

国家数字图书馆工程标准规范成果

网络环境下的知识 组织规范和应用指南

王 军 卜书庆 主编



國家圖書館出版社

网络环境下的知识 组织规范和应用指南

王 军 卜书庆 主编



圖 國家圖書館出版社

图书在版编目(CIP)数据

网络环境下的知识组织规范和应用指南/王军,卜书庆主编. —北京:国家图书馆出版社, 2012.9

(国家数字图书馆工程标准规范成果)

ISBN 978-7-5013-4848-0

I. ①网… II. ①王… ②卜… III. ①数字图书馆—知识管理—规范—指南 IV. ①G250.76-65

中国版本图书馆 CIP 数据核字 (2012) 第 189285 号

责任编辑:高爽

书名 网络环境下的知识组织规范和应用指南

著者 王 军 卜书庆 主编

出版 国家图书馆出版社(100034 北京市西城区文津街7号)

(原北京图书馆出版社)

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

印刷 北京佳顺印务有限公司

开本 787×1092(毫米) 1/16

印张 16.75

版次 2012年9月第1版 2012年9月第1次印刷

字数 300(千字)

书号 ISBN 978-7-5013-4848-0

定价 58.00 元

丛书编委会

主 编：国家图书馆

编委会：

主 任：周和平

执行副主任：詹福瑞

副主任：陈 力 魏大威

成 员（按姓氏拼音排名）：卜书庆 贺 燕 蒋宇弘

梁蕙玮 龙 伟 吕淑萍 申晓娟 苏品红

汪东波 王文玲 王 洋 杨东波 翟喜奎

赵 悦 周 晨

本书编委会

主 编：王 军 卜书庆

副主编：欧石燕 曾新红

编 委：喻 菲 刘华梅 郝嘉树 廖永霞 赵 悦 王广平

总 序

数字图书馆涵盖多个分布式、超大规模、可互操作的异构多媒体资源库群,面向社会公众提供全方位的知识服务。它既是知识网络,又是知识中心,同时也是一套完整的知识定位系统,并将成为未来社会公共信息的中心和枢纽。数字图书馆建设的最终目标是实现对人类知识的普遍存取,使任何群体、任何个人都能与人类知识宝库近在咫尺,随时随地从中受益,从而最终消除人们在信息获取方面的不平等。“国家图书馆二期工程暨国家数字图书馆工程”是国家“十五”重点文化建设项目,由国家图书馆主持建设,其中国家数字图书馆工程的建设内容主要包括硬件基础平台、数字图书馆应用系统和数字图书馆标准规范体系。

标准规范作为数字图书馆建设的基础,是开发利用与共建共享资源的基本保障,是保证数字图书馆的资源和服务在整个数字信息环境中可利用、可互操作和可持续发展的基础。因此,在数字图书馆建设中,应坚持标准规范建设先行的原则。国家数字图书馆标准规范体系建设围绕数字资源生命周期为主线进行构建,涉及数字图书馆建设过程中所需要的主要标准,涵盖数字内容创建、数字对象描述、数字资源组织管理、数字资源服务、数字资源长期保存五个环节,共计三十余项标准。

在国家数字图书馆标准规范建设中,国家图书馆本着合作、开放、共建的原则,引入有相关标准研制及实施经验的文献信息机构、科研机构以及企业单位承担标准规范的研制工作,这就使得国家数字图书馆标准规范的研制能够充分依托国家图书馆及各研制单位数字图书馆建设的实践与研究,使国家数字图书馆的标准规范成果具有广泛的开放性与适用性。本次出版的系列成果均经过国家图书馆验收、网上公开质询以及业界专家验收等多个验收环节,确保了标准规范成果的科学性及实用性。

目前,国内数字图书馆标准规范尚处于研究与探索性应用阶段,国家图书馆担负的职责与任务决定了我们在数字图书馆标准规范建设方面具有的责任。此次将

国家数字图书馆工程标准规范研制成果付梓出版,将为其他图书馆、数字图书馆建设及相关行业数字资源建设与服务提供建设规范依据,对于推广国家数字图书馆建设成果,提高我国数字图书馆建设标准化水平,促进数字资源与服务的共建共享具有重要意义。

