

郝金花 著

Xinxi Jiansuo Yu Zhenghe Liyong

信息检索与 整合利用

XINXI JIANSUO YU ZHENGHE LIYONG

中国矿业大学出版社

China University of Mining and Technology Press

XINXI JIANSUO YU ZHENGHE LIYONG

Xinxi Jiansuo Yu Zhenghe Liyong

信息检索与 整合利用

郦金花 著

中国矿业大学出版社

China University of Mining and Technology Press

图书在版编目(CIP)数据

信息检索与整合利用/酆金花著. —徐州:中国矿业
大学出版社,2006.12

ISBN 978 - 7 - 81107 - 601 - 1

I. 信… II. 酆… III. 情报检索 IV. G252.7

中国版本图书馆 CIP 数据核字(2006)第 022062 号

书 名 信息检索与整合利用

著 者 酆金花

责任编辑 罗时嘉 齐 畅

出版发行 中国矿业大学出版社

(江苏省徐州市中国矿业大学内 邮编 221008)

网 址 <http://www.cumtp.com> E-mail cumtpvip@cumtp.com

排 版 中国矿业大学出版社排版中心

印 刷 徐州中矿大印发科技有限公司

经 销 新华书店

开 本 880×1230 1/32 印张 12.25 字数 318 千字

版次印次 2006 年 12 月第 1 版 2006 年 12 月第 1 次印刷

定 价 26.00 元

(图书出现印装质量问题,本社负责调换)

前 言

文献检索利用的思想伴随着人类整个的文献组织及管理的各个阶段和过程中。随着分类法等检索语言的逐渐发展以及检索工具的日益繁荣,20 世纪初,各类图书馆才真正开展了利用书目工具等的参考咨询工作,此是正规的文献检索利用工作的起源。随后,信息技术不断发展,文献检索由手工检索发展到机械检索、单机光盘检索、联机检索以及目前的信息网络检索。由此,文献检索拓展到了信息检索;文献检索的理论、方法、技术都得到了前所未有的迅速发展。相应地,为了使用户快速适应新的文献(信息)组织和检索方法,各国积极开展有针对性的文献(信息)检索教育与培训。对于文献(信息)检索的研究主要集中在两个方面:① 信息检索方法:这方面的研究主要是针对狭义的信息检索过程,进行用户信息检索教育与培训,由此出版了大量的有关书籍,内容主要是对大量的检索工具和检索系统的检索方法的罗列。② 信息检索技术:在计算机信息检索的过程中,信息检索需求与检索系统中信息集合的匹配运算是信息检索的关键技术。目前,已经作为重要的技术研究,有少量专著对

信息检索的匹配算法和技术作专门的研究。

然而,广义的信息检索包括信息的存储与检索的整个过程。实际涉及到的相关理论比较广泛,尤其是信息检索语言、元数据的发展、信息检索协议以及文献(信息)计量学等理论的发展,对于信息在网络环境下的组织与共享、检索与利用都起着重要的作用。而随着各种数据库的数量和种类的快速增长,进一步提出和实践基于各种异质数据库和异构系统的信息整合与检索利用。本书将对信息检索全过程中涉及的一些主要相关理论与信息整合利用做一些研究与探讨。

本书分为两个部分:

第一部分为“信息检索”。包括六章:1. 信息检索概述;2. 信息检索语言;3. 元数据——信息著录标准;4. 信息检索协议:Z39.50 协议;5. 信息收集、检索利用中的计量方法——文献计量学;6. 信息检索教育。

第二部分为“信息整合与利用”。包括四章:7. 信息整合概述;8. 信息资源整合模式及系统;9. 信息整合实践;10. 信息检索利用——基于信息整合的图书情报平台。

本书的总体思路是,在阐述各种相关理论基础的同时,结合图书馆实践,进一步探讨其在信息检索、信息整合与利用中的应用及作用。

本书从文献检索思想的产生、文献检索理论的产生与发展到信息检索的实践与理论的拓展过程展开阐述,全书在论述相关理论及其在信息检索、信息整合与利用中的应用及作用的同时,力求将新的研究成果纳入其中。例如第二章信息检索语言。在重点论述检索语言类型、分类语言和主题语言的基础上,进一步探讨了网络环境

下检索语言的发展趋势。第三章信息著录标准:元数据。主要讨论 MARC 元数据、DC 元数据在信息组织中的应用与发展, MARC 与 DC 在信息检索中的作用。根据作者的学位论文成果,设计了基于 XML 和 CNMARC 字典库的 MARC 发布系统,使得网络环境下 MARC 数据可以面向 Web 检索利用。第四章,信息检索协议:Z39.50 协议。在介绍了 Z39.50 协议的概念及基本原理并结合图书馆实际探讨其在图书馆管理系统中的各种应用实践后,指出 Z39.50 协议基于 XML 的改进,新一代 ZING 与 Z39.50—1995 的关系,以及 ASN.1 数据到 XML 文档之间的映射。第五章,信息收集、检索利用中的计量方法——文献计量学。阐述文献计量学的由来基础上,还着重探讨了文献计量学中的三大定律在文献信息收集、检索利用中的作用,进而结合作者的研究成果,探讨了三大定律的分形意义和统一模式。同时,基于引文分析法在信息收集、检索利用中的广泛应用,结合实际,利用影响因子等指标对《大学图书馆学报》及近 5 年我国图书馆学情报学研究之影响做了案例分析研究。第六章,信息检索教育。在回顾我国开设文献检索课情况基础上,探讨了教材、师资建设等问题,根据作者研究成果提出教学改革——基于提高学生初步科研能力(创新能力)和信息素质的改革思路。在信息整合部分,结合新的研究成果和实践进行了探讨,等等。

