



知识集成方法与技术

—— 知识组织与知识检索

Knowledge Integration Methods and Technologies
—— Knowledge Organization and Knowledge Retrieval

王兰成 著



国防工业出版社
National Defense Industry Press

国家社会科学基金项目(05)

知识集成方法与技术

——知识组织与知识检索

Knowledge Integration Methods and Technologies

——Knowledge Organization and Knowledge Retrieval

王兰成 著

国防工业出版社

·北京·

图书在版编目(CIP)数据

知识集成方法与技术:知识组织与知识检索 / 王兰成著. —北京:国防工业出版社,2010.12

ISBN 978-7-118-06933-4

I. ①知... II. ①王... III. ①信息管理 - 研究②情报检索 - 研究 IV. ①G203②G354

中国版本图书馆 CIP 数据核字(2010)第 182421 号

※

国防工业出版社出版发行

(北京市海淀区紫竹院南路 23 号 邮政编码 100048)

天利华印刷装订有限公司印刷

新华书店经售

*

开本 850 × 1168 1/32 印张 11½ 字数 289 千字

2010 年 12 月第 1 版第 1 次印刷 印数 1—3500 册 定价 32.00 元

(本书如有印装错误,我社负责调换)

国防书店:(010)68428422

发行邮购:(010)68414474

发行传真:(010)68411535

发行业务:(010)68472764

致 读 者

本书由国防科技图书出版基金资助出版。

国防科技图书出版工作是国防科技事业的一个重要方面。优秀的国防科技图书既是国防科技成果的一部分,又是国防科技水平的重要标志。为了促进国防科技和武器装备建设事业的发展,加强社会主义物质文明和精神文明建设,培养优秀科技人才,确保国防科技优秀图书的出版,原国防科工委于1988年初决定每年拨出专款,设立国防科技图书出版基金,成立评审委员会,扶持、审定出版国防科技优秀图书。

国防科技图书出版基金资助的对象是:

1. 在国防科学技术领域中,学术水平高,内容有创见,在学科上居领先地位的基础科学理论图书;在工程技术理论方面有突破的应用科学专著。

2. 学术思想新颖,内容具体、实用,对国防科技和武器装备发展具有较大推动作用的专著;密切结合国防现代化和武器装备现代化需要的高新技术内容的专著。

3. 有重要发展前景和有重大开拓使用价值,密切结合国防现代化和武器装备现代化需要的新工艺、新材料内容的专著。

4. 填补目前我国科技领域空白并具有军事应用前景的薄弱学科和边缘学科的科技图书。

国防科技图书出版基金评审委员会在总装备部的领导下开展工作,负责掌握出版基金的使用方向,评审受理的图书选题,决定资助的图书选题和资助金额,以及决定中断或取消资助等。经评审给予资助的图书,由总装备部国防工业出版社列出出版。

国防科技事业已经取得了举世瞩目的成就。国防科技图书承担着记载和弘扬这些成就,积累和传播科技知识的使命。在改革

开放的新形势下,原国防科工委率先设立出版基金,扶持出版科技图书,这是一项具有深远意义的创举。此举势必促使国防科技图书的出版随着国防科技事业的发展更加兴旺。

设立出版基金是一件新生事物,是对出版工作的一项改革。因而,评审工作需要不断地摸索、认真地总结和及时地改进,这样,才能使有限的基金发挥出巨大的效能。评审工作更需要国防科技和武器装备建设战线广大科技工作者、专家、教授,以及社会各界朋友的热情支持。

让我们携起手来,为祖国昌盛、科技腾飞、出版繁荣而共同奋斗!

**国防科技图书出版基金
评审委员会**

国防科技图书出版基金 第六届评审委员会组成人员

主任委员	刘成海			
副主任委员	宋家树	蔡 锺	程洪彬	
秘书长	程洪彬			
副秘书长	邢海鹰	贺 明		
委 员	于景元	才鸿年	马伟明	王小谟
(按姓氏笔画排序)	甘茂治	甘晓华	卢秉恒	邬江兴
	刘世参	芮筱亭	李言荣	李德仁
	李德毅	杨 伟	肖志力	吴有生
	吴宏鑫	何新贵	张信威	陈良惠
	陈冀胜	周一宇	赵万生	赵凤起
	崔尔杰	韩祖南	傅惠民	魏炳波