国家图书馆馆长 周和平
2010年8月

前言

1 项目缘起

“国家图书馆知识组织标准规范”是中国国家图书馆于 2009 年启动的一项规范建设项目,它属于国家数字图书馆工程标准规范建设的主导项目之一(项目编号 D009)。国家图书馆的项目组经过一年多的调研、策划分析,于 2009 年 8 月制定了项目需求书,提出项目总体目标:基于国内知识组织工具的特点和发展需要,参考国际上网络知识组织系统的现有成果和发展趋势,研制国家图书馆网络知识组织系统的构建和应用相关规范,以提升国家图书馆数字资源的组织与整合能力,并为国内相关团体和个人构建各类网络化的知识组织工具提供参考。进一步,为国内网络知识组织系统提供一个共建共享平台,推动网络环境下国内知识组织系统的发展和应用水平。

知识组织是在传统文献信息环境下发展起来的信息组织与利用手段,在百余年的应用过程中,形成和完善了分类法、主题词表等知识组织工具。在今天的网络信息环境下,知识组织工具需要进一步的发展和创新,以适应网络化的信息获取手段,满足数字化信息资源组织的要求。随着数字图书馆的建设和发展,网络知识组织系统(NKOS)已成为图书馆和信息科学领域最重要的研究课题之一。网络知识组织系统是指在网络环境下可应用的,能够用于表示和组织数字信息资源的,不仅可以为人所利用而且还能为机器提供服务的知识组织工具。NKOS 可以分为两类:一类是从传统的知识组织工具发展演变而来但需要网络化,例如语义化描述的分类表、叙词表、主题词表、地名辞典等;另一类是在网络环境下产生和发展起来的,如本体、分众分类法等。前者在国际图书情报界已经有较多的研究和实践,如美国国会图书馆已发布了《国会图书馆主题词表》的 SKOS 语义化描述版本,OCLC 建立了术语服务平台向各类应用程序提供基于多种词表的词汇服务,美国自然科学数字图书馆(NSDL)开发了元数据注册平台来支持各类词表的发现和重用,联合国粮农组织将其管理的词表 AGROVOC 向本体的形式转换,欧盟在多语言词表和词表的互操作方面进行了多年的实践。后者,即本体和分众分类法,是信息科学和计算机科学共同关注的研究领域。本体方面,最新的研究和应用热点是关联数据,W3C 已经发布了《图书馆关联数据应用指南》(Library Linked Data Incubator Group Final Report);分众分类法方面,国外一些大学图书馆已在 OPAC 系统中嵌入了标签工具(如滨州大学、密歇根大学等),一些图书馆集成系统厂商提供了分众分类法的扩展模块(如 Aleph、Voyager、Millenium 等)。

与国际 NKOS 领域发展相呼应,国内 NKOS 的研究与应用渐成热潮。此时,国家图书馆启动 NKOS 构建规范的项目适时适机。本项目在汇总国际上 NKOS 构建方法、技术和经验的基

基础上,结合国内知识组织工具的特点和国内数字图书馆的应用环境的要求,开发制定了构建我国 NKOS 的参考规范和应用指南。一方面,为国家图书馆实施相关项目提供指导;另一方面,为国内相关机构研究和开发 NKOS 提供参考依据。

2 项目的目标和要求

随着数字资源的增长,对数字资源进行有效表示和管理的需要越来越强烈。作为实施数字资源表示、组织和管理的基础,急需加快国内 NKOS 的建设步伐。为此,国家图书馆启动了网络环境下的知识组织标准规范项目。

项目目标:

(1)传统的知识组织工具是提供给人使用,而 NKOS 不仅要便于专业人员描述与组织数字资源,而且要为网络应用程序(如 OPAC 系统、搜索引擎等)提供术语服务,以实现基于内容的智能信息服务,如概念检索、自动分类、学科导航等。

