

国家社科基金项目
“知识组织系统构建与知识服务研究”成果之一

RESEARCH ON KNOWLEDGE
ORGANIZATION SYSTEMS
CONSTRUCTION AND KNOWLEDGE
SERVICES

知识组织系统构建与 知识服务研究

卜书庆 等著



國家圖書館出版社
National Library of China Publishing House

国家社科基金项目“知识组织系统构建与知识服务研究”成果之一

知识组织系统构建与 知识服务研究

卜书庆 等著

 國家圖書館出版社

图书在版编目(CIP)数据

知识组织系统构建与知识服务研究/卜书庆等著. --北京:国家图书馆出版社,2014.3
ISBN 978-7-5013-5335-4

I. ①知… II. ①卜… III. ①自动标引—自动化系统—研究报告—中国
IV. ①G356.6

中国版本图书馆 CIP 数据核字(2014)第 043776 号

书 名 知识组织系统构建与知识服务研究
著 者 卜书庆 等著
责任编辑 高爽 杨晓利



出 版 国家图书馆出版社(100034 北京市西城区文津街7号)
(原书目文献出版社 北京图书馆出版社)

发 行 010-66114536 66126153 66151313 66175620
66121706(传真),66126156(门市部)

E-mail btsfxb@nlc.gov.cn(邮购)

Website www.nlcpress.com →投稿中心

经 销 新华书店

印 装 北京科信印刷有限公司

版 次 2014年3月第1版 2014年3月第1次印刷

开 本 787×1092(毫米) 1/16

印 张 14.25

字 数 160千字

书 号 ISBN 978-7-5013-5335-4

定 价 70.00元

知识组织系统构建与知识服务研究项目组

负责人 卜书庆

成 员 （按姓氏拼音排序）

陈树年 贺玲勇 刘华梅

汪东波 薛春香 喻 菲

前 言

该研究报告为国家社科基金项目“知识组织系统构建与知识服务研究”成果之一,同时也是本项目研究成果之二“中国分类主题一体化的网络化知识组织系统”平台软件设计的依据。该研究报告从网络环境下知识组织系统的构建理论与方法、构建原则及标准、维护及服务技术和研发实践等方面进行调研论证,研究并分析知识组织系统网络化的关键理论技术及其表现形式、面向概念检索与知识导航服务的关键理论和技术、面向自动标引的关键方法和技术,并在此基础上提出“中国分类主题一体化的网络化知识组织系统”平台的设计思路、结构框架、服务手段及每个系统的实施方案。

该研究报告对国内外知识组织系统的构建及服务现状进行分析,提出在网络环境下知识组织系统的发展趋势,及国内知识组织系统网络化构建维护需要从数据层、概念层解决的问题。根据网络用户需求特点分析,提出基于知识组织系统的知识服务模式及技术手段,进一步提出基于语义网的知识服务技术。

该研究报告就网络环境下知识组织系统的构建原则、构建规范与标准的变化分析,尤其对概念系统表示的数据模型 XML、表达语义的 RDF、表示本体的 OWL 和 W3C 推荐的简单知识组织系统(SKOS)规范进行比较分析,提出基于 HTML 和语义网技术的《中国分类主题词表》(以下简称《中分表》)概念描述格式,研制出适合于中国的分类法、汉语主题词表的 SKOS 规范,以及与《中分表》MARC 的转换规范。

该研究报告对计算机网络技术在知识组织系统自动构建、自丰富维护技术的应用进行总结,提出基于《中分表》的分类主题一体化映射结构的知识组织系统的概念关系和词间关系的自动识别技术、逻辑推理技术以及智能检索技术的应用方法。

该研究报告在《中分表》映射结构体系基础上,对面向自动标引服务的分类主题一体化知识组织系统构建与需求进行研究,提出基于《中分表》用于标引服务的知识组织系统的构建方法及文献自动标引的主要技术和方法。

该研究报告最后结合每部分提出的技术方法以及标准规范,提出“中国分类主题一体化的网络化知识组织系统”平台的设计理念、实现思路、三个子系统的结构框架、服务手段及每

个系统的实施方案。

“中国分类主题一体化的网络化知识组织系统”的设计必须以实用为前提,满足广大用户当前的实际需要。因此,该报告又对未来符合语义网的数据编码和交换语言 SKOS 化后的《中分表》进行应用研究,提出需要运用语义存储与检索技术以及语义网关联数据建立原则,设计 Web Service,实现第三方 API 调用的一种主流应用思路。

总而言之,该研究报告试图以国内图书馆、情报信息系统广为使用的《中国图书馆分类法》(以下简称《中图法》)和《中分表》为基础,对知识组织系统与知识服务进行研究,构建知识组织系统的管理平台及服务平台,使知识组织系统能成功运用到信息组织与检索服务等各个环节和领域中,为我国的科学研究、信息资源整合的知识服务奠定基础。

该研究报告框架与结构由卜书庆提出并设计,经项目组全体成员讨论并确定。项目组卜书庆、刘华梅、喻菲、薛春香、贺玲勇参与该报告不同章节及附录的撰写,汪东波、陈树年提出撰写意见或素材,卜书庆负责全书的统稿和审改,刘华梅协助整理编辑。

本项目研究过程中,得到南京农业大学侯汉清、南京图书馆全勤、上海空军政治学院戴维民、南京大学苏新宁等老师的悉心指导,在研究报告撰写过程中,得到四川大学范炜老师提供的撰写意见,在此一并致以诚挚的谢意!由于时间较短,课题规模较大、任务较重,研究报告中难免存在不足,敬请各位读者批评指正。