前 言

本书研究的学科数字信息群,是指数字信息环境中的以信息学科文献类型划分的多种数字信息对象集合。这里的学科文献类型是指图书馆学、情报学和档案学等信息学科的信息表现形态,数字信息对象即数字化的图书、报刊、论文、资料和档案等的描述性信息资源。

学科数字信息群与网络信息资源相比较,既有相同或相似的一些特点,又有明显的差异。它们共同的特点是:数量巨大,内容丰富,形式多样,更新频繁。数量巨大表现为网络上各种信息资源集为一体,并共享使用;内容丰富表现为信息资源覆盖各个学科或学科各个领域,来源广泛;形式多样表现为信息资源既有原始论文或电子报刊等一次文献,又有题录、摘评等二次、三次文献,还有文本、图形、声音等多种形式;更新频繁表现为网络信息量及发布点迅速增长,有用或无用的信息新陈代谢加快。学科数字信息群又是特定的网络信息资源,因此它与通常的网络信息资源相比有以下优势:虽然数量巨大,信息质量仍可控制,面向学科的数字信息群具有一定的标准和规范,许多信息经过必要的信息过滤、整理加工、质量控制和管理机制,已经处于有序状态;虽然形式多样,信息仍可有效组织,特别是学科数字信息群按学科的信息类型将网络信息资源进行划分,通过文献信息资源与网络信息资源的有机结合丰富了信息资源的体系,从而为数字信息群知识集成运用学科的理论和方法作了必要的准备。

基于知识集成方法与技术研究的理论与实践有其现实背景。知识的集成化处理是构成集成知识的前提,也是构成集成化智慧

的基础。学科数字信息群从信息集成发展到知识集成,不仅是研究异构数据的集成问题,也不仅是研究各类文献的信息集成问题,更重要的是要研究知识集成的技术与方法。数字信息群的知识共享是学科各领域协调发展和跨领域知识集成的重要基础,实现学科数字信息群的知识共享、集成和挖掘需要创新研究及应用相关的方法与技术,国际上已具备较成熟的学科知识集成条件,目前处于研究起步阶段,但在我国缺少统一的学科理论和技术体系支持,该研究基本属于空白。

本书在综述和分析国内外关于知识集成研究的大量成果基础上,从实现学科数字信息群知识集成的方法与技术两个方面提出研究的主要内容及其关键部分。主要的研究对象,一是对知识集成中知识组织的方法与技术进行研究,二是对知识集成中知识检索的方法与技术进行研究。但这种分别研究的结果并没有影响知识集成的方法与技术的整体研究。知识组织是传统的主题分析的继续和发展,是图书馆学、情报学和档案学中的分类法、标题表、叙词表等标引工具的延伸,也是信息机构服务、信息服务和知识服务自始至终的主题内容。知识组织体系是实现知识检索的前提和基础,知识检索则是基于知识组织体系的结果,又是对资源对象进行基于元数据的语义标注,而元数据是知识组织体系的语义基础,只有经过元数据描述与标注的资源才具有长期利用的价值。学科数字信息群的知识组织和知识检索的方法、技术之间的关系是相辅相成的,没有好的知识组织方法与技术,就不可能实现高效的知识检索,只有好的知识检索方法与技术,才能真正表达知识集成的结果。

全书共分9章。第一章是本书的概论;第二章对知识集成的实现基础进行论述;第三章对知识表示和本体管理进行论述;第四章研究知识集成中的元数据及其操作;第五章研究基于本体的知识集成框架和方法;第六章对知识集成中的知识组织问题进行具体研究;第七章在知识集成框架的知识组织基础上对知识集成中的知识检索问题进行系统研究;第八章主要对知识集成应用原型

系统的构建进行研究;第九章是对全文的总结和对未来工作的规划。在知识组织方法和技术的研究方面,相关课题研究所取得的主要成果:第一,构建了面向学科数字信息群的知识组织框架;第二,建立了学科数字信息群知识组织的元数据理论;第三,探索了学科数字信息群知识组织框架的实现方案。在知识检索方法和技术的研究方面,相关课题研究所取得的主要成果:第一,构建了基于本体的知识集成框架;第二,探索了应用于学科数字信息群的知识表示及推理方法;第三,建立了面向学科数字信息群的知识检索模型及知识集成原型系统。