(2)使得国家图书馆具备和国外相关机构进行知识组织工具互操作的能力,具备发布关联数据、进行语义数据整合的能力。

(3)为传统知识组织工具在网络环境下的改造、发展和完善提供建议和参考,推动 NKOS 的创新和发展。

(4)NKOS 是网络信息资源利用的重要手段。目前的网络信息利用主要依赖搜索引擎。为了更充分地发现与利用网络信息资源,需要知识组织手段的有力补充。将图书馆的知识组织工具转化为 Web 环境下普遍可获取的知识资源,是本项目的最终目标。

项目的设计要求:

(1)参考国际上相关的研究成果,与国际上重要的网络知识组织系统保持兼容性和互操作性,符合数字图书馆未来的发展趋势

NKOS 是图书馆学和信息科学目前的重点和热点研究领域,已有一些成果可以借鉴。如 W3C 公布了基于 RDF 的简单知识组织系统描述语言(SKOS),美国国会图书馆提供了 LCSH(美国国会图书馆主题词表)SKOS 版本的开放下载, OCLC 尝试建立术语服务平台。本项目应当充分借鉴已有的研究成果和实际应用,应和相关的国际规范保持一致,应与国际上主要 NKOS 兼容或互操作,在未来的发展中保持稳定性与可持续性。

(2)满足国家数字图书馆工程建设的需要,具有实用性和可操作性,为国内 NKOS 的建设提供参考和借鉴

本项目所研制的各种规范,一方面要靠拢国际上 NKOS 的构建规范,另一方面,要符合中文数字资源的特点,适应国内数字图书馆的环境。应具备实用性和可操作性,切实服务于国家图书馆组织整合各类数字资源、展开各种智能应用的需要。应覆盖主要的 NKOS 种类,为国内 NKOS 的建设起到指导作用。

(3)具有灵活性、可扩展性和前瞻性,为未来的发展和完善预留空间

NKOS 是一个尚处于发展阶段的研究领域,相关的标准和技术解决方案还没有进入大规模应用的阶段。在此情况下制定的中国 NKOS 构建规范,有些内容还属于探索研究范畴,具有一定的前瞻性,应根据现实的应用情况修改和完善。另外,知识组织工具现实的应用情况也十分复杂,目前这套规范主要是针对通用型、综合性的知识组织工具而制定的,未必完全适用于特定领域和特定环境下的知识组织要求。这就要求本项目所提出的规范和解决方案,具有灵活性和可扩充性,为未来的扩展和定制预留空间。

3 项目规划

1. 设计思想

NKOS 包括的类型很广泛,从术语表、分类表、词表到语义网络和本体。考虑到:①国家图书馆目前建设需要,②现有知识组织工具的基础,③目前国内外 NKOS 的发展水平、特别是 NKOS 的应用水平。本项目选择了受控表(包括分类法和主题词表)、本体和大众分类法这三种知识组织系统作为规范制定的重点,这些工具都是目前相对较成熟的、在数字图书馆中有较大应用潜力的 NKOS 类型。这些 NKOS 类型,代表了当前 NKOS 的研究和应用要解决的三大核心问题。

(1)如何基于传统 KOS 构建 NKOS:NKOS 是知识组织传统在网络环境下的延续和发展。NKOS 构建的一个重要途径是对传统 KOS 进行改造和语义化转换,尤其是对于规范档、分类表、词表这类受控表。在本项目中,以《中国分类主题词表》为主要应用对象,制定了“受控表语义描述规范”。

(2)如何基于 NKOS 向机器和各类智能应用提供服务:构建 NKOS 的主要目标是为搜索引擎、学科门户、内容导航、自动分类等应用程序提供知识查询服务。途径是提供开放的、统一的查询接口和访问协议,例如,SKOS API^①、ADL 词表查询协议^②等。参考已有的 NKOS 术语服务机制,本项目设计三类 SKOS 的术语服务包括基于 NKOS 元数据模型的术语服务、基于 SKOS 模型的术语服务和基于 RDFS/OWL 的术语服务。

(3)如何基于 NKOS 向最终用户提供服务:NKOS 要解决的另一个关键问题是如何跨越“用户鸿沟”。长久以来,各类 KOS 都是图书馆馆员等专业人员在后台使用的工具。如何让普通用户方便、有效地利用 NKOS 是 NKOS 应用和发展的生命力所在。本项目创造性地提出将大众分类法实现为 NKOS 的词汇服务渠道,在用户检索、标注等过程得以直接访问 NKOS。

