

# 网络环境下知识组织规范研究与设计

王 军 卜书庆

**摘 要** 网络环境下知识组织系统的设计与编制是当前国际图书情报学研究的重点课题。国家图书馆的知识组织规范项目对网络知识组织系统(NKOS)中语义描述、体系结构、编制原则和应用指南等内容进行规范设计和撰写,覆盖了目前数字图书馆应用的主要 NKOS 类型,包括:分类法、主题词表、本体和分众分类法。整体设计上,将传统知识组织系统和新兴社会化组织工具集成在统一的术语注册与服务框架内,使二者相互补充、相互融合,以克服知识组织工具在使用上的障碍。项目的长远目标是将数字图书馆的知识资源(包括 NKOS 和数字馆藏的元数据)转换为关联数据传输到互联网上,使数字图书馆成为语义 Web 的知识关联枢纽。图 2。参考文献 12。

**关键词** NKOS SKOS 数字图书馆 受控表 本体 分众分类法  
分类号 G254

## A Study and Design on the Standard for the Networked Knowledge Organization System

Wang Jun & Bu Shuqing

**ABSTRACT** The design and implementation of the Networked Knowledge Organization System (NKOS) is currently a hot research topic in the library and information science. The NKOS standard project, initiated by The National Library of China, systematically describes the architecture, semantic description mechanism, design principle and application guide for mainstream NKOS types used in China. The standard includes a localized semantic description language based on SKOS for Chinese classification systems and thesauri, an ontology generation mechanism based on metadata schema, a social bookmarking system architecture and a framework for terminology registration and service. This project is featured with its framework that integrates different types of knowledge organization tools together, esp. the formal knowledge organization tools (like thesauri and classification schemes) and the social tagging systems, solving the user-gap problem in the application of KOS. 2 figs. 12 refs.

**KEY WORDS** NKOS. SKOS. Digital Library. Controlled Vocabularies. Ontology. Folksonomy.

### 1 引言

知识组织是在传统文献信息环境下发展起来的信息组织和利用手段,在百余年的应用过程中,形成并完善了分类法、主题词表等知识组织工具。在今天的网络环境下,知识组织工具需要进一步发展和创新,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<sup>[1]</sup>。NKOS 是指在网络环境下能够用于表示和组织数字信息资源、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 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<sup>[2]</sup>。前者在国际图书情报界已有较多的研究和实践,如美国国会图书馆已发布《国会图书馆主题词表》的SKOS语义化描述版本<sup>[3]</sup>;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<sup>[4]</sup>;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<sup>[5]</sup>;联合国粮农组织将其管理的词表AGROVOC向本体形式转换<sup>[6]</sup>;欧盟在多语言词表和词表的互操作方面进行了多年实践。后者(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,如W3C已发布了“图书馆关联数据应用指南”(Library Linked Data Incubator Group Final Report)<sup>[7]</sup>;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(如宾州大学、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、Millennium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。在此背景下,国家图书馆于2008年启动了NKOS构建的规范项目,它是国家数字图书馆工程标准规范建设的主导项目之一。项目的总体目标是: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已有成果和发展趋势,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。本文阐述项目的整体设计思想,介绍其设计方案和实施要求,总结最终成果的核心内容及其特点,并探讨相关问题。

## 2 研究目标

随着数字资源的不断增长,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、组织和管理的基礎,急需加快国内

NKOS的建设步伐。构建NKOS规范是实现在数字图书馆中大规模应用NKOS的前提和基础,其作用与意义有:①NKOS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国际上已经提出了一些构建NKOS的标准草案和通例,如简单知识组织系统SKOS语义描述体系、OCLC的术语服务协议等。参考这些国际通例来设计和编制国内的NKOS,将使国内的数字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基于传统知识组织工具来构建NKOS,将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,推动知识组织领域的创新和发展。④NKOS是网络信息资源利用的重要手段,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语义化知识资源是NKOS建设的终极目标。

基于以上背景,国家图书馆于2009年底启动了“国家图书馆知识组织规范”(NLC-NKOS)项目。项目要求在汇总国际上NKOS构建方法、技术和经验的基础上,结合国内知识组织工具的特点和国内数字图书馆应用环境的要求,开发制定我国NKOS的参考规范和应用指南。项目的设计要求是:

(1)参考国际上相关的研究成果,与国际上重要的NKOS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是目前图书馆学和信息科学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)<sup>[8]</sup>;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。NLC-NKOS应当充分借鉴已有的研究成果和实践应用经验,和相关的国际规范保持一致,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。NLC-NKOS 所研制的各种规范, 一方面要向国际上 NKOS 的构建规范靠拢, 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、开展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 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, 应根据现实应用情况进行修改和完善。另外, 知识组织工具现实的应用情况也十分复杂, NLC-NKOS 主要是针对通用型、综合性的知识组织工具而制定的规范草案, 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 设计方案

NKOS 包括的类型广泛, 从术语表、分类表、词表到语义网络和本体<sup>[9]</sup>。考虑到国家图书馆目前的建设需要和现有知识组织工具的基础以及目前国内外 NKOS 的发展水平、特别是 NKOS 的应用水平, NLC-NKOS 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成熟的、在数字图书馆中有较大应用潜力的 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 是图书馆延续了上百年的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、词表这类受控表的转换。在本项目中, 以《中国分类主题词表》

(CCT) 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<sup>[10]</sup>、ADL 词表查询协议<sup>[11]</sup>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: 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 都是图书馆员等专业人士在后台使用的工具。如何让普通用户方便、有效地利用是 NKOS 应用和发展的生命力所在。本项目提出将大众分类法转化为 NKOS 的词汇服务渠道, 在用户检索、标注等过程中得以直接访问 NKOS。

NLC-NKOS 的整体设计思想如图 1 所示。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

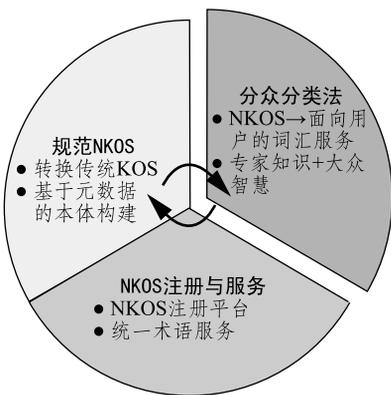


图 1 NLC-NKOS 的整体设计思想

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以下两点要求:第一,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;第二,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。为了满足要求,NLC-NKOS 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 词汇推荐服务:不仅推荐合适的标引词,还推荐词间关系,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。这样不仅将 NKOS 转化为方便用户使用的词汇服务,解决了“用户鸿沟”问题,而且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 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 平台注册、发现和服务的三大功能;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中实现知识服务可依赖的知识源。

## 4 规范构成

根据以上设计思想,NLC-NKOS 规划了以下四部分内容。

### 4.1 NKOS 注册和术语服务规范

本规范的目的是为国家图书馆建设 NKOS 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。

### 4.2 受控表语义描述规范及其在 CCT 上的应用指南

应用 W3C 提出的基于 RDF 的 SKOS 描述语言来表示图书分类法、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。作为应用指南,要求将所制定的语义描述方案用于描述《中国分类主题词表》(CCT),以检验其描述能力。CCT 是目前国内影响最大、使用最广泛、最权威的、一体化的分类法主题法知识组织工具,具有相当的复杂性。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。

### 4.3 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建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。

### 4.4 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,NLC-NKOS 针对国家图书馆的特定需求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知识组织系统应用指南,

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十一份文本。包括“国家图书馆知识组织规范”总体设计、NKOS 注册和术语服务规范及其应用指南、受控表语义描述规范及其在 CCT 上的应用指南、基于元数据的本体构建规范及其应用指南、面向数字图书馆环境的分众分类法需求分析与功能设计,以及各规范对应的调研报告。