卜书庆

2013 年 3 月于国家图书馆

目 录

前 言	(1)
1 知识组织系统构建与知识服务研究综述	(1)
1.1 引言	(1)
1.2 本项目研究的主要内容及方法	(5)
1.3 本项目的重点、难点、主要观点和创新之处	(11)
1.4 研究成果的学术价值、应用价值及不足	(12)
2 中国分类主题一体化的知识组织系统网络化研究及构建	(14)
2.1 知识组织系统网络化研究	(14)
2.2 国内知识组织系统网络化发展问题及研究	(31)
2.3 基于《中分表》的一体化创建维护系统的研发	(33)
2.4 基于《中分表》Web 版的“一体化在线更新获取信息系统”	(47)
3 基于分类主题一体化知识组织系统的网络服务研究及其在线服务系统的开发	(50)
3.1 信息用户研究与知识服务	(50)
3.2 基于分类主题一体化知识组织系统网络服务研究	(56)
3.3 基于《中分表》的分类主题一体化的知识服务系统设计及实施	(64)
4 分类主题一体化的自动标引服务系统研究及开发	(78)
4.1 文献自动标引研究	(78)
4.2 基于知识组织系统的文献自动标引	(83)
4.3 基于 CCT_KOS 的自动标引服务系统设计与实现	(88)
4.4 结语	(91)
5 基于语义网知识组织系统可视化试验及知识服务研究的展望	(94)
5.1 知识组织系统语义网络化和可视化研究	(94)
5.2 基于 SKOS 的《中分表》可视化试验研究	(96)
5.3 基于 SKOS 的知识服务展望	(103)
附录一 知识组织的受控标识系统编制规范及说明	(106)
附录二 基于分类主题一体化受控表的网络化语义描述规范及与 MARC 转换方案	(160)

1 知识组织系统构建与知识服务研究综述

知识组织系统(Knowledge Organization Systems, KOS),是指按照知识领域内的知识要素及其关联结构进行获取、描述、组织信息资源的概念系统和知识分类系统,以及对其进行控制的语义系统。知识组织系统的构建是获取、运用及其系统管理知识的关键,而基于知识组织系统的知识服务,则是知识组织系统构建的最终目标。网络环境下的知识服务包括以网络用户的知识需求为目的、以网络技术为手段,对各种知识产品或以知识类聚为主的信息资源的提供。

在数字图书馆和新一代万维网建设中,越来越需要适用于网络检索、知识导航、计算机可理解的、用于表示和组织各种信息的知识组织系统。传统的分类法、主题词表、规范文档、标题表等都是知识组织系统的核心部分,随着计算机技术和网络技术的普及与发展,传统知识组织系统的网络化及其应用研究,已经成为面向创新的图书馆知识管理和知识服务研究的重要课题。

1.1 引言

1.1.1 国内外知识组织系统构建与知识服务研究的发展特点

近年来,国内外以概念系统或知识分类系统为基础的网络知识组织系统发展迅速,其中构建方法的研究很多:在其结构形式方面,知识组织系统呈现出多元化、可视化、集成化或互操作化的发展特点;在其应用环境方面,呈现出网络化、自动化的发展特点;在组织方法与数据格式方面,呈现出标准化、网络化、兼容化等发展特点。在国内外出现基于知识组织系统的多模式的深层次资源服务、术语服务、知识学习的研究,并研制出一些相应的集成服务平台。这些发展与用户的需求、网络技术的应用分不开,下面具体分析其发展特点。

(1) 多元化与可视化

近年来,国内外针对网络信息资源以新结构方式构建 KOS 的研究很多,以概念系统或知识分类系统为基础的 NKOS(Networked Knowledge Organization System)的构建方法和应用服务趋于多元化。有面向领域概念空间架构的知识本体(Ontology)、有面向概念可视化检索的概念地图(Concept Maps)、有面向网页主题浏览检索的主题网关(Subject Information Gateway)等形式。这些形式基本是针对网络信息资源以新结构或新视觉方式创建 KOS,在应用实践中影响较大的有:2001—2006 年联合国粮农组织与泰国国立农业大学合作构建的 3 个面向渔业领域、食物安全领域、食物营养和农业领域的知识本体^①;2002 年美国国家科学数字化图书馆项目的组成部分 Get Smart 系统,是利用概念地图整合搜索工具,实现知识可视化、知识学习与管理的系统;英国资源发现网(The UK Resource Discovery Network)的组成部

^① AGROVOC[EB/OL]. [2010-05-10]. <http://www.fao.org/aims/aos.jsp>.

分——英国社会科学信息门户^①(Social Science Information Gateway, SOSIG), 是用来检索、获取、发现社会科学方面信息网页的主题网关。

总之, 这些 KOS 的范例在开展网络服务的建设初期, 基本都采用传统的分类法和主题词表, 如 AGROVOC 叙词表(联合国粮农组织与欧洲共同体委员会合作开发的用来标引和检索国际农业信息系统的多语种叙词表)、DDC(杜威十进分类法)、UDC(国际十进分类法)等, 但在应用服务过程中不断结构化、可视化和扩展兼容, 最终形成 NKOS。