① Simon Jupp, Sean Bechhofer, Robert Stevens. A Flexible API and Editor for SKOS. 6th Annual European Semantic Web Conference (ESWC2009), 2009: 506 – 520

② Greg Janec. ADL Thesaurus Protocol. <http://www.alexandria.ucsb.edu/thesaurus/>

项目的整体设计思想如图 1 所示。

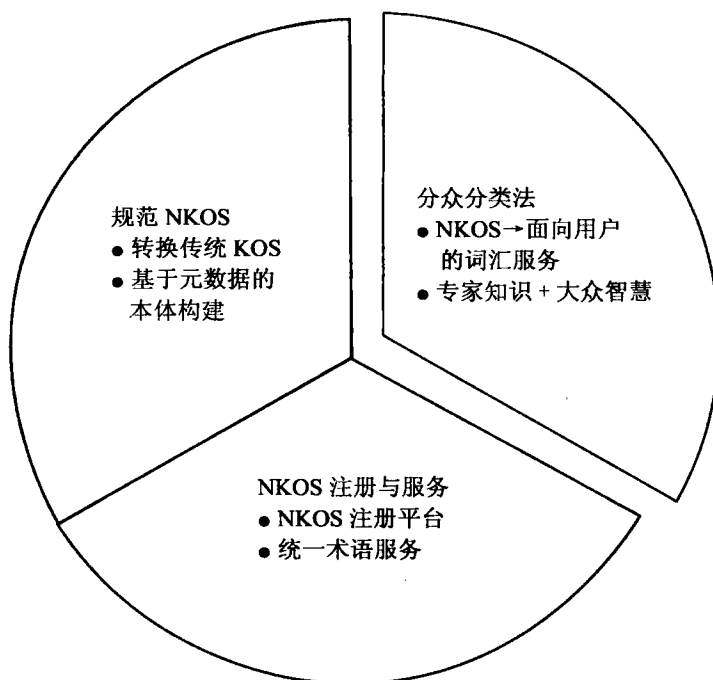


图 1 知识组织规范项目设计思想

首先是受控表、本体这一类规范 NKOS 的构建。其中包括两个子项：一是用 SKOS 来语义化描述分类法、词表等受控表系统，对于这些在文献信息环境下发展起来的传统知识组织工具，要全面描述它们的复杂结构和内容，就需要对 SKOS 进行本地扩展（例如分类表中的附表结构、交替类目、组配关系等），同时，要对其不适应网络环境和数字资源的方面加以变通调整。二是基于元数据来构建面向数字馆藏的元数据本体。其目的是基于所建的本体，将数字图书馆内各种不同格式、不同类型的元数据统一转换为 RDF 语义元数据，为关联数据发布奠定基础。

其次，是面向数字图书馆的分众分类法的设计。分众分类法是 Web 环境下发展起来的社会化自由标注工具，它具有社会化、自由化、扁平化（即没有等级结构控制）的特点。数字图书馆的应用环境提出了以下两点要求：①不仅要符合普通网民自由标注的习惯，同时要满足图书馆专业人员组织数字图书馆正规资源的要求；②作为数字图书馆的知识组织工具，要能够和已有的规范 NKOS 无缝整合。为了满足这样的要求，本项目提出了以下解决方案：①用户标注网络信息资源时，基于后台的规范 NKOS（如词表）向用户推荐合适的受控词汇。②上述的词汇

推荐服务,不仅推荐合适的标引词,而且推荐词间关系,辅助用户建立等级结构。这样就在大众分类法中引入等级结构以克服分众分类法结构扁平的缺陷。③在后台从分众分类法中提取新鲜词汇用来丰富后台的规范词汇,实现专家知识和大众智慧的双向融合。这样,一方面,将 NKOS 转化为了方便用户的词汇服务,解决了“用户鸿沟”的问题;另一方面,实现了专家知识和大众智慧的双向融合,解决 NKOS 词汇更新瓶颈。

