

中文知识组织系统形式化语义描述标准体系研究(一) ——扩展 SKOS 实现传统受控词表全描述*

曾新红

摘要 以 SKOS Core 为基础,以实现中文叙词表和分类法等传统受控词表的全描述为目标,研究传统中文知识组织系统的语义描述规范。其中,针对中文 KOS 描述需求中 SKOS 能够描述的部分,研究制定用 SKOS 描述的具体规范;对于现有 SKOS 词汇无法描述的部分,则利用 RDF/RDFS/OWL 的建模结构,基于 SKOS 的数据模型进行了本地(CKOS)扩展。本研究是国家图书馆知识组织标准规范的组成部分,同时也是中文知识组织系统形式化语义描述标准体系研究的基础。表 2。参考文献 33。

关键词 知识组织系统 SKOS CKOS 分类法 叙词表 形式化表示
分类号 G254 TP18

Formal Semantic Description Standard Family for Chinese KOS(I) : Realizing Full Description for Chinese Controlled Vocabularies by Expanding SKOS

Zeng Xinhong

ABSTRACT Based on SKOS Core,aiming to realize full description for traditional Chinese controlled vocabularies such as thesauri and Classification Schemes,this paper explores semantic description specifications. For the description requirements of Chinese KOS that can be described by SKOS,the study develops SKOS description specifications,and for those that cannot be described by SKOS,the local CKOS extension based on SKOS data model is put forward by using RDF/RDFS/OWL modeling constructs that also used by SKOS. The research results are parts of Knowledge Organization Standards of National Library,as well as constituent of Formal Semantic Description Standard Family for Chinese KOS. 2 tabs. 33 refs.

KEY WORDS KOS. SKOS. CKOS. Classification scheme. Thesaurus. Formal representation.

1 引言

国家数字图书馆工程“知识组织标准规范”项目的总目标是设计中文网络知识组织系统(Networked Knowledge Organization Systems, NKOS)的建设规范和共享共建平台,作为国家图书馆 NKOS 建设的指导规范和 NKOS 资源,提高国家图书馆应用 NKOS 的水平,推动国内

NKOS 的发展和应用。笔者承担了其中的 1.2 子项,即“基于传统知识组织系统的中文 NKOS 构建规范”,要求以传统的中文图书分类法和受控词表作为主要应用对象,设计“通用的中文网络知识组织系统(CNKOS) 构建规范”以及“《中国分类主题词表》(*Chinese Classified Thesaurus*, CCT) 语义化描述规范”。该项目于 2009 年底启

* 本文系国家数字图书馆工程“知识组织标准规范”(批准号: D009) 及广东省哲学社会科学“十一五”规划项目“中文知识组织系统的形式化语义描述标准体系研究”(批准号: GD10CTS02) 的研究成果之一。

通讯作者: 曾新红, Email: zengxh@szu.edu.cn

动 2010 年 5 月提交简略版初稿 8 月提交正式初稿, 历经小组审核、馆内专家审核、向社会公开质询 2011 年 7 月经过领域专家终审验收, 目前已全部完成。所提交的三份报告: 受控表语义描述规范 (D009 - 3)^[1]、“中国分类主题词表”语义描述规范及应用指南 (D009 - 4)^[2] 以及受控表语义描述规范之调研报告 (D009 - 8)^[3] 与其他七份报告一起, 共同组成了“国家图书馆知识组织标准规范”^[4]。

在该子项目的研究过程中, 要解决的最大难题就是传统受控表 (尤其是分类法) 的全描述问题。SKOS^[5-6] 主要针对叙词表的描述, 并兼顾各种 KOS 类型共性部分的描述。分类法作为情报检索语言的两大支柱之一, 与另一大支柱叙词表 (国内称为主题词表) 存在许多不同之处, 如果单纯用 SKOS 进行描述, 会丢失相当多的语义信息。另外, 叙词表中特种叙词概念的描述等许多问题, 也需要对 SKOS 进行扩展。笔者兼顾二者的需求, 充分借鉴国内外已有的相关研究, 在积累了多年的中文叙词表本体 (Onto-Thesaurus) 研究经验的基础上^[7-10], 采用 RDF/RDFS/OWL 等语义 Web 建模语言^[11], 在 SKOS 框架内对传统受控词表的语义描述进行了相应的扩展, 以期尽量实现全描述的目标。

2 国内外研究现状及本项目主要工作

2.1 国内外研究现状

W3C 发布的简单知识组织系统 SKOS 为知识组织系统的语义描述提供了基本的建模框架。SKOS 主要对应的是叙词表标准, 因此, 叙词表 (或结构更简单的主题标题表) 与 SKOS 的对应是直接的, 已有不少成功的应用, 如 AGRO-VOC^[12], LCSH^[13] 等。对于中文叙词表 (主题词表) 而言, 国外已有研究可以借鉴, 其特殊性主要在于拼音、英译名、族首词、附表的描述以及 URI 的确定。