(2) 集成化与互操作化

在互联网环境下, 为解决多种知识组织系统跨库、跨领域的同一主题“一站式”的信息检索和浏览, 很多传统的 KOS 开始集成化改造或构建概念标识的映射系统, 达到互检索操作的目的。早在 1981 年, 美国国会图书馆通过馆藏书目数据在 DDC 与 LCSH(美国国会图书馆标题表)和 LCC(美国国会图书馆分类法)之间建立了互换关系, 进入 2000 年后, OCLC(图书馆联机中心)已经通过 WebDewey(DDC Web 版)与 LCSH、CSH(加拿大标题表)等建立了映射系统, 启动术语服务项目(Vizine-Goetz, 2004, 2006), 开展在不同受控词表间术语映射的网络服务。2008 年, 美国国会图书馆在“书目控制的未来工作组”报告草案(Recommendations of LC WG on Future of Bibliographic Control)的答复中又提出: 要将 LCSH 与 MeSH(美国医学主题词表)、NALAT(美国农业图书馆叙词表)、AAT(艺术和建筑词表)等建立参照, 使各词表间可交叉检索和互操作^②。美国国家医学图书馆自 1986 年开始研制一体化医学语言系统(Unified Medical Language System, UMLS), 至 2007 年 11 月, 推出 UMLS 最新版本即超级叙词表, 该表集成了 100 多种词典、分类表、叙词表、专家系统等, 表达了 100 多万个概念, 包括 500 多万个概念名称^③。在德国, 由德国国家图书馆和科隆大学应用科学学院创建的基于多语种的叙词表项目即 CRISSCROSS 项目^④以及德国联邦教育与研究部资助 GESIS 社会科学信息中心的 KoMoHe 项目, 即术语映射方案^⑤等, 都是为达到语义集成的目的, 在多种 KOS 中建设映射系统和超级集成系统, 组织、创建和管理多领域的受控表之间的语词概念交叉索引问题。通过集成系统与不同的信息检索系统连接, 能够实现分布式的网络检索, 成为学科门户或主题网关的重要检索工具。

(3) 网络化与自动化

近十几年, 国内外大型文献信息或书目数据库中都有较完善的叙词表、分类表、主题规范文档或标题表及其编辑维护更新管理系统的支持, 如 DDC 电子版及支持编辑系统 ESS(Editorial Support System)在 OCLC 等数据库的应用。进入 2000 年后, 国外很多著名的叙词表、分类表、主题规范文档或标题表在电子化基础上构建适应网络服务的网络分类法、网络叙词表。如: DDC、LCC、UDC 的网络版, 即 WebDewey、Classification Web 和 UDC Online, 还有

① LC WG Rpt Response[EB/OL]. [2008-07-10]. http://www.loc.gov/bibliographic-future/news/LC-WGRptResponse_DM_053008.pdf.

② 同①

③ NLM. UMLS Metathesaurus[EB/OL]. [2007-12-22]. <http://www.nlm.nih.gov/research/umls>.

④ CRISSCROSS[EB/OL]. [2008-07-10]. <http://www.d-nb.de/eng/wir/projekte/crisscross.htm>.

⑤ KoMoHe[EB/OL]. [2008-07-12]. http://www.gesis.org/en/research/information_technology/komohe.htm.

AAT、NASA(美国航空航天局词表)的网络版等^①。它们建立了网络数据库和网络浏览器,在网络信息资源合作联机编目、服务系统中应用,对网络资源进行机助或自动分类和自动赋予主题词,提供与远程联机的本地公共检索目录系统,即 WebOPAC,并开展知识导航、知识检索服务等,如:2007 年,OCLC 的服务器开始运行的 WorldCat Local 系统,开展了为全球图书馆提供一站式发现和传递纸质和数字资源的服务,各终端图书馆通过该系统在本地图书馆开展编目、流通、资源共享以及 OpenURL 的解析互操作^②。KOS 利用网络技术、自动化技术把传统图书馆起着资源组织标准作用的分类法、叙词表继承深化和网络化拓展,以便适应于网络一站式的知识服务。

近年来,基于网络自动化系统构建与动态维护 KOS 的方法及技术越来越多,一方面是基于自组织、聚类分析的自然语言自动构建叙词表、分众分类法^③(Folksonomy)。如利用维基(WIKI)系统以大众标签为类称的信息组织方法建立的美味书签管理系统(del.icio.us)、闪亮图片夹(flickr)、维基百科全书(wikipedia)等分众分类法,这种基于多用户参与的大众分类体系通过再组织、不同层次的聚类系统建立起用于组织网络信息的分类法或词表。另一方面是基于自动丰富、自动发现技术的类表及词表的自动维护,即从文本数据中挖掘新的词汇并将它们定位到词表中,对词表进行自动更新。如 OCLC 启动的 WordSmith、Scorpion 和 FAST(主题术语的分面应用程序)等项目,主要是用来解决新主题术语在自由文档中自动发现和识别的途径以及自动分类建档的方法,达到对 DDC、LCSH 的主题自丰富问题。可见分众分类法与自动发现及识别技术的结合将是 KOS 自动丰富、自动维护的较好途径。

(4) 标准化

为适应 KOS 的网络化构建和网络化服务,进入 2000 年后,与知识组织标准规范有关的国际组织 ISO、IFLA、W3C 以及美国、英国等有关国家标准组织针对以前颁布的相关标准不断修订更新。如美国修订颁布的 ANSI/NISO Z39.19—2005《单语种受控词表编制、格式与管理规则》(Z39.19),英国修订颁布的 BS 8723——《用于信息检索的结构化词表指南》(BS8723),IFLA 修订起草的《多语种叙词表编制指南》《主题规范数据的功能要求》(FRS-AD),ISO TC46/SC4 发布的 MARC 格式的 XML 模式标准草案“ISO/DIS 25577 Information and documentation—MarcXchange”(ISO/DIS 25577)。自 2008 年,ISO TC46/SC9 陆续发布有关叙词表方面的标准草案:“ISO/DIS 25964 信息与文献—叙词表及其与其他词表的互操作(Information and documentation—Thesauri and interoperability with other vocabularies)”,其中该标准第一部分的 12—19 章是全新的内容,增补了词表在计算机网络环境下维护和应用方面的实用规范,包括概念的分面分析、叙词表的数据库表示和机读显示布局、叙词表构建和维

① WebDewey[EB/OL]. [2008-05-10]. <http://www.oclc.org/dewey/versions/webdewey/>.

