

蒋永福 李景正

论知识组织方法

摘要 知识组织有7种主要方法:知识表示、知识重组、知识聚类、知识存检、知识编辑、知识布局 and 知识监控。参考文献15。

关键词 知识组织 方法 知识聚类 知识存检

分类号 G302

ABSTRACT There are seven methods for knowledge organization: demonstration, reorganization, agglomeration, deposition and retrieval, editing, distribution and supervision. 15 refs.

KEY WORDS Knowledge organization. Method. Knowledge agglomeration. Knowledge retrieval.

CLASS NUMBER G302

知识组织方法多种多样:依知识的内部结构特征,可分为知识因子的组织方法和知识关联的组织方法^[1];依知识的不同存在形态,可分为主观知识的组织方法和客观知识的组织方法;依知识组织的语言学原理,可分为语法组织方法、语义组织方法和语用组织方法^[2,3],等等。具体的知识组织方法,归纳起来主要有7种:知识表示、知识重组、知识聚类、知识存检、知识编辑、知识布局 and 知识监控。

1 知识表示

所谓知识表示,是指把知识客体中的知识因子和知识关联表示出来,以便人们识别和理解知识。知识表示是知识组织的基础与前提,因为任何知识组织方法都要建立在知识表示的基础上。有主观知识的表示和客观知识的表示之分。

1.1 主观知识的表示

主观知识存贮于人脑中,对它的表示表现为复杂的人脑神经生理与心理过程。目前的科学发展尚未完全探明人脑主观知识表示的内在机制。但是,在人工智能的专家系统研究领域,对人脑的知识表示机制进行模拟研究,取得了可喜成果。专家系统的核心是知识库系统,知识库中的知识存贮方式及其推理输出规则,即为专家系统的知识表示方法。专家系统对专

家知识的表示主要采取以下4种方式^[4,5]:

1.1.1 逻辑(logic)表示法

这种方法运用命题演算、谓词演算等逻辑手段来描述一些事实的性质、状况、关系等知识。它利用命题逻辑中的联结词符号建立演绎逻辑系统,可进行事实推理、定理证明等运算。

1.1.2 产生式规则(Production Rules)表示法

这是一种前因后果式的知识表示模型,它由两部分构成:前一部分称为条件,用来表示状况、前提、原因等;后一部分称为结果,用来表示结论、后果等。其规则是:“IF(前件) THEN(后果)”,其意义是:如果(IF)前件满足,则(THEN)系统执行动作(或得出结论)。在一个专家系统中,专家知识、用户知识和背景知识,一般用产生式规则表示。如在文献检索专家系统中,用“IF...THEN...”规则能很方便地表达诸如索引规则、聚类规则、检索反馈策略等专家经验和思想。

1.1.3 语义网络(Semantic Network)表示法

知识的语义网络表现为某一领域知识概念之间关系的网式图。它由节点和弧构成:节点表示知识的基本概念(知识因子),弧表示节点间的联系(知识关联)。语义网络表示法能够把知识因子和知识关联同时生成和表示,并以图的形式直观地显示出来。这种表示方法符合人类联想记忆的思维模式,因此在专家

系统建设中得到广泛应用。在文献标引实践中,用语义网络来表示词表知识是一件比较容易的事情。

1.1.4 框架(Frame)表示法

它的基本思想是根据人们以往的经验 and 背景知识,来推理当前事物的相关知识。一个框架由多个槽(Slot)组成,每个槽又由一个或多个侧面(Facet)描述,若干个槽共同描述框架所代表的事物的属性及其各方面表现。框架表示法能够深入全面地揭示事物的内部属性,适用于知识的深层表达。

1.2 客观知识的表示

客观知识存在于各种类型的文献之中,具有确定的知识因子和知识关联结构。客观知识表示的任务就是把文献中的知识因子和知识关联用一定方式表示出来即可。对文献知识的表示,目前普遍采用分类标引法和主题标引法。这两种方法都属于揭示文献主题内容的方法,两者的基本原理相同:先编制标引用词典(或称标引语言),然后把文献知识特征(形式特征与内容特征)与词典中的标引词汇之间进行相符性比较,最后把相符的词汇用其代号(分类号或主题词)表示出来。这个代号为文献标识符。但两者的知识组织体例不同。分类标引法是语法组织和语义组织的综合,基本上属于族性组织体例。主题标引法是以语法组织为主、语义组织为辅的综合组织,基本上属于特性组织体例,其中词族索引和范畴索引由于展现了主题词之间的等级关系和学科关系,因而基本属于语义组织体例,而附表和语种对照索引则属于语法组织体例。