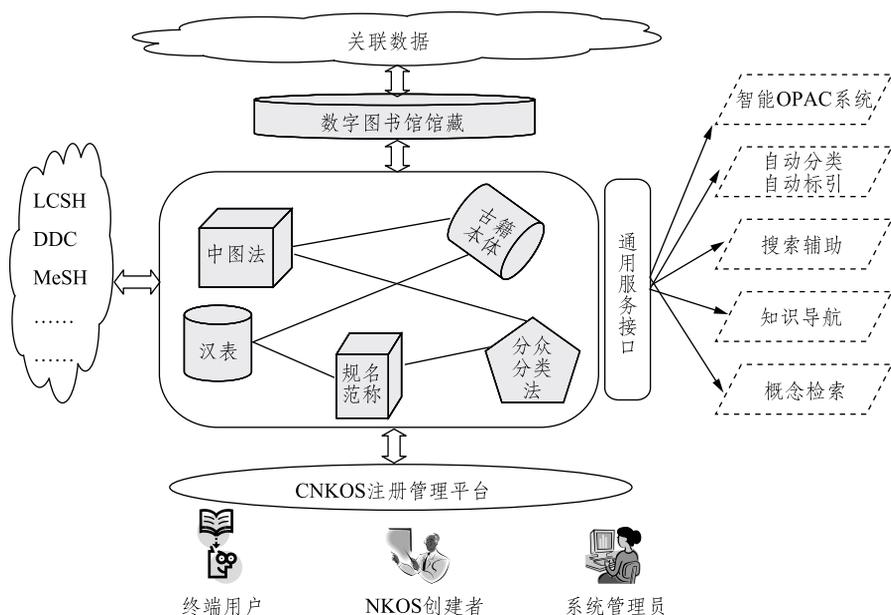


图 2 数字图书馆体系结构中的 NKOS

图 2 展示了所设计的 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 平台上注册、描述、管理自己创建的 NKOS;系统管理员对平台上注册的 NKOS 执行系统管理和维护任务;而用户可以在该平台上发现、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,而且具备与其他采用 SKOS 描述的知识组织工具的互操作能力,如 LCSH、DDC、MeSH 等。在互操作的框架下,各类 NKOS 都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,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互联网上,使数字图书馆成为语义 Web 的知识关联枢纽。

## 5 特色与总结

“国家图书馆知识组织标准规范”是国内图书馆界首次对网络环境下知识组织系统的设计、构建和应用进行全面系统的总结和编制。为了保证项目质量,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,并对项目需求进行了分析设计和严格的规范控制。该项目于2010年1月开始研制,由北京大学信息管理系、南京大学信息管理系和深圳大学图书馆三家单位共同组成了研制团队。2010年6月,研制团队提交了项目成果初稿,2010年底通过国家图书馆项目组的验收,2011年3月底通过馆内专家验收。最终成果于2011年5月进行了网上公开质询,2011年7月通过专家验收。在此过程中,国家图书馆项目组和验收质询专家先后多次提出修改意见,研制团队对项目成果进行了多次修改和完善,先后发布了第一版、第二版、第三版和最终版。项目于2011年11月正式结项。最终成果有以下四个特点:

(1)覆盖面广且重点突出。项目成果覆盖了当前主要的网络知识组织系统,既包括将传统的知识组织工具(如分类法、词表等)转换为NKOS的语义描述机制,也包括在数字图书馆中应用本体和分众分类法的实施方案。此外,还提出了一个网络环境下的术语注册和服务框架。这些内容一方面涵盖了目前主要的NKOS,另一方面强调了这些NKOS类型在数字图书馆中的重要地位。

(2)继承性与创新性兼备。项目在继承国际NKOS已有研究和应用成果的基础上,针对国内知识组织工具的特点和应用环境进行了适应性和独创性的设计。例如,考虑到国内的分类表和主题词表的复杂性,对SKOS的语义描述机制进行了本地扩展;为了解决用户鸿沟问题,在分众分类法中引入等级结构和词汇推荐服务;考虑到本体应用的复杂性,提出了基于元数据设计本体和批量生产语义元数据的方法。

(3)独立性和整体性互融。项目所包含的各个规范分别定义了受控表、本体和分众分类

法等NKOS系统,它们可以独立应用于不同的场景。同时,这些NKOS规范在命名空间、设计结构和描述机制上具有内在一致性和互操作性,可互相补充,形成一个有机互融的整体。如图2所示,从内部看,它们是不同的NKOS形态;从外部看,它们都通过NKOS术语注册和服务平台向外提供服务,联合起来将图书馆内部的数据资源转化为语义资源,融合到Web的大信息环境中。