Classification Web[EB/OL]. [2008-05-01]. <http://classificationweb.net>.

UDC Online[EB/OL]. [2008-05-01]. <http://www.udconline.net/>.

AAT Online[EB/OL]. [2008-05-01]. http://www.getty.edu/research/conducting_research/vocabularies/aat/.

NASA Thesaurus online[EB/OL]. [2008-05-01]. <http://mai.larc.nasa.gov>.

② 孙扬. OCLC WorldCat Local 发展综述[J]. 山东图书馆学刊, 2011(5): 46—49.

③ 分众分类法(Folksonomy)是美国信息架构专家 Thomas Vander Wal 和 Gene Smith 于 2004 年首先提出,含义是“由大众的一致意见而产生的基于用户的分类体系”。

护管理、叙词表管理软件指南、数据模型(元数据元素、关系标签等)、叙词表在检索—标引—互操作的应用集成指南、交换格式及其协议。

在信息服务方面,也有相关标准的发布,如“ISO Z39. 50”“OAI-PMH 元数据收割协议”等。制定了 XML(可扩展标记语言)、RDF(资源描述框架)和 OWLS(万维网本体语言)等影响深远的网络标准规范的万维网联盟即 W3C,在 2005 年又提出基于传统分类法、主题词表对网络信息组织、检索实际应用的语义网表示的推荐标准即简单知识组织系统(Simple Knowledge Organization Systems, SKOS),即作为受控词表和概念框架网络语义表示的推荐标准。2008 年后,又不断补充更新 SKOS 的三个主要部分:核心集(SKOS Core)、映射(SKOS Mapping)和扩展 XL(SKOS eXtension)。SKOS 发布后, NALAT、MeSH、LCSH、DDC 等词表及类表已经启用 SKOS 表示其数据元素,开展术语网络服务。

以上标准对在网络环境下受控表的构建、维护管理系统、知识服务等方面具有非常重要的现实意义,为实现分类法和主题词表网络共享和交互发展以及构建知识服务的网络技术平台奠定了共享基础,提供了知识组织系统的概念模型及其规范,满足了网络环境下的不同使用对象包括机读或机器理解的数据模型表示。

总之,这些发展趋势对解决分类法主题词表在传统图书馆长期使用中存在的问题带来生机,通过构建知识组织系统的网络实时更新平台,可以解决分类法主题词表修订周期长、更新速度慢、来源贫乏与知识和信息动态快速发展的矛盾,可以解决分类法主题词表修订与书目标引数据不能同步更新的矛盾,可以满足知识组织系统的网络共享和交互发展的需求,构建知识服务的网络基础平台。

1.1.2 国内知识组织系统构建与知识服务研究的现状及意义

国内影响最大、使用最为广泛的《中图法》《汉语主题词表》兼容互换的知识组织系统——《中分表》,在 20 世纪 90 年代中期,从印刷型工具的研制向机读型数据库建设方向发展,2001 年出版了《中图法》视窗电子版,2005 年出版了《中分表》视窗电子版。从此,国内诞生了第一部可应用于局域网服务的分类主题一体化、电子化的知识组织系统,为我国综合性信息资源的知识组织系统网络化打下了坚实的基础。这期间国内开发的其他的知识组织系统多为非网络化的、某一领域的、机助编辑的、单一的叙词表或分类表及其管理应用系统,或单一的自动标引研究试验系统,如“电子政务叙词表及管理应用系统”“军用主题词表及管理机助标引系统”“基于《中图法》知识库的自动标引和自动分类系统”等。

近年来,业界尤其以中国科学技术信息研究所为代表,依托国家社科基金项目及“十二五”国家科技支撑计划,开展了“新能源汽车”领域知识组织系统构建、汉语叙词表修订项目,“面向外文科技文献信息知识组织体系建设与应用示范”等^①研究,试图对海量数字资源抽取相关专业术语并利用机器学习、自然语言处理等技术获取词间关系,应用网络技术自动构建词表等。

国内在网络环境下的知识组织系统构建以及网络服务方面的研究成果与国外先进的知

^① 曾建勋,常春,吴雯娜,等. 网络环境下叙词表编制模式与应用方式研究(国家社科基金项目 10BTQ048)[C]//2011 年全国知识组织与知识链接学术交流会议论文集. 杭州:中国科学技术信息研究所,中国科学技术情报学会,2011.