最后,本项目提出了一个中文 NKOS 的注册和管理平台的设计方案,实现网络环境下类表、词表、规范文档、本体等各类 NKOS 的集中存储、发布、管理和发现,以促进不同机构间的 NKOS 的统一管理、共享和共建。这部分有三个重点:①定义描述各类 NKOS 的元数据标准;②描述 NKOS 平台的三大功能:注册、发现和服务;③定义统一术语服务及访问协议。目标是将各类网络知识组织系统统一在 NKOS 注册平台之下,通过术语服务机制向各类应用提供统一服务。

这三部分形成一个有机的整体。规范 NKOS 与分众分类法双向集成,融合专家知识和大众智慧,通过 NKOS 注册和术语服务将知识组织系统——图书馆积累了上百年的知识资源——转化为用户触手可及的服务和网络应用实现知识服务的可以依赖的知识源。

2. 规范构成

根据以上设计思想,本项目共规划以下六份规范和设计报告书。

(1) 网络知识组织系统注册与术语服务规范

本规范的目的是为国家图书馆建设网络知识组织系统注册平台提供设计参考。NKOS 注册平台是一个集中登记和管理 NKOS 的系统,它为用户提供了一个发现 NKOS 的场所,为开发者提供了一个共建、共享和发布 NKOS 的平台。术语服务定义了应用软件访问 NKOS 平台中的任意知识组织系统的标准接口、通信协议和操作原语。

(2) 受控表语义描述规范

应用 W3C 提出的基于 RDF 的简单知识组织系统描述语言 SKOS 来表示图书分类法、主题词表等受控词表。要求该规范具备描述国内常见受控词表的能力,并针对国内受控词表的特点可对 SKOS 进行适当扩展。

(3) “中国分类主题词表”语义描述规范与应用指南

《中国分类主题词表》(CCT)是目前国内影响最大、使用最广泛、最权威的分类法主题法一

体化的知识组织工具。该规范要制定采用 SKOS 对 CCT 进行语义描述的方法和规则。可根据需要对 SKOS 进行扩展,并对 CCT 未来的发展提出调整或改造的建议。

(4) 基于元数据的本体构建与应用规范

根据国家图书馆的核心元数据方案,来设计和构造面向数字馆藏的元数据本体。其目的是基于构建的元数据本体,将数字图书馆内各种不同格式、不同类型的元数据统一转换为 RDF 语义元数据,为关联数据发布奠定基础。

(5) 面向数字图书馆环境的分众分类法需求分析和功能设计书

分众分类法是在 Web2.0 环境下出现的一种大众化、社会化的简单信息组织工具。在数字图书馆的环境下应用分众分类法,要考虑到数字馆藏不同于网页资源的特点和图书馆环境的特殊需求,这就要求对根植于 Web 环境的分众分类法系统进行改造。考虑到分众分类法系统和应用环境紧密结合的特点,本报告针对国家图书馆的特定需求而起草了一个社会书签系统的需求分析和功能设计书。

为了更好地服务国家图书馆的建设,在规范报告之外还撰写了在数字图书馆内部署这些知识组织系统的应用指南,并对国内外相关研究和应用现状进行了调研,撰写了综述报告。最终项目成果共包含 11 份文本:

- D009—0 “国家图书馆知识组织规范”总体设计
- D009—1 网络知识组织系统注册与术语服务规范
- D009—2 网络知识组织系统注册与术语服务规范应用指南
- D009—3 受控表语义描述规范
- D009—4 “中国分类主题词表”语义描述规范及应用指南
- D009—5 基于元数据的本体构建规范
- D009—6 基于元数据的本体构建规范应用指南
- D009—7 面向数字图书馆环境的分众分类法需求分析与功能设计
- D009—8 受控表语义描述规范之调研报告
- D009—9 术语注册和术语服务系统调研报告
- D009—10 本体和关联数据在数字图书馆中的应用调研报告