在本书的写作过程中,得到来自多方的帮助和支持。首先要特别感谢国家社会科学规划办对我们的信任和支持;感谢南京政治学院上海分院为课题研究所提供的优越环境;感谢本书参阅和引用的参考文献的作者们,是他们的成果给了我们许多启迪;感谢课题组所有的成员为申报和完成本课题所作出的贡献。尤其是感谢我的两位博士生敖毅和曾琼,他们在知识的组织和知识的检索研究方面付出了辛勤的劳动,敖毅博士对知识组织的实现技术颇有研究,曾琼博士对知识检索的实现技术也颇有研究,按计划两位博士研制的原型系统为课题的顺利完成起了重要的作用。在我反复修改和调整他们的博士论文内容的同时,逐步形成了课题最终成果的框架和主体内容。在课题的评审过程中,一些专家在肯定成果的同时也提出好的建议,随后我进一步对本书的部分内容作了完善和调整。可以这样说,本专著的许多内容是在知识组织技术和知识检索技术的研究成果上经过有机结合后的再一次“知识集成”。

研究课题已经在2009年6月顺利结项,由衷感谢众多的专家同行们的热情鼓励。就在我准备出版这本拙著时,获悉得到了国防科技图书出版基金的支持,特别要感谢国防工业出版社和陈洁主任,感谢国防科技图书出版基金评审委的各位专家。有专家建议在知识集成研究成果的基础上,把数据、信息、知识进行跨层次集成以形成智慧,在学术思想上站得更高,继续研究智慧集成,对

此再次表示深深的感谢。集成知识是在知识网络的基础上所构成的集成化知识,集成智慧更具有超大规模性、整合性和融合性,其产生的知识集成技术具有更为强大的知识加工处理智能,能够为知识的超大规模处理提供有力的保障。本课题完成了面向学科数字信息群的集成智慧的一系列“构件”,已逐渐地构成了集成化智慧的基础。

我们的研究暂时告一段落,虽然感到有很多的结论要下,更感到有许多的研究工作还要深入下去。尽管经过努力,本书仍然可能存在缺点甚至错误,恳请各位专家及广大读者不吝批评、指正。本书撰写过程中广泛参考了相关文献,为减少篇幅按出版社意见将页下注释也合在参考文献中,网上引用的一些资料因历时较长或网页内容变更可能没有详细列清它们的来源,特向这些作者表示歉意和感谢。联系的 EMAIL 是 wanglancheng@163.com。

作者
2010 年于上海

目 录

第一章 绪论	1
1.1 学科数字信息群	1
1.2 知识集成	4
1.3 面向学科数字信息群的知识集成	8
1.4 研究目标与基本路向	13
1.5 内容概览	17
第二章 知识集成的实现基础	20
2.1 信息组织与检索技术	20
2.1.1 分类组织及检索法	21
2.1.2 主题组织及检索法	23
2.1.3 集成组织及检索法	26
2.2 知识集成中的信息技术应用	29
2.2.1 XML 和 RDF(S)技术	29
2.2.2 数据仓库技术	34
2.2.3 自然语言处理技术	37
2.2.4 语义网技术	39
2.2.5 搜索引擎技术	41
2.2.6 本体方法	43
2.3 本章小结	49

第三章	知识表示和本体管理	50
3.1	知识集成概述	50
3.1.1	知识的概念	50
3.1.2	知识集成的特点	52
3.2	领域知识表示	54
3.2.1	知识表示的几种方法	55
3.2.2	因素和因素空间	56
3.2.3	领域检索请求的描述架建立	59
3.2.4	基于真值流的知识推理	65
3.3	知识集成中的多本体管理	75
3.3.1	本体进化的方法	76
3.3.2	面向知识集成的本体概念整合	86
3.4	本章小结	89
第四章	知识集成中的元数据	90
4.1	元数据概述	90
4.1.1	元数据元素的语义层面	91
4.1.2	元数据元素的分类	92
4.1.3	元数据之间的关系	93
4.2	元数据在知识集成中的应用	96
4.2.1	元数据在知识表示中的应用	96
4.2.2	元数据之间的多种类型互操作方式	99
4.2.3	元数据登记注册管理机制	101
4.3	知识集成中的主要元数据比较	104
4.3.1	都柏林核心元数据(DC)	104
4.3.2	机读目录元数据(MARC)	108
4.3.3	学科数字信息群的其他元数据	110