2 知识重组

知识重组是对相关知识客体中的知识因子和知识关联进行结构上的重新组合,形成另一种形式的知识产品的过程。知识重组的目的是,通过对知识客体结构的重新组合,为用户克服因知识分散而造成的检索困难提供索引指南;为人们提供经过加工整序后的精炼性知识情报;为用户便于理解和吸收知识,提供评价性或解释性知识。它又包括知识因子的重组和知识关联的重组。

2.1 知识因子的重组

它是指将知识客体中的知识因子抽出,并对其进行形式上的归纳、选择、整理或排列,从而形成知识客体的检索指南系统的过程。这一重组过程,实际上是对知识因子在结构上的整序或浓缩的过程。在这个过程中,知识因子间的关联并未改变,没有产生新知

识^[6]。在文献情报工作中,往往利用知识因子的重组手段,形成文献知识的索引系统,例如,主题索引系统和分类索引系统的形成。它们的产品形式就是各种类型的二次文献,包括目录、索引、文摘、题录、书目之书目、文献指南等。

2.2 知识关联的重组

它是指在相关知识领域中提取大量知识因子,并对其进行分析与综合,形成新的知识关联,从而生产出更高层次上的综合的知识产品的过程。由于改变了知识因子间的原有联系,所以其结果可以提供新知识,也可以提供关于原知识的评价性或解释性知识。它所形成的产品主要是各类三次文献,如综述、述评、词典、手册、年鉴、类书、百科全书、专题讲座等。

无论是知识因子的重组还是知识关联的重组,都要遵循客观性原则,即都不能改变原知识客体的语义内容。由此可以看出,知识重组基本上属于语法组织的范畴。从 N. Chomsky 的生成语言学(Generative Linguistics)原理看,知识客体内容的不改变指的是知识语言的深层结构(Deep Structure)的不改变,而知识的语法重组实际上是形成或求得基于深层结构的各种表层结构(Surface Structure)形式的过程^[7],如图 1 所示。

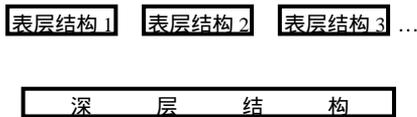


图 1 知识语法重组过程

3 知识聚类

它也可称为知识分类组织法。聚类和分类是一个过程的两个方面:分类的结果产生了聚类,聚类的结果产生了分类。知识聚类组织法,是指将知识按一定的聚类标准分门别类地加以类集和序化的过程。它的基本原理是“事以类聚”,即根据事物的不同属性,将属性相同或相近的事物集中在一起,将属性不同的事物区别开来。按所采用的聚类标准的不同,它可分为 5 种类型。

3.1 以学科聚类

这是以学科分野作为聚类标准,将相同学科的知识类集在一起。这是古今中外各种文献分类法普遍采用的知识聚类方法。它的基本原理是:以学科分野作为区分知识集合的标准,以类目——具有相同属性知识的集合性概念——表征知识因子,以类目的等级体系作为表征知识关联的主要方法。

3.2 以主题概念聚类

它是知识客体的主题概念为聚类标准,将相同或相近主题概念的知识聚集在一起。它为各种类型的主题词法所采用。基本原理是:以主题概念作为类集知识的标准,用主题词表征知识因子,用参照系统显示知识关联(主题词之间的关系)。

3.3 以人聚类

这是以人物作为标准来类集知识。在文献分类法和主题法中,经常采用以人聚类的组织方法,如在《中图法》(第4版)中,“A 马列主义、毛泽东思想、邓小平理论”、“I210 鲁迅著作及其研究”、“K81 传记”等,就采用了这种方法。在词典、手册、百科全书以及地方志的编纂过程中,也经常以人设条。