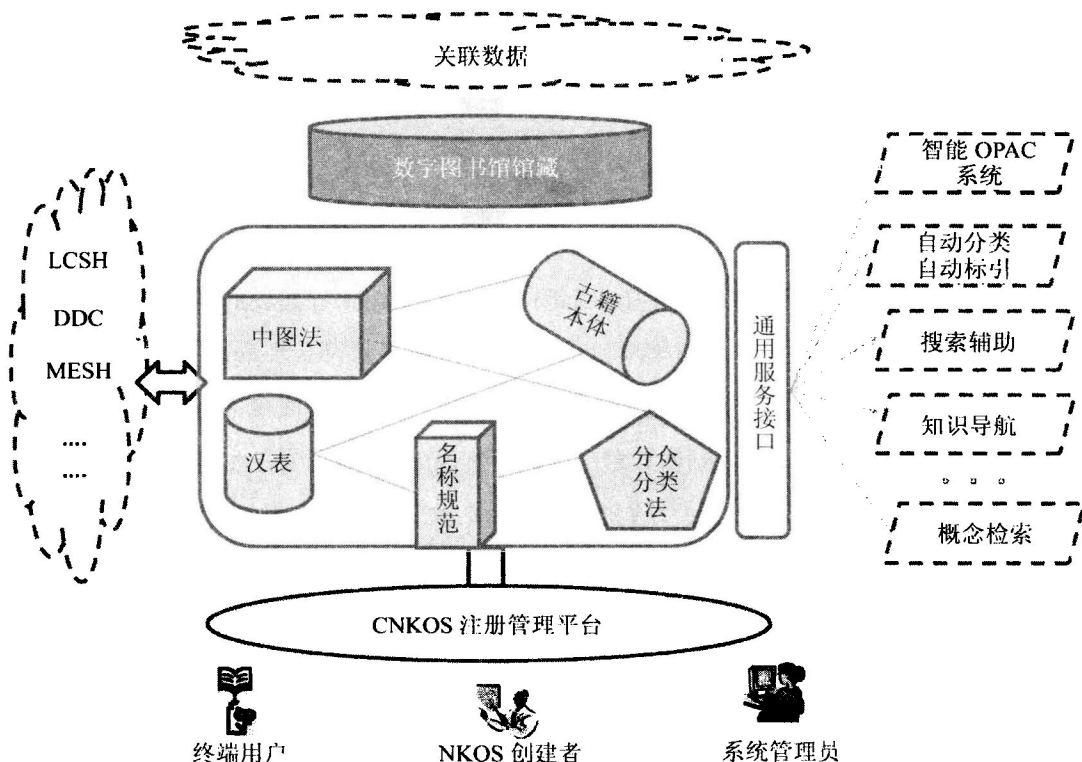


图2 数字图书馆体系结构中的 NKOS

图2展示了本项目所设计的 NKOS 在数字图书馆体系结构中的位置和作用。各类知识组织工具(包括分类法、词表、名称规范、本体、分众分类法等)在 NKOS 规范的作用下具备了互操作的能力,例如,“受控表语义描述规范”采用统一的方式来描述分类法、主题词表和名称规范。NKOS 的创建者在 NKOS 注册平台上注册、描述、管理自己创建的 NKOS;系统管理员对平台中的注册的 NKOS 执行系统管理和维护任务;而用户可以在该平台中发现、浏览和查询所需要的工具。除了向用户提供管理和查询界面,NKOS 注册管理平台通过通用服务接口和查询协议向各类网络应用提供服务,包括:OPAC 系统、搜索引擎、门户导航、自动分类与标引等。平台中管理的 NKOS 都是符合规范的,这样,它们就具备与其他采用 SKOS 描述的知识组织工具互操作的能力,如 LCSH、DDC、MeSH 等。在互操作的框架下,基于各类 NKOS 可以将数字图书馆的数字馆藏转化为关联数据发布到 Web 上。

总之,本项目的所有规范作为一个整体,其长远目标是:①将各类知识组织工具转化为机器可理解的语义化的知识库,使它们具备和其他知识组织系统广泛互操作与关联的能力;②以

NKOS 注册中心为平台,向数字图书馆内外的各类智能应用提供知识服务;③将数字图书馆的知识资源(包括 NKOS 和数字馆藏的元数据)转换为关联数据输送到整个互联网上,使得数字图书馆成为语义 Web 的知识关联枢纽。

