



大学数字图书馆国际合作计划
CHINA ACADEMIC DIGITAL ASSOCIATIVE LIBRARY



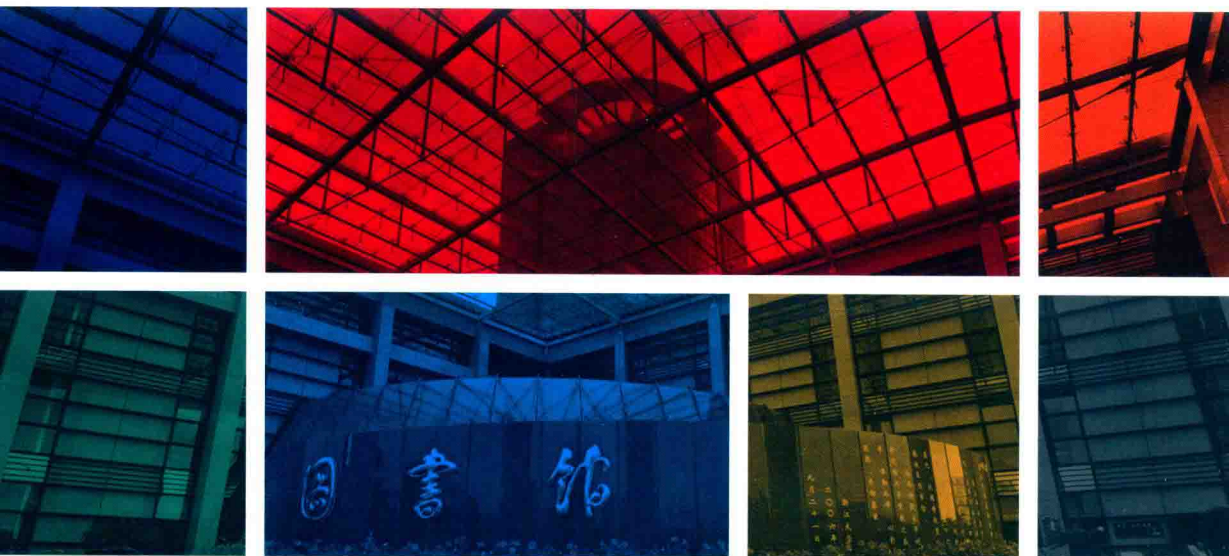
CADAL
项目标准规范丛书



基于“学术单元”的 知识组织新框架： 多维度标签构建研究

潘云鹤◎丛书主编

韩松涛◎编 著



ZHEJIANG UNIVERSITY PRESS

浙江大学出版社



NUAA2017031855

合作计划
ATIVE LIBRARY



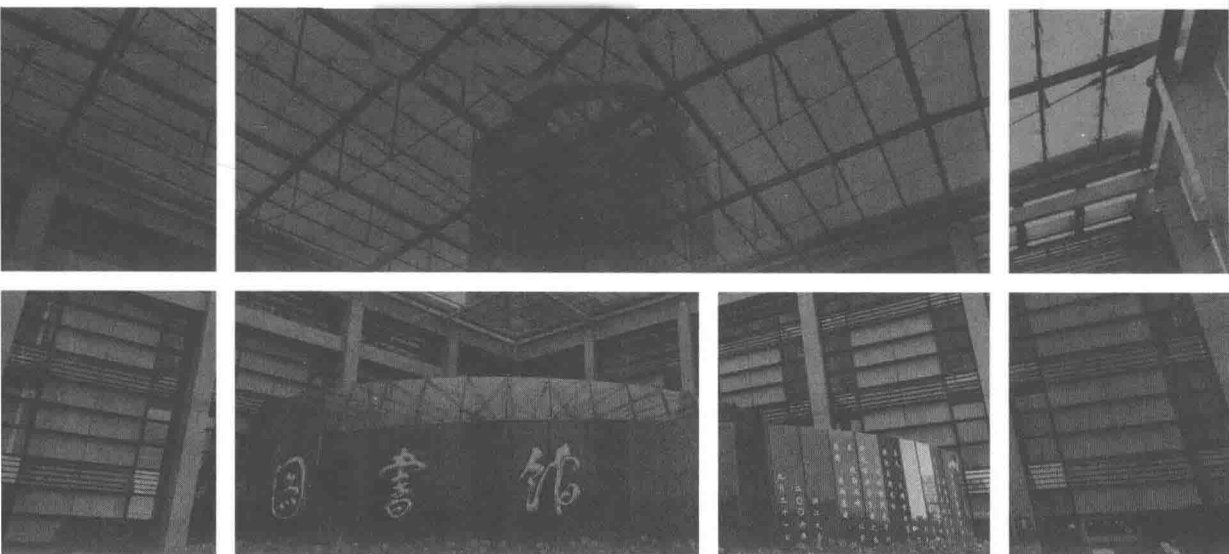
项目标准规范丛书



基于“学术单元”的 知识组织新框架： 多维度标签构建研究

潘云鹤◎丛书主编

韩松涛◎编 著



ZHEJIANG UNIVERSITY PRESS

浙江大学出版社

图书在版编目 (CIP) 数据

基于“学术单元”的知识组织新框架:“多维度标签”构建研究 / 韩松涛编著. —杭州:浙江大学出版社, 2017. 6

(CADAL 项目标准规范丛书 / 潘云鹤主编)

ISBN 978-7-308-16815-1

I. ①基… II. ①韩… III. ①院校图书馆—数字图书馆—图书工作 IV. ①G258.6 ②G250.76

中国版本图书馆 CIP 数据核字 (2017) 第 075712 号

基于“学术单元”的知识组织新框架:

“多维度标签”构建研究

Jiyu “Xueshu Danyuan” De Zhishi Zuzhi Xinkuangjia:

“Duoweidu Biaoqian” Goujian Yanjiu

韩松涛 编著

责任编辑 张凌静 (zlj@zju.edu.cn)

责任校对 杨利军 陈思佳

封面设计 续设计

出版发行 浙江大学出版社

(杭州市天目山路 148 号 邮政编码 310007)

(网址: <http://www.zjupress.com>)

排 版 杭州中大图文设计有限公司

印 刷 杭州日报报业集团盛元印务有限公司

开 本 787mm×1092mm 1/16

印 张 14

字 数 341 千

版 次 2017 年 6 月第 1 版 2017 年 6 月第 1 次印刷

书 号 ISBN 978-7-308-16815-1