识组织系统相比还有差距。首先,自动构建维护更新技术还不能适应网络环境的发展。其次,网络服务研究还处于起步试验阶段。再次,基于网络技术的知识组织系统实用研发较少,可创建、维护、服务的网络知识组织系统集成研究还属空白。

随着数字图书馆的快速发展和网络信息检索的需求,本项目旨在国内通用的分类法、叙词表即《中分表》网络化研究的基础上,构建知识组织系统并实现知识服务,具有重要的理论研究意义和实践应用意义:

(1)从理论和技术上,探索国内图书馆、档案馆、情报信息领域中以叙词表、分类表或兼容转换结构为主的知识组织系统网络化构建方法的突破口和关键点,推广和促进传统知识组织工具在网络环境下的发展和应用,为我国各类信息资源的知识组织、检索和利用提供一个全面的、可扩展更新的、通用的网络数字型知识组织系统,使之在我国信息资源整合、重组、共享中发挥不可替代的作用,从而为我国长期保存、利用网络资源作出应有的贡献。

(2)解决数字图书馆等数字化信息资源的自动标引及各种信息资源的知识导航、知识检索服务的实际需求。

(3)解决国内知识组织系统修订周期长、更新速度慢、来源贫乏与知识和信息动态发展之间的矛盾。

(4)解决国内知识组织系统与信息资源语义元数据不能同步更新的问题。

(5)提供国内同结构不同领域或内容的网络知识组织系统自动创建、维护、应用等技术方法。

1.2 本项目研究的主要内容及方法

本项目主要是针对分类表、主题词表及其一体化的知识组织系统的网络化构建和开展网络化服务进行研究,研发基于《中分表》的分类主题一体化的知识组织系统。该系统的研发旨在国内通用的分类法、主题词表及其一体化体系的基础上构建动态的知识组织系统,并在互联网环境下开展知识服务。它包括分类主题一体化知识组织系统的维护构建、知识服务网络(平台)系统的搭建、数字资源的自动标引试验系统的研发,以及基于语义网环境下知识组织系统的可视化研究。

网络语义工具与传统的电子化的知识组织工具相比,最大的特点之一是它具有互操作性,即不同系统能够无缝地进行通信和共享数据。这也是 Web 服务追求的终极目标。最大的特点之二是它具有计算机可理解推理的语义关系形式,这是网络语义工具自动创建、丰富维护及应用的理想条件。所以,网络语义工具是基于 Web 服务的自动创建、维护和用于标引、浏览、检索一体的和开放式的语义工具。因此,就我国知识组织系统的发展和使用现状,如何实现国内知识组织系统由电子化向网络化的转变,需通过研发网络化系统解决如下问题:

(1)在《中图法》《中分表》等使用广泛的传统知识组织工具电子化成果的基础上,为我国综合性信息资源的组织、检索和利用提供一个全面、即时、通用的网络数字型知识组织工具的交换和应用的技术平台,使之在我国信息资源整合、重组、共享中发挥不可替代的作用。

(2)从网络互操作技术出发,应在 MARC 与 XML、RDF、SKOS、OWL 等语义网相关标准

的基础上,建立可互操作和兼容的知识组织工具的网络数据库,从网络层、数据层、语言标识层刻画数据的使用与表述模式,便于计算机的理解和逻辑推理。

(3)在网络数据库和网络技术平台基础上,建立基于网络环境的、分类主题一体化的、通用的知识组织工具的编辑、维护、应用集成系统,包括自动创建编辑、逻辑校验系统、自动丰富维护系统、网络资源的自动标引系统、网络知识导航及概念检索系统等,发布《中分表》的Web版,对外开展知识服务。

(4)通过图书馆长期积累形成的大型文献数据库的语义元数据和网络信息资源主题术语的自动统计分析,利用计算机技术、自然语言处理技术等方法,开展自动标引服务。进一步改造传统知识组织工具的精细分类体系,确定网络知识组织工具的粗放多面的概念体系,细化叙词的术语逻辑体系以及与自然语言的转换方法和技术,结合智能标引和智能检索的技术来实现网络知识组织工具自丰富、自维护及可视化的功能。

1.2.1 本项目研究的基本思路和方法

针对以上问题进行专题调研,对国内外的该专题文献、研发项目进行调研分析,总结其解决问题的办法及研究方法,并结合先期我们对中国知识组织工具的研制成果,借鉴国内外相关项目的实施经验,确定了以下研究思路和方法:

首先,在《中分表》数字化(电子化)基础上,对分类表主题词表一体化的知识组织系统网络化进行研究,包括以下研究论题。

(1)知识组织系统网络化的条件研究,包括网络环境下词表、类表构建原则及标准;数据存储、共享交换方式及格式,从标识层、数据层、网络层刻画知识组织系统的表述模式,对MARC与XML、RDF、OWL、SKOS之间的数据表示和应用特点进行研究,形成MARC与XML转换表、MARC与SKOS转换表;网络环境下知识组织系统的构建方法及自动构建与自丰富技术,确定《中分表》网络化系统构建应采取的措施。

(2)基于知识组织系统网络化服务研究,包括网络用户需求特点分析、基于语义网的知识服务特点分析、基于知识组织系统的网络服务的服务模式及技术研究,确定《中分表》Web版的服务功能。

其次,在《中分表》映射结构体系基础上,对面向自动标引服务的分类主题一体化知识组织系统构建与需求研究,包括以下论题。

(1)文献自动标引研究,包括文献自动标引原理、文献自动标引主要技术和方法、基于知识组织系统的文献自动标引的理论基础与数据基础。

(2)面向一体化标引服务的知识组织系统的需求研究,包括构建结构、构建方法、基于《中分表》用于标引服务的知识组织系统的构建研究。

再次,在以上研究基础上,设计出“中国分类主题一体化的网络化知识组织系统”的架构及其软件系统的研发。“中国分类主题一体化的网络化知识组织系统”架构应包括以下三部分(见图1-1),每部分既相互联系、相互补充,又可独立调用,分别完成不同环境下的任务。