4 关于项目成果的规范性说明

需要特别说明的是本项目的规范性要求和国家数字图书馆工程建设的现实需求之间的矛盾。作为国家数字图书馆工程的规范类项目,本项目成果应尽量符合规范文本的形式要求,具备通用性、规范性、概括性和简洁性。但是,由于以下两方面的因素,使得本项目成果很难完全符合规范文本的撰写要求,在文本结构和撰写风格上呈现出多样性、综合性、现实性的特点。

(1)项目需求书明确要求项目成果直接服务于国家数字图书馆工程的建设,具备实用性和可操作性。这使得若干子项目必须针对国家图书馆的具体对象、特殊需求和特定环境来设计,如 D009—1/2 可以直接作为国家图书馆未来建设 NKOS 注册平台的系统设计方案,D009—4 要针对“中国分类主题词表”进行语义描述,D009—5/6 实质上是基于元数据设计本体和生产语义元数据的操作指南,而 D009—7 是一个实现分众分类法与规范知识组织工具有机结合的解决框架。

(2)本项目中的不少内容尚属于学术界的前沿研究课题。一般地,应该在方案得到广泛检验、行业已有普遍应用之后,再将在实践中普遍遵循的惯例、准则等固定为规范。但是,国家数字图书馆工程建设的现实需求是:基于现有的研究成果和尝试性应用,在工程项目实施之前提供有关的设计参考、开发指南和应用范例。因此,本项目所产生的规范需要在实践应用中不断修正、补充与完善。

(3)本项目的研究内容和研究对象大多是网络信息环境下新出现的事物,如语义描述、关联数据、术语服务等,这些内容 and 应用背景已经超出了图书馆界传统的标准规范所覆盖的范围。为这些内容和对象撰写规范,所遵循的是相关的国际标准和规范。例如,本项目的 D009—3/4 在文本结构和撰写风格上沿用了 W3C 的语义描述规范 SKOS 的结构和风格,与国内标准规范文本在章节结构和撰写风格上有所不同。

5 出版说明

本项目共产生了 11 份文本,其中 D009—8/9/10 是项目调研报告,不作为规范文本发布。

考虑到分众分类法和应用环境紧密结合的特性,D009—7 以需求分析和功能设计书的方式向国家图书馆提交。为了出版和阅读的方便,在本书中,将 D009—1/2 合并为“网络知识组织系统注册与术语服务规范及应用指南”,将 D009—5/6 合并为“基于元数据的本体构建规范及应用指南”。由于 D009—4“中国分类主题词表”语义描述规范及应用指南与 D009—3 受控表语义描述规范的主体内容基本一致,故而略去。这样,本书由以下四篇构成:第一篇 网络知识组织系统注册与术语服务规范及应用指南;第二篇 受控表语义描述规范;第三篇 基于元数据的本体构建规范及应用指南;第四篇 面向数字图书馆环境的分众分类法需求分析和功能设计。

王 军 卜书庆
2011 年 12 月

目 录

前言	(1)
----------	-----

第一篇 网络知识组织系统注册与术语服务规范及应用指南

引言	(3)
1 范围	(4)
2 引用的规范性文件	(4)
3 术语和定义	(4)
4 总则	(6)
5 结构	(7)
6 NKOS 注册功能需求规范	(7)
7 NKOS 元数据规范	(13)
8 术语服务规范	(22)
9 NKOS 注册体系架构设计指南	(46)
10 NKOS 注册与术语服务应用案例	(62)
参考文献	(66)

第二篇 受控表语义描述规范

1 绪论	(71)
2 CNKOS 的命名域和词汇表	(76)
3 概念体系层次的描述	(80)
4 附表的描述	(81)
5 概念的描述	(85)
6 概念的语言标签的描述	(87)
7 等同关系的描述	(89)
8 概念的标记符号的描述	(91)
9 概念之间语义关系的描述	(93)
10 注释的描述	(96)
11 不同概念体系中概念之间映射关系的描述	(102)
12 特殊元素的描述	(106)
附录 A CNKOS 的抽象数据模型	(113)