定 价 56.00 元

版权所有 翻印必究 印装差错 负责调换

浙江大学出版社发行中心联系方式 (0571)88925591; <http://zjdxcs.tmall.com>

“CADAL 项目标准规范丛书”编委会

主 编 潘云鹤

副主编 庄越挺 黄 晨

编 委 (按姓氏笔画排序)

邢造宇 朱腾云 刘 翔 刘柏嵩

孙 锐 孙晓菲 肖丽梅 吴江琴

张 浩 张 寅 陈 蕾 竺海康

金更达 贾雪梅 彭 佳 韩松涛

潘 晶

总 序

春华秋实,“CADAL 项目标准规范丛书”即将正式出版了。该丛书不仅汇编了历年来 CADAL 项目标准规范建设的成果,而且包含了 CADAL 人在标准规范建设中的经验总结和理论探索,是 CADAL 项目建设的基石,也是 CADAL 项目成果中最璀璨的明珠之一。

CADAL 项目的建设始于 2000 年。那时,全球的数字图书馆建设热潮初起,包括卡内基梅隆大学、中国科学院和浙江大学等机构在内的中美两国计算机科学家共同发起了“中美百万册书数字图书馆合作计划(China-US Million Book Digital Library Project)”,目标是建设一百万册电子图书系统,并逐步向用户提供服务。2000 年 12 月,“百万册书计划”项目正式启动,项目定名为“高等学校中英文图书数字化国际合作计划(China-America Digital Academic Library, CADAL)”。

2002 年 9 月,国家计划委员会、教育部、财政部在《关于“十五”期间加强“211 工程”项目建设的若干意见》的文件中,将“高等学校中英文图书数字化国际合作计划(CADAL)”列入“十五”期间“211 工程”公共服务体系建设, CADAL 与“中国高等教育文献保障系统(China Academic Library & Information System, CALIS)”一起,构成了中国高等教育数字化图书馆的基本框架。2005 年 11 月,作为首个有外资参与的“211 工程”项目, CADAL 完成了百万册资源的加工工作,拥有了较为齐全和完整的数字化特藏,初步建成了中国高等教育数字化图书馆。作为当时全球最大的公益性数字图书馆, CADAL 不仅数据量大,而且学术性强、开放程度高,100 万册图书向全世界开放提供数字化信息服务, CADAL 一期建设圆满完成。