3.4 以用聚类

这是以知识本身的用途或效用为标准来类集知识。例如,在文献知识的分类组织中有一个实际效用原则,此原则要求“在文献归类时要考虑本单位的专业性质和本单位专业读者的需要,把文献归入到最大用途的类目中去”^[8],这一原则就体现了以用聚类的思想。

3.5 以时空聚类

知识具有时空结构。知识的时空聚类就是根据知识产生的时间和空间属性为标准来类集知识。文献知识的分类组织中所采用的“时代复分”和“地区复分”,就体现了以时空聚类知识的原则。在知识创新研究中,如果正确使用知识的时空聚类方法,将会产生显著效果,甚至可能取得意想不到的创造性成果,这是因为知识的时空聚类不仅仅是一种“归类”,更重要的是一种“整合”。在知识创新研究中,可采用如下4种知识时空聚类方法^[9]:

(1) 将时空相近的知识整合在一起。不仅可以实现知识的归类,可以了解某一学科的发展现状,还可以进行某一学科研究成果的横向比较研究。

(2) 将时间上相近、空间跨度较大的知识整合在一起。这可以实现某阶段不同知识成果间的交叉渗透,有利于新知识的产生。知识空间的跨度越大,其知识单元的差异就越大,各知识单元间的优势互补就越明显,其价值也就越大。

(3) 将时间跨度较大、空间上相近的知识整合在一起。这可以看出某一学科、某一领域的发展脉络,预测其发展趋势,也可以找出某一事物发展的主导因素和其中的关键问题。

(4) 将时间跨度和空间跨度都较大的知识整合在

一起。这可以极大地发挥各知识单元间的“杂交”优势,可以使一些老的学科焕发青春,又促进新学科的产生。许多新兴的交叉学科的诞生,就是通过这一途径实现的。

4 知识存检

知识存检是由“存贮”和“检索”两方面构成的系统或过程。这里所说的“检索”,是指检索系统的建立过程,而不是用户的实际检索过程,因为用户的实际检索过程处于知识的查找和利用环节,而不处于知识组织的环节。知识因能够存贮而得以积累和延传,因能够检索而得以吸收利用。建立科学有序的知识存检系统,是知识组织活动的重要任务。知识存检可分为脑内存检和脑外存检。

4.1 知识的脑内存检

知识的脑内存检表现为复杂的大脑神经生理与心理过程。目前认知心理学对大脑知识存检机制的研究,可谓富有成果。在认知心理学研究中,知识的存检被当作“记忆”过程来对待。一个完整的知识记忆过程包括知识的存贮、编码和提取(回忆)3个环节。那么知识是如何被记忆的呢?对此,认知心理学家阿特金森(R. C. Atkinson)和希夫林(R. M. Shiffrin)提出了由感觉记忆、短时记忆和长时记忆构成的三级记忆系统理论:感觉记忆是指从感官刺激接收的知识信息以感觉形式保持瞬间的记忆形式,其中被“保留”的部分进入短时记忆系统;短时记忆是指从感觉记忆中转移过来的知识信息在头脑中的短暂存贮,其存贮容量不多,约为 7 ± 2 个记忆单位,其保持时间也不长,约1分钟,其记忆内容通过复述进入长时记忆系统;长时记忆是指关于外界知识的长久存储,它的容量可谓无限多,它的保持时间从几分钟、几小时、几天、几年到一个人的终生^[10]。

知识的脑内存检(记忆)过程,实际上就是主观知识的形成过程。尽管人们目前还不能完全探知其内在机制,但它是一切脑外知识存检方法均应遵循的基本依据,因为一切脑外知识存检方法的根本目的就在于为个体的知识记忆服务^[11]。

4.2 知识的脑外存检

这又分个体的脑外存检和公共的脑外存检。前者表现为个体的随机行为过程,难以把握和描述其普遍规律,因此本文只论述知识的脑外公共存检。知识的公共存检属于客观知识的存检范畴,亦即属于文献存检范畴。社会的文献情报部门就是从文献的公