国内外学者对中文 KOS 的形式化语义描述研究主要集中在《中国分类主题词表》(即 CCT) 或《汉语主题词表》的 SKOS 转换上, 其中有代表性的四项研究是: ①旅美学者曾蕾和秦健建立

了讨论 CCT 的 SKOS 化 wiki 网站^[14], 与国内外学者合作, 致力于 SKOS 标准文本的汉化, 研究讨论 CCT 到 SKOS 的转换方案。该网站中的相关讨论涉及 CCT 的 SKOS 化过程中需要解决的许多具体问题, 为笔者了解 SKOS 在表示 CCT 的某些特性上的不足、可能的解决方案以及 SKOS 的一些最新发展动向提供了良好的平台。②国家图书馆的喻菲研究了将 CCT (电子版) MARC 数据转化为 SKOS 的方法, 提出了 CCT “主题词—分类法对应表”部分的“主题词规范数据格式”与 SKOS 词汇之间的大致对应关系^[15]。③山西大学的贾君枝对《汉语主题词表》的 SKOS 表示进行了研究^[16]。④河北大学刘丽斌等对 CCT 电子版的 SKOS 描述的自动转换进行了研究^[17]。在初稿完成之后, 笔者又继续追踪了其他密切相关的研究成果, 如中科院文献信息中心张士男和宋文研究了《科图法》的 SKOS 描述方案设计^[18]; 南京政治学院上海分院段荣婷基于 SKOS 对《中国档案主题词表》进行的语义网络化应用研究^[19]。

笔者对以上国内外已有的相关研究进行了深入分析, 结论是: 应用 SKOS 词汇描述叙词表 (或主题标题表) 的叙词概念 (或标目概念)、首选语言标签、可选语言标签、上位/下位/相关词、注释的转换方案比较一致, 可以直接利用; 但现有研究对于拼音、英译名、主题词对应的分类号、附表、专类概念的进一步区分, 组配概念、分面及 Node label、组配概念的分面组成、等级和相关关系的进一步细化的语义化描述, 由于 SKOS 的“简单”定位以及研究者的不同理解, 要么没有涉及, 要么作了一些不尽相同的展望或建议。而笔者所研究的中文叙词表本体 (Onto-Thesaurus) 虽然解决了上述一些问题^[9-10], 但它是基于 OWL DL 的, 本项目需要针对 SKOS 的数据模型重新设计描述方案。

与叙词表 (或主题标题表) 相比, 国内外对分类法的语义描述研究要薄弱得多。分类法中与叙词表相似的共性部分, 可以采用 SKOS 来进行描述。但分类法中有大量不同于叙词表的特性, 与这部分描述相关的研究非常少, 目前已有的成果主要是针对杜威十进制分类法 (Dewey

Decimal Classification, DDC) 进行的研究,还处于起步阶段。其中两项研究较有代表性:①参考文献[20]针对 DDC 中的号码范围、附表、索引术语、类—主题关系、注释的内部结构,中图法(CLC)中的交替类号以及分类系统中的顺序等特殊表示问题进行了研究和探讨,提出了一些解决方案。这些方案虽然不是很成熟,也未被 SKOS 新版本所采纳,但其中的一些思想使笔者深受启发,在其基础上,笔者针对中国传统分类法的特点就以上问题提出了新的建议(借鉴与区别详见 D009-8 报告第 1.3.1 节^[3])。②OCLC 发布的 dewey.info^[21]将每一语种 DDC 的每一个版本视为一个独立的 Scheme,每一类(头三级类目)定义为 skos: Concept 的个体。目前提供了十个语种 DDC Summaries 的关联数据版本。该项目的 URI 方案为本项目的研究提供了参考。

2.2 项目主要工作

笔者依据我国情报检索语言的权威理论^[22]和构建标准^[23-24],参考《中国分类主题词表》^[25]、《汉语主题词表》、《中图法》^[26]、《军用电子分类表》^[27]的编制说明或规范,以及已有的国内外研究成果,首先以 SKOS 标准文本的定义为准绳,对中文叙词表及分类法中各项语义元素的 SKOS 描述进行了具体的规定或建议。

对于 SKOS 现有词汇无法描述的问题,则采用语义 Web 建模语言 RDF/RDFS/OWL,对 SKOS 进行扩展。扩展部分被称为 CKOS 词汇,规范中采用与 SKOS 标准文本类似的体例对其进行了说明、形式化定义,并给出了描述示例。

本着尽量尊重 SKOS 原有数据模型的原则,大部分 CKOS 词汇以原有 SKOS 词汇的子类或子属性的形式出现。CKOS 词汇与 SKOS 词汇采用的都是 RDF/RDFS/OWL 的建模结构,它们可以混合使用。SKOS 词汇加上 CKOS 词汇,共同组成了 CNKOS 扩展数据模型(D009-3^[1]的附录中给出了 CNKOS 的 UML 模型)。使用者可以根据自身需要,选择采用纯 SKOS 版本进行描述,或采用 CNKOS 扩展版进行描述。