(1)基于《中分表》的一体化创建维护系统,包括基于《中分表》Web环境的一体化在线更新获取信息系统。该系统主要完成局域网与互联网等不同环境下的知识组织系统的构建维护及自丰富任务。

(2)基于《中分表》的一体化网络知识服务系统,即《中分表》Web版。该系统主要完成

互联网环境下知识及相关知识学习、辅助标引与主题文献跨库检索及导航服务任务。

(3) 基于《中分表》知识组织系统的自动标引服务系统。该系统主要完成数字化文献的自动分类、主题词自动标引的任务。

最后,在以上实用系统、互联网发布的服务系统和自动标引试验系统的开发基础上,通过用户使用和测试,最大可能地改善和优化系统功能,满足国内用户的实际需要,同时对未来语义网环境下,中国知识组织系统需要语义化和可视化服务功能的需求进一步实验研究,形成基于语义网环境下的知识组织系统可视化试验及知识服务研究的专题报告。

该专题研究当图书馆资源成为新兴语义网的一部分时,知识组织系统的语义化、网络化的关键点和关键步骤。它不仅需要该项目已完成的知识组织系统 SKOS 转换输出,使其符合语义网基本的数据编码和交换语言以及资源语义描述框架,而且需要运用语义存储与检索技术,遵守互联网关联数据建立原则,设计 Web Service,实现第三方 API 调用的一种主流应用思路。该专题就面向用户应用需求的两个方面:主题概念的 HTML 呈现和用户知识检索发现可视化方面进行试验研究。以《中分表》SKOS 化后为例进行应用试验,总结实现以上两个方面的需求时,在技术上需要的关键步骤和功能模块,并对将来可以提供的更多的应用服务,如术语服务、关联数据服务、互操作服务,进行展望和研究。

1.2.2 本项目研究成果及特点

本项目研究成果包括两种类型:软件工具及研究报告。

研究报告即“知识组织系统构建与知识服务研究报告”,其主要研究内容已在“1.2.1”小节中列出,各专题详细内容见本报告第 2、3、4 章节。该报告特点在于它从理论、技术和研发实践三方面对基于分类表、主题词表及一体化表的知识组织系统的网络化构建、新技术环境下用户需求的变化、知识服务包括自动标引服务等问题系统研究,对“中国分类主题一体化的网络化知识组织系统”平台设计思路进行总结,对平台结构功能及应用情况进行分析,充分说明了该项目研究是成功的,研究成果达到了预期研究目标。

软件工具即“中国分类主题一体化的网络化知识组织系统”平台正是在“研究报告”中各专题研究的基础上,开发完成的实用系统和试验系统,通过这些系统的应用价值分析,又充分证明了本项目研究的科学性和正确性。该平台包括以下三个相互关联的软件系统:①基于《中分表》的一体化创建维护系统;②基于《中分表》的一体化网络知识服务系统(<http://cct.nlc.gov.cn/login.aspx>);③基于《中分表》知识组织系统的自动标引服务系统。见图 1-1,分别说明。

1.2.2.1 基于《中分表》的一体化创建维护系统

基于《中分表》的一体化创建维护系统是“中国分类主题一体化的网络化知识组织系统”的基础平台及核心部分。该系统是以《中分表》数据为基础,适合简单和复杂结构的知识组织系统的创建、机助维护和发布的网络管理系统。除了全面支持《中分表》,还支持与它结构相同(或类似)的分类法、叙词表和名称规范的概念系统的创建与维护,包括分类法与叙词表系统之间的相互映射系统的同步创建与维护。该系统可满足类目(类号与类名等)、叙词、语词(术语、关键词等)等概念单元的自动或半自动的增删改、查重,支持概念逻辑关系的自动校验,并能根据系统内部设定的逻辑关系规则进行自动推理生成相应的逻辑关系。