共存检工作的专门机构。

文献情报部门对文献的存贮,从静态角度看只是表现为文献的存储与积累,但从动态角度看却是一种知识的历时纵向传递过程,其实质是知识的社会公共记忆。人类的知识就是通过这种公共记忆系统而得以世代传承和发展。Crawford 和 Gorman 就说过^[12]:“图书馆采集、检索和维护所有形式的知识和信息载体,并为用户检索馆藏提供说明和辅佐,简言之,图书馆的存在为在知识进步和文化保存中超越时空而进行的不懈努力赋予了深远的意义。”从全社会角度看,图书馆作为知识的公共记忆系统,汇集了大量客观知识,使知识得到了有序化存贮与积累,避免了知识存在的散乱状态,这本身就是知识组织的一种体现^[13]。文献情报部门通过文献揭示了编排的多种方法,建立起一整套文献存检方法体系,已成为知识公共存检的主体系统。目前迅速发展的因特网,其实也是一种知识信息的脑外存检系统,只不过它以虚拟形式存在。因特网对知识信息的存检方法,从基本原理上看也与一般的文献存检方法大体相似,只是它的网络化、自动化程度远比一般的文献存检系统高。

5 知识编辑

本文的“知识编辑”,是指对知识客体进行的搜集、整理、加工制作等编辑活动。知识编辑的过程,一般表现为先“辑”后“编”^[14],即先搜集相关资料,然后加工制作成特定形式的知识产品。

从知识编辑的功能上看,它是知识产品能够汇入到知识流海洋之中的一道道关闸,在这里,知识编辑起到了“编辑筛”的作用,即符合编辑标准(包括内容标准和形式标准)的知识产品得到生产和流通,而不符合编辑标准的知识产品则被筛掉。某一知识产品在未经编辑前被称为初级产品(或称编前产品),经过编辑加工后的产品则被称为成样产品(或称编后产品)。在这里不能称为终级产品,因为该产品可能以后继续被编辑)。某一知识产品,经过“编辑筛”的筛选,成为成样产品,这就是所有编辑活动的微观过程。从宏观上看,知识编辑的筛选功能,可以实现知识生产的“优生”,从而保证知识生态的优化组合。可见,知识编辑是关系到知识生态环境优劣的知识组织方法。因特网上的知识信息之所以庞杂混乱,主要是由于许多知识信息未经组织编辑而直接上网所导致。

知识编辑和知识重组具有相似性,两者之间有时相互包容、相互交叉,例如,大部分二、三次文献的形

成过程,就是通过知识编辑活动来完成的。正因如此,有些知识组织活动既可以说成是知识重组,也可以说成是知识编辑。尽管如此,知识编辑和知识重组之间还是有比较明显的区别:知识编辑活动往往是由比较固定的编辑组织,且由相对职业化的编辑人员来完成。而在知识重组活动中一般不需设有固定的编辑组织,即使有时设有编辑组织,也是临时性的;有些知识编辑活动就不宜或不能说成是知识重组活动,如丛书编辑、译文编辑等。

本文根据编辑体例和编辑产品的不同,将知识编辑分为 7 种类型:辑录性编辑,如类书的编辑;撰写性编辑,如辞书、百科全书的编辑;汇集性编辑,如丛书、文集的编辑;审读性编辑,如专著、论文的编辑;译介性编辑,如翻译作品的编辑;评介性编辑,如综述、书评的编辑;专辑性编辑,如专题知识讲座的编辑。

6 知识布局

知识布局是一种宏观的知识组织方式,它是指对社会上的知识资源进行调配和布局,以实现知识资源的合理配置,满足社会、经济、文化发展的需要。根据知识载体的不同,知识布局可分为主观知识的布局和客观知识的布局。

6.1 主观知识布局

主观知识的布局,主要是通过对主观知识的拥有者——人的配置活动来实现。主要有自然性布局和政策性布局两种形式。科教文卫部门和政府机关聚集大量知识分子(即拥有大量主观知识资源),这是社会发展需要所造成的自然性布局。发达国家和发达地区聚集大量优秀知识分子,也是一种自然性布局。而向落后部门或落后地区支援、调配所需人才,则属于政策性布局。