2008 年正值我国数字图书馆建设规模、数字资源的海量管理乃至数字图书馆技术皆迅猛发展的时期。为了更好地实现“构建拥有多学科、多类型、多语种海量数字资源的,由国内外图书馆、学术组织、学科专业人员广泛参与的,具有高技术水平的学术数字图书馆”的项目建设总体目标,教育部决定继续投资 CADAL 建设,并将二期更名为“大学数字图书馆国际合作计划(China Academic Digital Associative Library, CADAL)”。建设方案几经酝酿、修改,2009 年 8 月, CADAL 项目二期可行性研究报告通过了教育部专家的评审和论证。作为“211 工程”“高等教育文献保障体系”的两大专题之一,2010 年 4 月,项目二期正式启动,并于 2012 年 5 月通过项目二期验收。通过一、二期十年的建设, CADAL 项目以 250 万册的数字资源总量,继续保持国内外公益性数字图书馆规模的领先地位,实现了对科学技术与文化艺术的多种类型媒体资源的数字化整合,有能力向读者提供一站式个性化知识服务,已成为国家创新体系中重要的学术信息基础设施之一。

CADAL 的建设,见证了中国数字图书馆事业的成长历程:从蹒跚学步到健步如飞,而今不仅风华正茂,而且更充满了自信和创新的活力。

标准为人类文明的发展提供了重要的技术保障, CADAL 标准规范研究始于其建设之

初,并贯穿了 CADAL 建设的全过程。CADAL 人深知,当今世界,标准化已经成为衡量国家、行业、企业、项目核心竞争力的基本要素,“得标准者得天下”。

在 CADAL 项目建设之初,国际上已经有一些涉及数字图书馆的初步标准规范成果,中国的图书馆界正开始这方面的研究和实践。作为一个由海量数据、众多应用软件组成的复杂层次系统,完全依靠“拿来主义”,不仅不能满足项目建设的需要,而且还会给项目的发展带来更大的障碍。

因此,CADAL 项目的标准规范建设,始终立足于项目实践。在吸收、借鉴国内外数字图书馆标准研究已有的规范、标准成果的基础上,CADAL 密切关注数字图书馆快速发展过程中出现的新技术、新介质、新形式,根据数字资源加工、发布、存储、服务、应用等流程,围绕数字资源创建、描述、组织、检索、服务和长期保存的整个过程,针对文本、图像、音频、视频、动漫数据等多媒体的数字资源,构建了一个由数字对象、资源集合、信息组织、知识组织和辅助等5大框架,包含元数据标准、加工规范、多维度标签分类标准、知识标引标准、数字版权规范、数字图书馆评估规范等16个标准规范集组成的 CADAL 项目标准规范体系,全面覆盖了数字资源从内容策划到创建、组织描述、保存管理、获取和整合,再到维护和提供服务的完整流程。CADAL 项目标准规范的建设过程,是一个典型的“从实践中来,到实践中去”的过程。

CADAL 项目的标准规范建设遵循的是对数字图书馆发展趋势的正确判断。因而,CADAL 项目的标准规范,有的是遵循国际数字图书馆建设的互操作规范,将国内外的相关标准规范经过改造为我所用的;有的是根据项目建设的实际需要,经过项目建设检验的实践经验固化成文的;有的是适应新的发展需求形成的新规范。这些标准规范遵从一定的规则,形成系列,并构筑起一个完整的体系,既能指导实践,又具开放性、规范性,更保持可持续发展;从另一角度科学合理地诠释了数字资源从无序的集合到有序的信息组织,继而成为知识本体的生命周期;指导了 CADAL 项目的建设实践,引导了数字图书馆的发展方向。

“CADAL 项目标准规范丛书”的“汇编”部分,系统完整地展示了 CADAL 项目标准规范建设的成果;而丛书的另一部分,则收集了建设者们关于标准规范的部分理论研究成果。读者可以发现在将海量资源从信息变为知识,进而形成知识与情报网络的建设实践中,融入了大量对数字图书馆的新认知和新探索,诠释出 CADAL 项目服务新模式的建立和关键技术的突破。

信息技术的快速变革,为数字图书馆标准规范的制定增添了相当的难度。目前,包括 CADAL 项目在内的数字图书馆标准规范,大多是事实标准,需要随着技术的发展不断更新和修订。本丛书编入的部分标准规范,已经过多次修订和完善。如何兼容频繁变化的信息技术,使其对数字图书馆建设的当前与未来皆具指导意义,是 CADAL 项目标准规范建设的一个新挑战。