4.3.4	主要元数据的比较	122
4.4	元数据之间的映射关系	124
4.4.1	MARC 与 DC 元数据	124
4.4.2	ERM 与 DC 元数据	132
4.4.3	MARC AMC 与 EAD 元数据	135
4.5	本章小结	137
第五章	基于本体的知识集成	139
5.1	知识集成中的本体方法	139
5.1.1	本体在知识集成中的主要功能	139
5.1.2	基于本体的知识集成	141
5.2	语法层面的知识集成	144
5.2.1	语法层的知识集成框架	144
5.2.2	语法层的知识集成方法	145
5.3	语义层面的知识集成	150
5.3.1	语义层的知识集成框架	150
5.3.2	语义层的知识集成方法	151
5.4	本章小结	154
第六章	知识集成中的知识组织	156
6.1	知识组织研究现状及比较	157
6.1.1	国内外关于知识组织的研究	157
6.1.2	相关项目的介绍及比较	159
6.1.3	学科数字信息群的知识组织需求	186
6.2	知识组织的需求分析及其工具	188
6.2.1	需求分析	188
6.2.2	知识组织工具	192
6.3	知识组织框架 LIA-CKOF 研究	198

6.3.1	学科知识工程中知识组织的定位	198
6.3.2	知识组织框架的分层结构	200
6.3.3	各层的内部组织结构	201
6.4	基于RDF的LIA-GKOF知识表示技术	204
6.4.1	基于RDF的元—元数据定义	204
6.4.2	基于DC元数据的互操作机制	211
6.4.3	应用知识组织工具的元数据内容组织	215
6.5	基于OAI-PMH协议的LIA-GKOF知识 发现技术	217
6.5.1	元数据收割协议OAI-PMH	217
6.5.2	LIA-GKOF中的资源注册与评估机制	228
6.5.3	LIA-GKOF的OAI-PMH协议实现	232
6.6	基于数据仓库的LIA-GKOF知识存储策略	237
6.6.1	联机事务处理与联机分析处理	238
6.6.2	数据仓库中的维度模型	238
6.6.3	LIA-GKOF中基于多维主题的知识组织	241
6.7	本章小结	244
第七章	知识集成中的知识检索	246
7.1	知识检索的研究和发展	246
7.1.1	从信息检索到知识检索	246
7.1.2	知识组织与知识检索	250
7.2	知识检索模型	254
7.2.1	信息检索模型	254
7.2.2	三类知识检索模型	255
7.2.3	基于本体的知识检索模型 Ont-KRM	256
7.3	知识检索呈现	259
7.3.1	知识检索呈现与信息检索呈现	259

7.3.2	基于本体的知识检索呈现方式	261
7.4	知识检索评估	263
7.4.1	衡量检索效率的主要标准	263
7.4.2	具有参与度的知识检索评估机制	264
7.5	本章小结	270
第八章	知识集成原型系统的设计与实现	271
8.1	概述	271
8.1.1	知识集成原型系统试验环境	271
8.1.2	LIA-GKOF 试验框架	274
8.2	知识组织框架设计原型	275
8.2.1	元—元数据定义	275
8.2.2	数据源注册与状态监测	276
8.2.3	数据提供者质量评估	277
8.2.4	元数据采集	278
8.3	基于本体的知识集成原型系统	280
8.3.1	系统总体介绍	280
8.3.2	一个领域本体原型	283
8.3.3	知识管理系统	285
8.3.4	知识检索系统	288
8.4	本章小结	291
第九章	结论与总结	293
9.1	全文总结	293
9.1.1	知识集成的主题研究	293
9.1.2	知识组织的方法和技术	294
9.1.3	知识检索的方法和技术	298
9.2	进一步工作	301

附录	304
图索引	319
表索引	323
参考文献	324
后记	342

Contents

Chapter 1 Overview	1
1.1 Disciplinary Digital Information Group	1
1.2 Knowledge Integration	4
1.3 Knowledge Integration Oriented to Disciplinary Digital Information Group	8
1.4 Research Objectives and Basic Approach	13
1.5 Content Overview	17
Chapter 2 The Basic Realization of The Knowledge Integration	20
2.1 Information Organization and Retrieval Technology	20
2.1.1 Classification Organization and Retrieval Method	21
2.1.2 Subject Organization and Retrieval Method	23
2.1.3 Integrated Organization and Retrieval Method	26
2.2 The Application of Information Technology on Knowledge Integration	29
2.2.1 XML and RDF (S) Technology	29