》改名的建议，该建议得到 ISKO 执委会的赞同，委托 ISKO 科学咨询委员会提出 4 种方案，最终确定使用《知识组织》（Knowledge Organization），从 1993 年的第 20 卷起正式更名。自此知识组织的概念在图书情报界广泛传播开来。从分类法到知识组织是个飞跃，这表明，人们的研究领域已经突破了分类法的原有范围，延伸到知识组织所能概括的所有问题。

由上可知，知识组织一词首先是在图书情报学领域提出和使用的，而且其意义是指文献分类、编目等文献组织活动的统称或概括。事实上，这种知识组织概念，只能说是狭义的知识组织，因为文献组织活动只是一种客观知识的组织活动，而知识组织还应包括主观知识的组织。广义的知识组织应包括主观知识组织和客观知识组织两方面。所以，知识组织是指对知识客体所进行的诸如搜集、整理、加工、整序、揭示、控制、提供等一系列组织化过程及其方法，它包括主观知识（隐性知识）的组织和客观知识（显性知识）的组织两方面。对隐性知识的组织表现为知识的自组织过程，对显性知识的组织表现为外在的、社会的控制与组织过程。

第二节 知识组织的内容、实质和原理

虽然图书情报学界是最早开始研究知识组织问题的，但一直局限在文献的分类、标引、编目的范畴之内，这种以文献知识为对象的知识组织理论，我们可称之为狭义的知识组织理论。而从一般意义上直接探索知识组

织的理论与方法，即广义的知识组织理论研究是从1956年人工智能（artificial intelligence）或称知识工程学（knowledge engineering）创立之后开始的。

知识组织的理论基础是[6]：对知识进行的任何组织都必须建立在知识单元的基础上，而知识单元无非就是概念。知识不能靠自己组织和表示，除非用知识单元及许多词语或句子的可能组合来表示。知识组织所研究的最小单元是概念及其词语表达。知识组织的任务是寻求抑制知识存取无序化的方法，其目标是使知识（资源）处于有序化状态，并提供有序知识，保证客观知识主观化过程的顺利进行。

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，满足人类的客观知识主观化的需要是知识组织活动的最终目的。因此可以说，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。

如何组织客观知识，这就需要首先了解和掌握客观知识的结构特征。知识的结构呈现为一种网状结构，它由众多结点（知识因子）和结点联系（知识关联）这两个要素构成。知识因子是组成知识的基本单位，一个概念、一个语词、一种事物都可成为知识因子，知识关联是若干个知识因子间建立起来的特定联系。从知识的这种结构特征看，知识组织的基本原理就是用一定的方法把知识客体中的知识因子和知识关联揭示出来，以便于人们认识、理解和接受。客观知识都是用一定形式的语言表示出来的，语言是知识的直接承担者。这样，知识的内部结构直接表现于语言的外部结构之中。所以，人们揭示知识的内部结构，自然采取了通过语言结构去“探测”的方法。知识的语言结构一般表现为语法、语义和语用三方面，所以知识组织的原理也可从以下三方面分别考察[7]：

①知识重组（语法学原理）。它是知识组织的初级但很重要的方法，

是对知识对象内的知识因子和知识联系进行语法结构上的重新整合，结果生产出新的知识产品，它包括知识因子的重组和知识联系重组。

②知识表示（语义学原理）。它是将知识对象中的知识因子和知识联系表示出来，便于人们识别和理解。知识表示是知识重组的前提，包括采用分类标引表示法和主题标引表示法表示的知识因子表示法、一阶谓词逻辑表示法、产生式规则表示法、框架式表示法、语义网络表示法表示的知识联系表示法。

③知识记忆（语用学原理）。知识组织是以实现客观知识的主观化为目的的。客观知识的主观化实际上是个体的知识记忆过程。从语言学角度看，知识的语法规则和语义表示都要为知识的语用服务。这就说明，知识组织最终要为个体的知识记忆服务。

第三节 知识组织模式

目前来看，分类法、主题法和元数据法是传统的对文献或信息的内容属性进行组织、有序化的主要模式。

一、传统的知识组织模式——分类法、主题法

分类法和主题法是传统最重要的知识组织模式，是数代图书情报人员智慧和经验的积累，它们的知识组织能力在一百多年的发展和应用过程中得到了充分证明和不断的丰富。

1. 分类法

分类是人类的思维方式，是从本质上提示和把握事物之间的区别与联系的重要手段。分类法将表示各种知识领域（学科及其研究问题）的类目按知识分类原理进行系统排列并以代表类目的数学、字母符号（分类号）作为文献主题标识的一类情报检查语言。知识经过分类组织后，就

能揭示知识的全貌及其内在联系，提供分门别类查询知识的途径。由于信息被分别组织在不同的类中，起到了过滤和筛选作用。它的主要特点是按学科、专业集中文献，并从知识分类角度揭示各种文献在内容上的区别和联系，提供从知识分类检索文献的途径。将事物和学科概念纳入知识分类体系，是对知识进行系统组织的合理的方法。

国内的图书情报部门主要采用《中国图书馆分类法》、《中国科学院图书馆图书分类法》和《中国人民大学图书馆图书分类法》。这三部分类法在传统图书馆中的文献分类、建立分类著录卡片、建立分类索引工具方面一直发挥重要作用。这三种分类法都属于线性分类法，应用方面主要体现在分类排架和目录组织，即可以按学科门类和知识体系进行浏览和检索。分类法不仅在文本秩序阶段发挥了重大的作用，传统图书馆中专业馆员对其极为熟悉并乐于使用，而且在数字环境下，分类法为网络信息组织和访问提供了一种解决办法，表现在类目按一定的主题进行组织，并辅之以年代、地区等分类形成分类主题树状结构目录。使用分类法来组织知识具有以下几个优点[8]：一是主题分类列表可以作为一种导航工具，帮助用户通过浏览查找所需要的信息资源；二是分类表是等级式的，因而易于扩展和缩小检索范围，能够提高查全率和查准率；三是如果一个网站所使用的是某部比较通用的分类表，那么它能够比较容易对其他使用了相同分类表的网站实现跨数据库浏览和主题检索；四是类目与信息记录之间通过超文本技术直接连接，可以加强交替类目、参见与注释类目之间的横向联系，加强多重列类的纵向联系，可以揭示知识空间的多维联系。

目前在组织知识方面大致有以下分类组织方法：

（1）主题分类法

主题分类法是依事物分类，而不是依学科分类，其结构是以一个主题充当一个类目，类目像主题词表一样按字顺排列，而不是根据逻辑顺序排列，这样，能够将相关的资源集中在一起。它的优势在于适用于交叉学科的主题；不足是容量太小，对学科知识的覆盖率极为有限。