而 CADAL,已经做好准备,迎接新的挑战。

潘雲鵬

2014年9月

前言

构建多维度标签的任务,起始于跟 CADAL 管理中心的黄晨研究馆员的一次基于数字图书馆的聊天。我们谈到了关于信息的空间距离问题,然后谈及冒号分类法的五个维度,并且提到了五个维度可以构建一个多维度标签——可以用多个体系构建一个信息的组织体系,形成立体的信息组织体系,于是信息之间就有空间距离的问题;又谈到阮冈纳赞真是一个天才,虽然他的“冒号分类法”在当时不能够被广泛使用,但他的分面组配的方法却好像是为互联网时代定制的。在互联网时代,以之脱胎而来的多维度标签一定是一个优秀的信息与知识的组织方法。

在黄晨研究馆员的支持和鼓励下,我接受了一个任务。一个具有海量资源的数字图书馆,里面有 200 万册的图书、期刊、学位论文等文献,有元数据,但没有全文。我的任务是对其中的中文图书用知识标引的方法将每一种书中的知识进行标引,并用一个知识组织方法将其组织起来。这个数字图书馆的名称是 China Academic Digital Associative Library(CADAL)。

用知识管理的形式对数字图书馆的文献进行组织、揭示,是这个时代提出的要求。现有的知识管理方法有知识本体、主题地图、知识挖掘等,但是这些研究无法为我完成任务提供有效的帮助。这些知识管理方法的出发点、方法、目标等,都与我要完成的任务不相契合,因此我只能另辟蹊径。CADAL 中用了 Academic,也就是“学术”这个词,说明学术是 CADAL 这个数字图书馆的最大特色。在《冒号分类法》的启发下,我们准备构建一个基于“学术单元”的知识组织新框架——“多维度标签”。在 CADAL 管理中心的支持下,我们开始了艰难的研究。初步的成果是制定了 CADAL 标准——《CADAL 多维度分类标准》,并用这个标准对 CADAL 百万量级图书进行了多维度标签自动标引的初步试验。现在我们又 CADAL 项目的支持下,准备将这个研究进行一个全面的阐述,于是就有了这篇篇幅不长的著作。

我想本书的创新点有两个:一是提出了“学术单元”这个概念。知识单元是现在知识组分的重点,知识单元强调的是对知识的划分,形成知识组织的最小单位。从文献单元到知识单元,人们关注的是从大单元到细胞级单元的变化。学术单元概念的提出,脱离了思维的限制,从另一个角度来讨论知识组分的问题。个人认为脱离了从大到小的思维限制,是这个创新的闪光点。第二个创新点是构建了多维度标签。对于过于自由的标签的控制,已经成为标签研究的一个热点,也有学者提到过用多面组配的方式对标签进行控制。个人认为第二个创新点是以学术单元为基础,找到了属于学术单元的多个分面,或称多个范畴,从而构建了以学术单元为基础的多维度标签。为学术资源的组织、学术数字图书馆的建设提供了一个新的知识组织标准。

期望基于“学术单元”的知识组织新框架——“多维度标签”构建研究能在知识组织的研究与实践中打开一扇新的大门。

韩松涛

2016 年 8 月

目 录

第1编 知识组分研究	1
第1章 当前知识组分研究概述	3
1.1 文献单元、信息单元和知识单元	3
1.1.1 文献单元	3
1.1.2 信息与信息单元	4
1.1.3 知识单元	5
1.1.4 文献、信息、知识辨析	7
1.2 文献单元及其价值	7
1.2.1 文献服务与文本服务	7
1.2.2 信息机构与互联网的信息服务比较	9
1.2.3 文献单元新论	10
第2章 学术单元理论	12
2.1 知识产生的主要活动——学术活动及其特征	12
2.1.1 知识的产生	12
2.1.2 作为知识产生的学术活动及其特征	14
2.2 学术单元概念及其意义	17
2.2.1 学术单元概念	17
2.2.2 学术单元与文献单元、知识单元的关系	18
2.2.3 提出学术单元概念的意义	19
第2编 知识组织框架研究	21
第3章 当前知识组织方法简述	23
3.1 形式本体论	23
3.1.1 形式本体论的哲学背景与基本概念	23
3.1.2 本体的构建语言与工具	24
3.1.3 本体的应用研究	26
3.2 主题地图	30
3.2.1 主题地图的产生	30
3.2.2 主题地图的结构	31
3.2.3 主题地图的应用	32