6.2 客观知识的布局

它基本上属于文献资源布局的范畴。所谓文献资源布局,就是从宏观上制定目标和规划,进行协调和分工,以指导文献情报部门的文献收集工作,突出各自优势,形成比较完备的收藏,并将其作为社会的知识资源,共享共用,从而建立文献资源保障体系^[15]。文献资源的布局分为自然性布局和政策性布局。社会上的文献情报部门收集有大量文献资源,发达国家和地区拥有世界上大部分文献资源,这是一种自然性布局状况。而诸如“知识工程”、“捐书工程”、“送书下乡”、“全国图书协调方案”等,就属于政策性布局措施。

7 知识监控

这里所说的知识监控,主要是指政策性监控,它是指政策主体按照自己的意愿和利益,制定相关的政策法规,对知识主体(包括生产主体、管理主体和利用主体)的行为活动加以限定和监督。知识监控是一种知识的外在组织方法,其目的是为了完善知识系统的内在秩序。知识监控可分3种类型。

7.1 思想监控

这是指政策主体对知识主体及其知识产品的政治倾向性所实施的限定与监督。其主要对象是人文社会科学领域。对知识主体及其知识产品实施思想监控,是阶级社会的一个特点,只要阶级存在,这种监控就不可避免,不管其历史后果是正确的还是错误的^[16]。

7.2 法律监控

就是指法律主体为了保证其意志与利益的实现,为了保障知识主体的合法权益,为了保证社会的知识管理活动正常秩序,而对知识主体和知识管理活动加以法律限定与监督。知识的法律监控,主要表现为相关法律法规,如知识产权法、专利法、图书馆法、档案法、信息保密法等制定与实施。知识的法律监控具有强制性与统一性特点。

7.3 标准监控

是为了保证知识组织活动的标准化和规范化而实施的一种监控。它主要表现为一系列相关标准、规程的制定与实施过程,如文献著录标准、文献分类与标引规则、数据库数据格式标准、参考文献著录规则、期刊编辑规则、文献情报工作评估标准、索引款目格式及其编排规则、索引语言词典等的制定与实施。知识标准监控具有统一性、规范性和操作性特点。

参考文献

- 1,6 王知津. 文献演化及其级别划分——从知识组织角度进行探讨. 图书情报工作, 1998(1)
- 2 霍国庆. 论信息组织. 情报资料工作, 1997(6)
- 3 蒋永福. 论知识组织. 图书情报工作, 2000(6)
- 4 杨列勋. 情报检索专家系统的知识表达. 情报科学, 1991, 12(2)
- 5 杨学山. 专家系统及其在管理中的应用. 北京:清华大学出版社, 1991
- 7 [美]J 格林著;方立,张景智译. 乔姆斯基. 北京:中国社会科学出版社, 1990
- 8 刘延章. 文献信息分类学. 北京:中国科学技术出版社, 1996
- 9 邱均平,段宇锋. 论知识管理与知识创新. 中国图书馆学报, 1999, 25(3)
- 10 周谦. 学习心理学. 北京:科学出版社, 1992
- 11 蒋永福. 图书馆与人类的记忆——从知识记忆角度理解的图书馆学. 中国图书馆学报, 2000, 26(2)
- 12 刘嘉. 走向消亡还是长足发展:论网络环境下的图书馆. 图书情报工作, 2000(5)
- 13 蒋永福. 图书馆与知识组织——从知识组织角度理解的图书馆学. 中国图书馆学报, 1999, 25(5)
- 14 [台湾]张觉明. 现代杂志编辑学. 北京:中国书籍出版社, 1987
- 15 杨沛超等. 从藏书建设到文献资源建设. 情报资料工作, 1991(5):5~8
- 16 蒋永福. 文献信息控制方法浅论. 图书馆建设, 1992(3)

注 本文系黑龙江省教委资助项目“支撑知识创新体系的知识组织系统”的研究成果之一。

蒋永福 黑龙江省牡丹江市图书馆副馆长、副研究馆员。通讯地址:牡丹江市。邮编 157000。

李景正 黑龙江大学信息管理系主任、教授。通讯地址:哈尔滨市。邮编 150080。

(来稿时间:2000-07-03)

本刊 迁址

中国图书馆学报编辑部已迁至国家图书馆东 618 室办公。
通讯地址:北京市海淀区中关村南大街 33 号中国图书馆学报编辑部。
邮编:100081。电话:010-68415566-5141。传真:010-68417815。