本书是江苏省教育厅“高校哲学社会科学基金项目 05SJB870001”的成果之一;是以作者的学位论文和发表的相关研究论文为基础,结合以往的文献检索教学、研究的实践,并参考了有关学者的研究成果撰写而成的。在此,对所有标注或遗漏的参考文献的著者表示衷心感谢!

感谢 CALIS、CASHL、JALIS、南京大学、徐州师范大学等网站为本书提供了许多有价值的资料。感谢我的先生在我研究和写作过程中给予的关心和支持,感谢中国矿业大学出版社的领导和编辑同志的支持。我的研究成果能够面世,与他们的热情相助是分不开的。

由于作者水平所限,写作时间仓促,书中内容难免有不妥之处,恳请专家学者批评指正。

目 录

前言	1
----------	---

第一篇 信息检索

1 信息检索概述	3
1.1 信息检索理论的起源	3
1.1.1 文献检索思想的产生	3
1.1.2 目录的文献检索利用功能	4
1.1.3 目录学的内容与文献检索	6
1.2 “文献检索”理论的形成及其影响	6
1.2.1 文献信息存储理论的形成	7
1.2.2 文献检索利用理论方法的形成	11
1.2.3 “文献检索”理论的形成作用、意义及影响	13
1.3 信息检索形成的现实背景	15

1.4	“文献检索”向“信息检索”的拓展	17
1.5	信息检索概念	21
1.5.1	信息的概念及类型	21
1.5.2	信息资源的概念及类型	26
1.5.3	信息检索	33
2	信息检索语言	35
2.1	检索语言的概念	35
2.2	检索语言的作用	36
2.3	信息检索语言的类型	37
2.3.1	人工语言	38
2.3.2	自然语言	39
2.3.3	人工语言与自然语言的关系	40
2.4	分类语言	41
2.4.1	国外重要分类法	41
2.4.2	中国重要图书分类法	44
2.5	主题检索语言	56
2.5.1	主题检索语言概述	56
2.5.2	关键词语言	57
2.5.3	标题词语言	59
2.5.4	单元词语言	61
2.5.5	叙词语言	62
2.6	信息(情报)检索语言的发展趋势	64
2.6.1	分类主题一体化——学科检索与事物检索 一体化	65

2.6.2	受控语言与自然语言的融合·····	66
2.6.3	情报(信息)检索语言的计算机技术化方向·····	67
2.6.4	关于“本体语言”的研究·····	69
3	元数据——信息著录标准·····	71
3.1	元数据概念·····	71
3.2	元数据格式的分类·····	72
3.3	传统元数据·····	73
3.3.1	MARC 元数据的产生 ·····	73
3.3.2	MARC 的记录结构 ·····	74
3.3.3	MARC 元数据的特点 ·····	76
3.4	现代元数据·····	78
3.4.1	DC 元数据 ·····	79
3.4.2	XML 标记语言 ·····	83
3.5	MARC 与 DC、MARC 与 XML ·····	86
3.5.1	MARC 与 DC 的相互转换 ·····	86
3.5.2	基于 XML/ RDF 的 MARC 元数据描述 技术·····	87
3.5.3	基于 XML 的 MARC 发布 ·····	89
3.6	元数据理论在信息检索利用中的作用·····	95
3.6.1	查询工具——搜索引擎、网络指南的困境 ·····	95
3.6.2	MARC 元数据在网络信息组织和检索中 的应用·····	97
3.6.3	DC 元数据在信息组织与检索中的重要作用 ·····	99
3.6.4	XML 在信息检索中的作用 ·····	105

4 信息检索协议:Z39.50 协议	106
4.1 Z39.50 协议的概念及产生	106
4.1.1 Z39.50 协议的概念	106
4.1.2 Z39.50 协议的产生	106
4.2 Z39.50 协议的原理与特征	109
4.2.1 Z39.50 协议的原理	109
4.2.2 Z39.50 协议的特征	111
4.3 Z39.50 协议提供的服务	112
4.4 基于 Z39.50 的分布式信息检索系统	115
4.4.1 基于 Z39.50 的分布式信息检索系统的 体系结构	115
4.4.2 基于 Z39.50 的分布式信息检索系统 实例	117
4.5 Z39.50 协议在分布式信息检索系统中的 具体应用	122
4.5.1 Z39.50 协议在国外图书馆中的应用	122
4.5.2 Z39.50 协议在我国图书馆中的应用	123
4.6 Z39.50 协议的发展前景	132
4.6.1 新一代 Z39.50:ZING	132
4.6.2 从 ASN.1 到 XER 编码规则	134
5 信息收集、检索利用中的计量方法——文献计量学	140
5.1 文献计量学的由来	140
5.2 文献计量学研究对象、目的、意义及作用	141
5.3 文献计量学与信息检索利用	142