3.3	数据挖掘与知识发现	33
3.3.1	数据挖掘与知识发现的概念	33
3.3.2	数据挖掘与知识发现的研究热点	34
3.3.3	以知识元为核心的数据挖掘与知识发现	36
3.4	当前大型数据库的组织体系	36
3.4.1	引文分析理论	37
3.4.2	科学计量学	39
3.5	知识管理诸概念辨析	40
3.5.1	知识组织	41
3.5.2	知识标引	42
3.5.3	知识挖掘	44
第4章	多维度标签的理论来源	46
4.1	标签及其应用	46
4.1.1	大众分类法——标签	46
4.1.2	标签与学术资源标引	49
4.1.3	标签存在的问题及解决方案	50
4.2	分面组配分类法	51
4.2.1	体系分类法与分面组配分类法的历史回顾	51
4.2.2	《冒号分类法》概述	53
4.2.3	布利斯《书目分类法》第二版简述	59
4.3	知识组织框架下分面组配与标签研究	60
4.3.1	分面组配分类法应用研究	60
4.3.2	知识组织框架下的分面组配与标签研究	61
第5章	多维度标签的构建——基本框架与标注标准	64
5.1	多维度标签构建概述	64
5.1.1	各维度的确立	64
5.1.2	多维度标签的属性	66
5.2	多维度标签基本维度	67
5.2.1	本体	67
5.2.2	对象	68
5.2.3	方法	69
5.2.4	时间	71
5.2.5	空间	72
5.3	多维度标签扩展维度	73
5.3.1	语种	73
5.3.2	学术等级	73
5.3.3	作者状况	73
5.4	维度与标注标准	73
5.4.1	单层多维度标签	74

5.4.2 多层多维度标签	74
5.5 多维度标签的人工标注与自动标注	76
5.5.1 多维度标签的人工标注	76
5.5.2 多维度标签的自动标注实验	77
第3编 多维度标签应用研究	89
第6章 以多维度标签作为数据源的分析研究展望	91
6.1 标签本体研究展望	91
6.2 利用标签构建网络词表研究展望	92
6.2.1 标签词表与叙词表	93
6.2.2 标签词表与自动标引	95
6.3 多维度标签的可视化研究展望	96
6.3.1 信息可视化与知识可视化	96
6.3.2 标签可视化研究现状——标签云	97
6.3.3 多维度标签可视化研究展望	99
第7章 多维度标签的领域应用——以道教学科为例	101
7.1 本体维度	101
7.1.1 道教分类法研究综述	101
7.1.2 《新编道教分类法》	107
7.2 对象维度	110
7.3 方法维度	111
7.3.1 田野调查	111
7.3.2 文献	111
7.4 时间维度	112
7.5 空间维度	115
7.5.1 十大洞天	115
7.5.2 三十六小洞天	117
7.5.3 七十二福地	120
7.5.4 三十六靖庐	126
7.6 扩展维度	127
7.7 多维度标签标注举例	128
第8章 多维度标签在电子商务中的应用	130
8.1 电子商务网站的多维度标签现状	130
8.1.1 淘宝网、当当网多维度标签使用现状	130
8.1.2 亚马逊中文网站的多维度标签使用现状	133
8.2 电子商务网站多维度标签维度重设及其意义	134
8.2.1 电子商务网站多维度标准维度设计	134
8.2.2 电子商务网站设立多维度标签的意义	135

附录 1 CADAL 标准——《多维度标签分类标准》 137

附录 2 词表 151

1-1 GBT 13745—2009《学科分类与代码》(只录一级学科) 153

1-2 《非学术文献分类表》(附综合性学术文献代码表) 156

1-3 《中图法类目与学科对照表》(以“B9 宗教”为例) 158

3-1 《类型名称汇总表》 163

3-2 《方法论名称汇总表》 163

4-1 《宇宙空间名称表》 165

4-2 《地形地貌名称表》 170

4-3 《世界国家及首都名称简表》 171

4-4 《中国各省市县名称简表(以浙江省为例)》 178

5-1 《地质年代表》 182

5-2 《世界时代表》 183

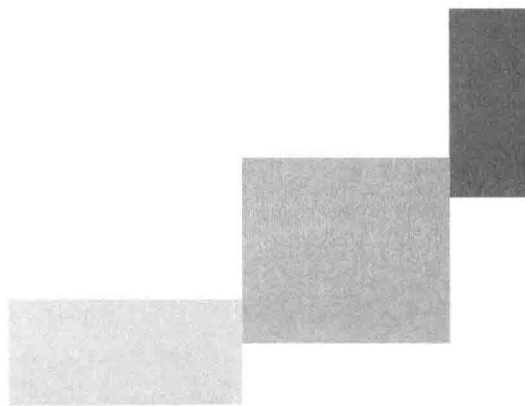
5-3 《中国朝代帝王年号简表》 183

5-4 《六十甲子表》 212

索引 213

第 1 编

知识组分研究



第1章 当前知识组分研究概述

知识组织是以知识组分为基础的。讨论知识组分的学者,如王子舟、文庭孝等,将知识组分分为文献单元、信息单元、文本单元、知识单元等,也有学者如温有奎提出“知识元”^①的概念,其概念基本等同于狭义的知识单元。现有研究主要集中于探讨文献单元和知识单元。本课题组提出“学术单元”的概念,虽然具有创新的意义,但不是另辟蹊径,而是在文献单元与知识单元研究上的进一步拓展。