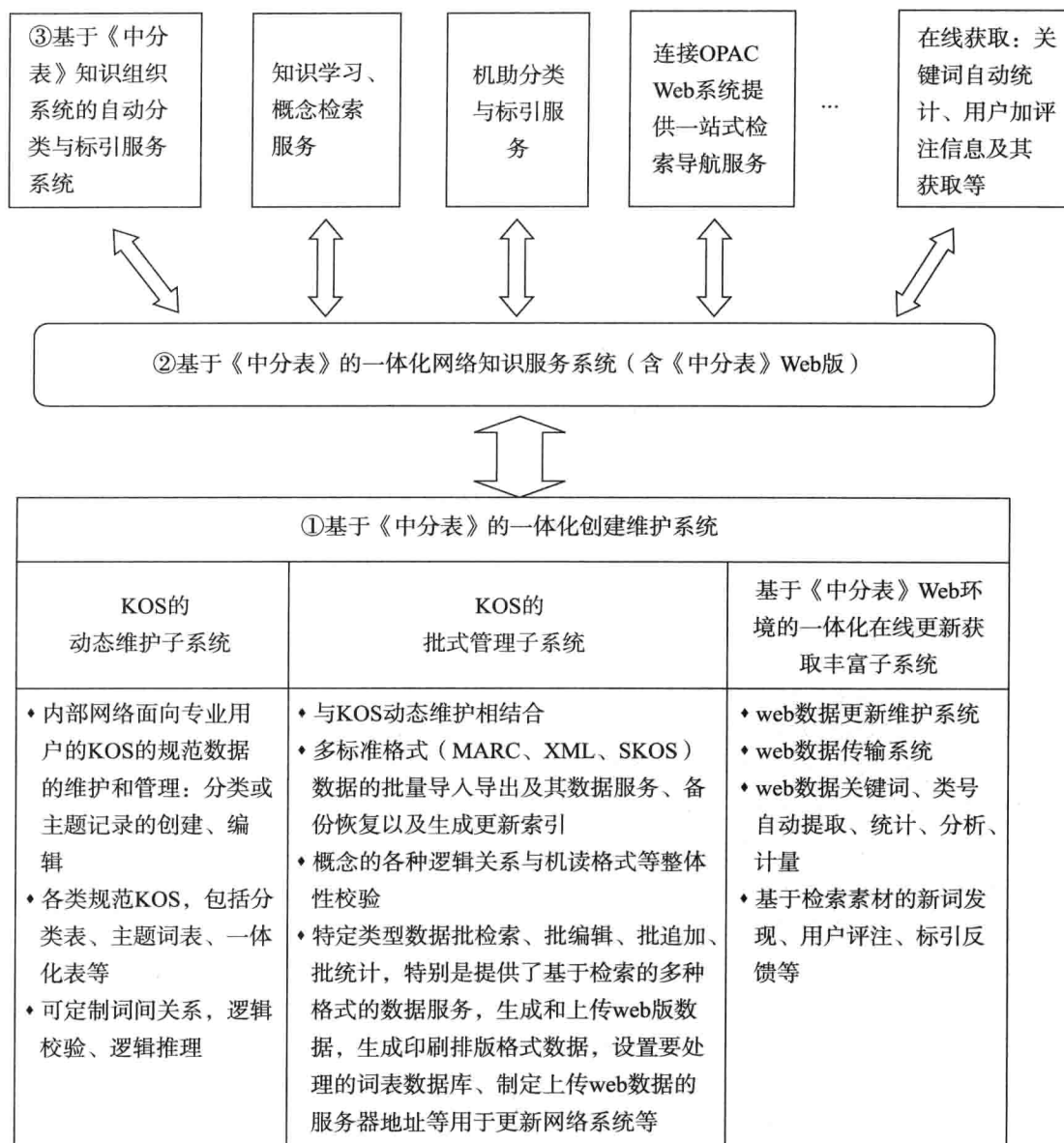


图 1-1 中国分类主题一体化的网络化知识组织系统平台

该系统包括三个子系统：动态维护子系统、批式管理子系统、Web 环境的一体化在线更新获取丰富子系统。三个子系统既相互关联、相互补充，又可相互独立调用，分别解决不同时间、不同类型、不同规律、不同权限和不同应用环境下的复杂数据网的管理问题。

动态维护子系统用于分类表、主题表以及一体化表的编制及维护，主要功能包括：在局域网环境下分类或主题记录的创建、编辑，类目概念间逻辑关系、主题词概念间逻辑关系、类目和主题概念间映射关系的自动获取和动态建立以及逻辑关系的自动校验、自动推理等。

批式管理子系统用于分类表、主题表规律性问题的批处理和批统计功能需求，主要功能包括：多标准格式（MARC、XML、SKOS）数据的批量导入导出、备份恢复以及生成更新索引，

各种概念逻辑关系与机读格式的整体性校验,特定类型数据批检索、批编辑、批追加、批统计,特别是提供了基于检索的多种格式的数据服务,生成和上传 Web 版数据,生成印刷排版格式数据,设置定义被处理的数据库,以及上传数据的 Web 服务器等用于更新网络系统等。

Web 环境的一体化在线更新获取丰富子系统用于通过互联网技术为《中图法》《中分表》等知识组织系统服务平台实时更新数据的系统。该系统需要保障分类法、主题词表以及一体化表的数据实时传输于 Web 环境中并安全更新,即时获取 Web 用户使用信息、修订反馈信息并自动统计,有条件地提供与文献分类主题同步更新的元数据服务。因此,该系统主要功能包括 Web 数据更新维护、Web 数据传输、用户评注和检索词统计等几部分内容。一方面,利用 Web 更新系统实时更新数据并安全传输至 Web 服务器上;另一方面,利用用户评注系统和检索词匹配统计系统发现新概念、新词及修订信息,进而用于更新《中分表》Web 数据、《中图法》Web 数据。

本系统的分类主题一体化管理方法和设计理念适用于中国知识组织工具实际应用的现状,并为国内外首创。本系统为我国综合性的分类主题一体化的知识组织系统创建维护搭建了可行及可应用的平台,为专业性及个性化服务的知识组织系统构建维护奠定了坚实的基础。本系统完成了知识组织系统构建的预期研究计划,达到了预期研发目标,有以下特点:

(1)先进的开发设计理念。该系统从面向过程到面向对象,再到面向服务的信息系统设计理念,设计出三个子系统,完成知识组织系统创建维护过程中各阶段的不同任务。在动态系统中,面向每个概念单元(标识层)实施编制规则;在批处理系统中,面向概念体系即数据库进行管理与功能增值;在 Web 获取与自丰富系统中,面向开放服务(网络层)的用户获取使用信息及丰富维护系统;最后通过有机合成,共同完成系统创建维护任务。