(4)规范性与实用性并重。一方面,作为国家数字图书馆工程的规范类项目,NKOS规范应符合国家标准文本的格式化要求,具备通用性、规范性、概括性和简洁性;另一方面,项目的成果又需直接服务于国家数字图书馆工程建设的现实需要,更要具备实用性和可操作性。因此,许多内容都是针对国家图书馆的具体对象、特殊需求和特定环境来设计的,如“NKOS术语注册和服务规范”可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;“基于元数据的本体构建规范”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“分众分类法需求与设计书”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。

由于以上特点,本项目的最终文本呈现出多样性、综合性、现实性的特点。此外,NKOS的研究内容和研究对象是网络信息环境下出现的新事物,如语义描述、互联数据、术语服务等,这些内容和应用背景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需要参考和遵循相关的国际标准和规范,如W3C所发布的SKOS语义描述规范,这也使得NLC-NKOS的最终成果与国内标准规范文本在章节结构和撰写风格上有所不同。

NKOS是一个正在迅速发展和不断成熟的前沿研究课题<sup>[12]</sup>。为了满足国家数字图书馆工程建设的现实需求,为了向国内NKOS工程项目的实施提供有关的设计参考、开发指南和应用范例,我们基于NKOS领域现有的研究成果和尝试性应用设计起草了网络环境下知识组织的系列规范文本,这些内容需要在实践应用中不断修正、补充与完善。

参考文献:

[ 1 ] Gail Hodge, Linda Hill, et al. Next generation knowledge organization systems: Integration challenges and strategies [C] // Proceedings of the 5th ACM/IEEE - CS Joint Conference on Digital Libraries, ACM New York, NY, USA 2005.

[ 2 ] 王军, 张丽. 网络知识组织系统的研究现状和发展趋势 [J]. 中国图书馆学报, 2008 (1). (Wang Jun, Zhang Li. The status and trends of NKOS research [J]. The Journal of Library Science in China, 2008 (1).)

[ 3 ] Ed Summers, Antoine Isaac, Clay Redding, et al. LCSH, SKOS and linked data [C] // Proceedings of the 2008 International Conference on Dublin Core and Metadata Applications (DCMI'08). Dublin Core Metadata Initiative, 2008: 25 - 33.

[ 4 ] Terminology services technical overview [OL]. [2012 - 02 - 10]. <http://tspilot.oclc.org/resources/overview.pdf>.

[ 5 ] Hillmann D I, Sutton S A, Phipps J, et al. A metadata registry from vocabularies UP: The NSDL registry project [C] // the Dublin Core International Conference, 2006.

[ 6 ] Soergel D, Lauser B, Liang A, et al. Reengineering thesauri for new applications: The AGROVOC example [J]. Journal of Digital Information 2004, 4 (4).

[ 7 ] W3C. library linked data incubator group final report [OL]. [tor/llid/XGR-llid-20111025/.

\[ 8 \] SKOS simple knowledge organization system \[OL\]. \[2012 - 01 - 20\]. <http://www.w3.org/2004/02/skos>.

\[ 9 \] Gail Hodge. Systems of knowledge organization for digital libraries: Beyond traditional authority files \[M/OL\]. CLIR Publications, 2000 \[2012 - 02 - 10\]. <http://www.clir.org/pubs/abstract/pub91abst.html>.

\[ 10 \] Simon Jupp, Sean Bechhofer, Robert Stevens. A flexible API and editor for SKOS \[C\] // 6th Annual European Semantic Web Conference \(ESWC 2009\), 2009. 6: 506 - 520.

\[ 11 \] Greg Janec. ADL thesaurus protocol \[OL\]. \[2012 - 02 - 10\]. <http://www.alexandria.ucsb.edu/thesaurus/>.

\[ 12 \] 卜书庆. 近年来国际知识组织规范的进展研究 \[J\]. 中国图书馆学报, 2010, \(5\). \(Bu Shuqing. The progress of the research in the international knowledge organization standards \[J\]. The Journal of Library Science in China, 2010\(5\).\)](http://www.w3.org/2005/Incuba-</a></p></div><div data-bbox=)

王 军 北京大学信息管理系教授。通讯地址: 北京市海淀区颐和园路 5 号。邮编: 100871。

卜书庆 国家图书馆研究馆员。通讯地址: 北京市海淀区中关村南大街 33 号。邮编: 100081。

(收稿日期: 2012 - 02 - 15)