1.1 文献单元、信息单元和知识单元

从广义上讲,文献单元、信息单元都是知识单元的一种类型,但从狭义上讲,三者各有不同。从时间上来说,文献单元是最早出现的,也是经历了数千年的一个重要的知识管理的单位。信息单元、内容单元是文献单元和知识单元研究中的一个过渡或者中间概念,虽然存在时间不长,但“信息”的概念已经发展为“信息资源”的概念,成为所有文献、信息、知识的总称,并转化为图书馆学的新名称“信息资源管理”。随着知识管理的兴起,知识单元作为知识管理的最重要元素,成为主要的研究对象。由于目前认为知识单元是组成知识的最小单位,故知识单元也被称为知识元。

1.1.1 文献单元

文献从外在形态上讲,自成一个独立性的单元,故文献单元与文献的概念是基本等同的。我们在讨论文献单元时,也可以将文献作为一个相同概念进行讨论。

国家标准对文献的定义是较为权威并具有广泛影响的,GB 3469—83《文献类型与文献载体代码》与 GB 379211—83《文献著录总则》均将文献定义为“记录有知识的一切载体”。王子舟认为这个文献的定义存在某些疏漏,指出记录有知识的人工载体很多,除了图书以外,还有钱币、邮票、商周时期的青铜器等。王子舟认为文献必须是专用于传播知识、交流知识的,它在用途、功能上具有知识传播的唯一性,而上述钱币、邮票、青铜器等并不具有传播知识的本身属性。王子舟因而提出了“并非记录了知识的都是文献”这样一个概念,并将“文献”重新定义为“专门记录和传递有知识的人工载体”^②,强调其传递知识的功能。这样的定义有其合理性,但我们也可以把国家标准对文献的定义看作是广义的定义,而王子舟的则是狭义的定义。

① 温有奎,徐国华,赖伯年,等.知识元挖掘[M].西安:西安电子科技大学出版社,2005.

② 王子舟,王碧滢.知识的基本组分——文献单元和知识单元[J].中国图书馆学报,2003(1):5-11.

文献单元以载体形式存在,是一个普遍被公认的观点。同时,也认为文献单元往往以独立的载体形式存在,举例如简帛、图书、光盘等,这个认识却是有偏差的。

期刊中的学术论文是一种文献单元,而且是现实存在的数量巨大的文献单元,而期刊则是由多篇论文集合成的一个独立载体。学术论文之所以被认定为文献单元,其实是根据其外在的独立形式来判断的,如开头为题名,其后为责任者,再则后有文摘关键词,中间是正文,最后是引文。所以论文作为文献单元,与图书的不同点是其不具有独立的载体形式,共同点应该是与图书一样,具有外在的独立形式。

所以图书与论文为代表的文献单元,其共同点是具有外在的独立形式,这种形式或者是物理上的,或者是逻辑上的。国际文献工作标准化组织批准的《文献工作术语标准:情报与文献用术语(草案)》(ISO/DL 5127)对文献的解释为:“Document,是指在存贮、检索、利用或传递记录信息的过程中,可以作为一个单元处理的,在载体内、载体上或依附载体而存贮有信息或数据的载体。”^①从“依附载体”来看,其概念包括期刊论文。期刊论文不具有独立的载体形式,而是“依附”于期刊这个“载体”。所以论文这种文献,虽然不具有独立的载体,但依然离不开载体。

从上述的讨论可以看出,文献的关键内容是载体。如对古籍的修复,其实是指对古籍载体——纸张的修复,载体存在了,载体上记载的知识也就存在了。所以图书馆对文献的管理,其实是对载体的管理;对文献的著录,部分内容是对载体的描述,如文献的大小(以前用开本,现在用书脊长度)、页数;部分内容是对文献出版情况的描述,如作者、出版地、出版社、出版年,等等;只有少部分涉及文献的内容,但也往往只有书名一项。所以图书馆对文献的管理是一种形式上的管理,而非内容上的。随着时代的发展,这种管理已经不能满足需求了,于是出现了信息单元、知识单元的研究。