(2)满足复杂和简单的知识组织系统类型的一体化构建与管理。既满足专业用户对分类法、主题词表等不同结构类型的一体化逻辑编辑、自动校验、自动推理的机助动态构建与批式维护,又满足了互联网环境下社会用户的参与以及多种规范数据(MARC、XML、SKOS)的国际兼容互换,为知识服务、数据服务、再构知识组织工具奠定了实时更新数据的基础。

(3)与现代信息处理的先进技术结合,研发各种实用手段。一体化类表词表维护管理系统最理想的工作模式应该是在国际互联网支持下的多点协同工作。出于一体化类表词表维护管理系统的复杂性和安全性、智能性与服务的稳定性等多种因素的考虑,现阶段将系统建立在局域网环境下,批更新于网络环境中,在逐渐应用完善的前提下,移植到国际互联网。

本系统利用现代计算机可视化的强大功能,通过一体化类表词表丰富的语义(上下文)环境确定其表现形式,利用多窗体操作界面 MDI(Multiple Document Interface),非常好地映射类表词表各个语义记录本身及其相互关系,且能非常方便地操控这些记录之间的关系,符合人工创建词表类表的编制思路与编制模式。利用超文本技术,可十分准确的表达分类和主题概念之间的各种互见关系,符合人们联想思维方式。利用文本控件字体颜色,可以清晰表示新增类目、修改类目,停用类目等不同性质、状态的类目等。

(4)该系统为实践验证的实用可行系统。《中分表》是目前国内使用最广泛的、适合各种信息资源的、分类主题一体化的知识组织工具,研究以其为原型的构建维护系统,基本可以满足国内各类型分类表、词表、规范文档等知识组织工具的需求。该系统是以国内第一部分类主题一体化的电子化知识组织工具——《中分表》电子版为基础,构建适合多结构、多资

源类型、多种用户应用的知识组织系统。该系统研发过程即为知识组织工具“生产”过程,构建完成了《中图法》第五版(数据库、印刷版)、《中分表》Web 服务版、《中图法》Web 服务版、《中图法·简本》第五版、《中图法·期刊分类表》第三版等知识组织工具。因此,该系统是在生产不同结构与功能的知识组织工具过程中逐渐完善优化,并证明它可行及可推广应用。该系统能满足可更新扩展的实用分类体系的建设需求,很方便地推出新的行业或专业知识组织系统的构建平台,同时能保证与国外有影响力的分类法主题词表建立映射关联的机制,从而可进一步形成一个有影响力的知识本体。通过特殊授权可下载数据,为网络用户、专业用户、多语种用户服务。

1.2.2.2 基于《中分表》的一体化网络知识服务系统

广义讲,基于《中分表》的一体化网络知识服务系统是“中国分类主题一体化的网络化知识组织系统”中面向专业用户与社会用户以及机器用户的服务平台。该平台面向机器用户开展数字资源及网络资源的自动标引、自动分类系统是相对独立的系统,是以本系统数据和用户反馈信息为基础构建机器用的知识库,通过匹配运算完成资源内容的自动标引任务,即“基于《中分表》知识组织系统的自动标引服务系统”,为明确系统的用户任务,该系统详细特点见 1.2.2.3 小节。

狭义讲,基于《中分表》的一体化网络知识服务系统就是在一体化创建维护系统支持下,重点为最终用户包括专业用户与社会用户开展互联网环境下的知识服务的平台。通过本平台(<http://cct.nlc.gov.cn/login.aspx>)完成用户的知识学习、相关概念检索任务;系统辅助专业用户开展文献分类与主题词标引服务;系统提供多个 Web OPAC 连接,面向社会用户提供各种主题知识的文献检索服务,包括全国各类型图书馆收藏的相关主题、专指主题、学科或专业范畴主题的各类型文献的一站式检索,实现学科文献导航服务,学科文献馆藏或门户建设,扩大或缩小相关主题文献的可视化检索。面向社会用户开展关键词检索的自动统计,开展评注标签服务,自动获取用户使用信息,实现实时更新数据服务等任务。

本系统是针对中国现行知识组织工具维护与服务过程中存在的实际问题,提出网络环境下为“最终用户”服务以及社会参与维护知识组织工具的先进理念,结合国外分类法主题词表网络服务技术而设计出基于《中分表》的中国分类主题一体化网络知识服务系统,通过采取的必要手段和技术解决了以下问题,达到了预期研究目标。

(1)通过互联网技术集成多个 OPAC 检索系统,满足用户一站式的检索需求,解决了我国使用《中图法》《中分表》《汉语主题词表》组织的各种信息资源的知识分类导航、知识检索服务的实际需求,提供了一个集成式的图书馆信息资源的知识门户,使用户通过一个互联网平台界面,能够从知识角度、学科角度跨库浏览和检索分布在我国图书馆各主题网关内的信息资源,同时也有助于新知识的学习和发现,帮助实现知识获取、知识发现。

(2)通过互联网技术开展实时更新《中图法》《中分表》《汉语主题词表》等数据服务,解决了国内知识组织系统修订周期长、更新速度慢、来源贫乏与知识和信息动态发展之间的矛盾,不仅同时解决了在 OPAC 系统内嵌入知识组织系统不能同步更新的知识导航方式,而且通过集成方式解决了国内知识组织系统与信息资源语义元数据不能同步更新的问题。

(3)通过互联网技术为《中图法》《中分表》《汉语主题词表》的广大用户(包括图书馆业界和读者)提供各类信息资源的知识组织、内容标引、知识检索、学科导航和实时更新的一个通用的网络数字型的知识共享服务平台及多知识组织系统映射关联的机制,为不同的词表