5.3.1	布拉德福定律与信息检索利用	143
5.3.2	洛特卡定律与信息检索利用	149
5.3.3	齐夫定律与信息检索利用	151
5.3.4	三大计量定律的分形模型及意义	155
5.3.5	引文分析与信息检索利用	162
6	信息检索教育	185
6.1	“文献检索与利用”课程的开设及问题	186
6.2	课程开设的简要历程、范围及要求	186
6.2.1	大力普及、推广发展阶段——我国 20 世纪 80 年代的文献检索与利用课阶段	187
6.2.2	理性建设与发展阶段——我国 20 世纪 90 年代的 文献检索课	191
6.2.3	信息素质教育阶段——21 世纪的信息检索课 阶段	193
6.3	教材问题	194
6.3.1	文献检索教材的发展	196
6.3.2	文献检索教材普遍的拓展——信息检索	200
6.3.3	教材改革对策	203
6.4	师资问题	205
6.4.1	师资状况	205
6.4.2	解决师资困境的对策	206
6.5	课时、教学问题及影响教育效果的因素	208
6.5.1	课时问题	208
6.5.2	教学问题	208

6.5.3	影响教育效果的因素	208
6.6	教学改革	209
6.6.1	文献(信息)检索课应注重学生初步科研 能力的培养	209
6.6.2	文献(信息检索)课的教学目标应由检索 技能型向信息素质型转变	215

第二篇 信息整合与利用

7	信息整合概述	225
7.1	信息整合概念	225
7.1.1	整合概念	225
7.1.2	信息资源整合	226
7.2	信息整合与信息组织	228
7.2.1	信息组织概念	228
7.2.2	信息整合与信息组织的相互关系	230
7.3	信息资源共享理论、实践与信息资源整合	231
7.4	信息资源整合与检索利用	233
7.5	信息整合的原则	234
7.5.1	递进性原则	234
7.5.2	科学性原则	235
7.5.3	实用性原则	237
7.5.4	共享性、协调性原则	238
7.5.5	经济性原则	238

7.5.6 层次性原则	239
7.5.7 动态性原则	239
7.5.8 效益性原则	240
7.6 信息资源整合研究的内容	241
7.6.1 信息用户	242
7.6.2 信息资源的组织和结构	246
7.6.3 信息技术与系统	247
7.6.4 信息整合的信息机构和管理机制	247
8 信息资源整合模式及系统	249
8.1 信息资源整合的必然性	249
8.1.1 信息组织管理的必要性	249
8.1.2 信息服务的必要性	250
8.2 信息资源整合的模式	250
8.2.1 信息资源整合模式	251
8.2.2 信息资源整合模式的发展	258
8.3 信息整合系统	266
8.3.1 国外信息整合系统	267
8.3.2 国内信息整合系统	270
8.3.3 几种软件比较	274
8.4 影响信息整合的因素	277
8.4.1 技术因素	277
8.4.2 机制因素	278
8.4.3 历史因素	278
8.4.4 管理水平因素	279

9 信息整合实践	280
9.1 以图书馆实体与虚拟资源为主的信息整合	281
9.1.1 基于 OPAC 的信息资源整合:馆藏文献与 电子文献的整合	281
9.1.2 电子资源一站式检索:电子数据库资源的 整合	284
9.1.3 学科导航或学科门户:网络学科专题资源 的整合	285
9.1.4 信息整合服务创新:个性化信息推送服务	290
9.2 地市级区域性信息保障与共享整合	291
9.2.1 服务基础构架	292
9.2.2 资源建设与共享	293
9.2.3 信息服务管理	294
9.3 国家级或行业性学术信息资源整合	296
9.3.1 高等教育系统的 CALIS 概况	297
9.3.2 CALIS 信息资源整合与服务实践	302
 10 信息检索利用——基于信息整合的图书情报平台	 308
10.1 图书馆平台可检索利用的信息资源与服务	309
10.1.1 图书馆信息资源类型	309
10.1.2 图书馆提供的信息服务	314
10.2 基于图书馆信息整合的信息检索利用	315
10.2.1 基于 OPAC 信息整合的检索利用	315
10.2.2 基于馆际信息整合与资源共享的检索 利用	323

10.2.3 基于数据库整合的信息检索利用·····	341
10.2.4 基于网络信息整合的学科导航·····	347
10.2.5 基于共建、共享的特色数据库的信息 检索·····	350
10.2.6 虚拟联合参考咨询·····	351
10.3 其他开放学术信息的检索利用·····	354
10.3.1 电子预印本·····	354
10.3.2 开放获取期刊·····	361
参考文献·····	367

第一篇

信息检索