从广义的知识单元来说,文献单元是一个确定的知识单元,可以作为一个独立的知识单元处理。同时,文献单元包含多个狭义的知识单元,由于不能直接将物理的文献单元中的知识单元提取出来,所以有学者认为文献单元是知识管理的一种间接单元。

1.1.2 信息与信息单元

与文献及文献单元不同,“信息”与“信息单元”两个概念有很大的差别。原始的信息概念主要是指动态的数据和消息,如股市的行情、经济运行的数据、新闻报道等。“信息”一词在图书馆界的广泛运用,与互联网、电子资源等资源形式的出现有关。1992年以来,国内学术界将“情报”改为“信息”,“文献情报”“文献组织”“文献著录”“文献检索”“文献计量”等也相应变为“文献信息”“信息组织”“信息描述”“信息检索”“信息计量”,“信息”开始成为图书馆界一个普遍和通用的概念,并形成广义的信息的概念。这个信息概念包括了文献,也包含了数字资源中概念的数据,还包括了管理信息时产生的数据,如元数据,等等。经过一段时间的发展,“信息”的概念延伸为“信息资源”的概念,基本成为所有文献、信息、知识的总称。从某种意义上说,广义的“信息”与“信息资源”的概念基本等同。“信息资源”的概念转化为“图书馆学系”的新名称“信息资源管理系”。也有的高校将图书馆学建成的学院从“图

^① 文庭孝,罗贤春,刘晓英,等.知识单元研究述评[J].中国图书馆学报,2011(5):75-78.

书情报学院”改名为“信息管理学院”。这个名称虽然省略了“资源”两字,但与计算学科组建的“信息学院”中的“信息”还是有区别的,还是指信息资源的管理。

“信息资源”,即广义的“信息”概念,与“信息单元”是两个不同的概念。信息单元的概念来源于邱均平的《信息计量学》一书。该书认为随着“情报”一词被“信息”所替代,情报计量学也发展成为信息计量学,同时探讨了计量单元的问题,认为“从计量单元来说,文献计量学已经不仅仅停留在篇、册、本为单位的文献单元的计量上,而开始深入到文献的内部对知识单元和文献的相关信息进行计量研究,如题名、主题词、关键词、词频、知识项、引文信息、著者、出版者、日期、语言、格式等都已成为计量的对象。”^①虽然其文中没有出现信息单元的名称,但对信息计量学的计量单元进行了罗列。文庭孝在此基础上总结出了“信息单元”的概念,并认为“文献的外形特征及其标识即信息单元,它不是文献本身,也不能有效揭示文献的知识内容,只是我们控制和处理文献的信息标识,便于我们通过信息标识来组织管理文献。”也就是将“信息单元”定义为“表达和获取文献的重要信息线索,独立存在时没有实质性的价值和意义,只有同文献单元结合使用时才能发挥出特殊的作用”^②。虽然与知识单元的研究相比,信息单元的研究沉寂得多,但信息单元也有自己的内涵和存在的价值。

与信息单元内容相近的,出现过一个“内容单元”概念。这个概念出现在马费成《科学信息离散分布规律的研究:从文献单元到内容单元的实证分析》^③系列文章中。从字面意思上讲,内容单元是与文献单元相对立的,文献单元是通过外在形式识别的单元,而内容单元主要就文献的内容而言。马费成文章中提到“在内容单元(而不是文献单元)层次上的研究基本上是空白”,说明马费成也是将两者对立看待的。但根据文中的叙述“选择文献中的关键词或主题词作为基本计量单元切入科学信息离散分布的研究,即在文献分布规律成既定的情况下,揭示其中所含的关键词或主题词(内容单元)的分布状态”来看,文中提到的内容单元,基本等同于邱均平提到的信息单元,并且马费成利用关键词和主题词进行运算的过程,也基本可以算是信息计量学的过程,故“内容单元”并不是一个新的单元内容。

1.1.3 知识单元

知识单元有广义和狭义两种,文献单元、信息单元都是广义的知识单元。本节只讨论狭义的知识单元问题。

早在1980年,我国科学计量学专家赵红洲先生就曾经提出:“知识单元是已经获得科学共同体认同的,堪称‘粒子’形态的科学概念,它是科学的细胞,科学大厦的砖块。大量的知识单元经过重组,按一定思路可凝聚成知识纤维,知识纤维又可在更高层次上组成复杂的知识系统。”^④赵红洲先生堪称我国知识单元研究的倡导者。

① 邱均平. 信息计量学(一):信息计量学的兴起和发展[J]. 情报理论与实践,2000(1):75-78.

② 文庭孝,罗贤春,刘晓英,等. 知识单元研究述评[J]. 中国图书馆学报,2011(5):75-77.

③ 马费成,陈锐. 科学信息离散分布规律的研究:从文献单元到内容单元的实证分析(1):总体研究框架[J]. 情报学报,1999(1):79-84.

④ 赵红洲. 初论“潜科学”[J]. 潜科学,1980(1):25-28.

我国图书情报学界对知识单元的探讨则略显滞后。1988年,张德芳先生撰文指出:知识单元是寓藏于出版单元(某一文献)之中的,一个出版单元可能贮存一个以上的知识单元。“科学知识单元是科学劳动的产品,是在事物、实物、现象、过程等零散信息基础上,经过系统归纳、分析、整理,能够反映其本质规律的概念、定理或定律。”^①这是一个非常有远见的论点。对于知识单元的具体运用,1995年左秀英提出建设“知识单元专题数据库”的意见,并指出知识单元“内容变化很大,它可以是一个简单的结论、一个具体的数字,也可以是一段较长的论述。”^②左秀英论文中提出的其实是从文献中提取知识单元的具体做法,并没有过多地从理论上探讨知识单元的问题,但从其论述上看,无不有后来知识单元的影子,该文可作为图书馆界对知识单元问题的一种无意识的指向,体现了我国图书馆员的思考与创新能力。

到了20世纪90年代末,图书馆界终于认识到,如何从文献单元服务深入到知识单元服务,已经是图书馆实践面临着的新的重大课题。

随后的一系列研究,如2003年王子舟在《中国图书馆学报》上发表的《知识的基本组分——文献单元和知识单元》、温有奎在《情报学报》上发表的《知识元链接理论》等重要研究出现后,知识单元及知识组织的研究进入了一个“全民时代”。

具有普遍性的知识单元概念主要认为:知识单元是知识管理、知识计量与知识评价的最小单元。狭义的知识单元目前主要用知识元来表示。知识元是指文献中相对独立的、表征知识点的一个元素,它可以是一段文字、一幅图表、一个公式、一章或一节、一段动画、一个程序等。知识元也往往直指知识产品(图书、论文、专利等)中的概念、论点(观点)、论据(数据、资料)、论证(方法、模型)、结论等知识核心和知识创新点,也可称为元知识。^③温有奎等人对知识元问题进行了大量的研究,其主要贡献在于提出了“知识链”的概念,并在《知识元链接理论》中提出:“文本内容的组织排列,是由一个个独立知识元素的逻辑排序结构。这种独立的知识元素我们称之为知识元,逻辑依存关系称之为知识链。”^④这个理论已经被广泛认可。

温有奎将知识元分成两大类10小类,分别是:①描述型,包括信息报道型、名词解释型、数值型、问题描述型和文献引证型;②过程型,包括步骤型、方法型、定义型、原理型和经验型等。^⑤此外,也有不同的分类方法,如将知识元按照知识元表达的内容差别可以分成理论与方法型知识元、事实型知识元和数值型知识元:①理论与方法型知识元包括思想、方法论、公理、原理、定律以及正在探索中的观念、观点、方法与技巧;②事实型知识元包括自然、社会存在和演变的事实信息;③数值型知识元包括各种数据类知识和科学数据,具有数值分析和知识推理功能,其中包含了大量的社会经济数据。^⑥

上述的分类方法,其实将本书中的信息与知识两个概念的内容都包括在了“知识元”这个概念中。知识元与文献单元共同组成了信息资源。文献单元以具有形式上的独立性及其具

① 张德芳. 激发和活化凝固在文献中的知识:论图书馆改革[J]. 四川图书馆学报, 1988(6):1-7.

② 左秀英. 开发文献知识信息单元的方法和途径[J]. 江苏图书馆学报, 1995(3):47-48.

③ 文庭孝. 知识单元的演变及其评价研究[J]. 图书情报工作, 2007(10):72-76.

④ 温有奎, 徐国华. 知识元链接理论[J]. 情报学报, 2003(6):665-670.

⑤ 温有奎, 温浩, 徐端颐, 等. 基于知识元的文本知识标引[J]. 情报学报, 2006(3):282-288.

⑥ 原小玲. 基于知识元的知识标引[J]. 图书馆学研究, 2007(6):45-47.