3 CNKOS 总体介绍

本规范中,CNKOS 有两个含义:一是“中国网络知识组织系统”的简称;二是特指本语义描述规范,以区别于 SKOS。本规范中涉及的所有语义描述词汇统称为 CNKOS 词汇。

3.1 CNKOS 的命名域和词汇表

CNKOS 中扩展词汇的命名域(namespace)缩写为 ckos,所采用的 SKOS/RDF/RDFS/OWL 词汇使用其原有的命名域及缩写,详见表 1。

表 1 命名域 URI 缩写

URI	缩写
http://www.nlc.gov.cn/2010/06/ckos#(暂定)	ckos:
http://www.w3.org/2004/02/skos/core#	skos:
http://www.w3.org/1999/02/22-rdf-syntax-ns#	rdf:
http://www.w3.org/2000/01/rdf-schema#	rdfs:
http://www.w3.org/2002/07/owl#	owl:

CNKOS 的词汇表包括 SKOS 词汇表(SKOS vocabulary,即 SKOS Core 定义的类和属性)(详见 D009-3^[1]表 2)和扩展的 CKOS 词汇表两部

分(见表 2,因篇幅所限本文只列出少量的说明和描述示例,详细说明和更多的描述示例请见 D009-3 报告本身)。原则上 SKOS 的词汇都可

以在 CNKOS 的描述中应用,规范中列出了在传统中文 KOS 语义描述中可能会经常使用的词汇。扩展的 CKOS 类和属性分别是 owl: Class,

owl: ObjectProperty, owl: DatatypeProperty 或 owl: AnnotationProperty 的实例,大部分以 SKOS Core 已有类和属性的子类或子属性的形式出现。

表 2 CNKOS 中扩展的 CKOS 词汇

URI	说明	定义
ckos: Auxiliary	附表(类) 附属于某一概念体系、不能独立使用的子概念体系	定义 C1: ckos: Auxiliary 是 owl: Class 的实例。 定义 C2: ckos: Auxiliary 是 skos: ConceptScheme 的子类。
ckos: CompoundConcept	复合概念(类)	定义 C3: 这些扩展类都是 owl: Class 的实例。 定义 C4: 这些扩展类都是 skos: Concept 的子类。 例: < 锗\有机金属化合物\化工生产 > rdf: type ckos: CompoundConcept. < 哲学学派—研究 > rdf: type ckos: CompoundConcept. < 陆军战役+军兵种运用 > rdf: type ckos: CompoundConcept; skos: altLabel "陆军战役军兵种运用". < 1 世界 > rdf: type ckos: LocationConcept; skos: inScheme < 2WorldRegionTable >.
ckos: GeneralConcept	一般通用概念(类)	
ckos: PersonConcept	人物概念(类)	
ckos: LocationConcept	地名概念(类)	
ckos: OrganizationConcept	机构概念(类)	
ckos: FamilyConcept	家族名称概念(类)	
ckos: TitleConcept	题名概念(类)	
ckos: TimeConcept	时间概念(类)	
ckos: NationalityConcept	民族概念(类)	
ckos: LanguageConcept	语种概念(类)	
ckos: ResourceConcept	资源类型概念(类) 指信息资源的类型,如图书、文章、期刊等。	(注:为便于理解,本规范用“< 1 世界 >”表示类目概念“1 世界”的 URI(其他概念体系、附表、叙词概念等的 URI 表示类似),在真正的 RDF/XML 文件中,须用实际的 URI 替换此简略表示。)
ckos: MediaFormatConcept	媒体格式概念(类) 指媒体类型,如图像、文本等计算机格式。	
ckos: altClassEntryOf	是…的交替类目(属性)	定义 C7: ckos: altClassEntryOf 是 owl: ObjectProperty 的一个实例。 定义 C8: ckos: altClassEntryOf 的 rdfs: domain 和 rdfs: range 都是 skos: Concept。
ckos: topConcept	顶级概念是(属性) 揭示词族等级中的某一概念与其顶级概念之间的关系	定义 C9: ckos: topConcept 是 owl: ObjectProperty 的一个实例。 定义 C10: ckos: topConcept 的 rdfs: domain 和 rdfs: range 都是 skos: Concept。ckos: topConcept 是 skos: semanticRelation 的子属性(参见并比较 skos: hasTopConcept 的定义 S5 - S8 ^[5])。

续表

URI	说明	定义
ckos: crossClassNote	交叉类目注释。以“与 xxxx 交叉”显示,例如:与“M91 后勤训练”交叉。(军用电子分类表) 扩展 ckos: crossClassEntry 揭示其中隐含的链接关系。	定义 C11: ckos: crossClassNote ,ckos: altClassNote ,ckos: relatedClassNote ,ckos: referenceNote ,ckos: classGuideNote ,ckos: combineNote ,ckos: imitateNote 和 ckos: stopNote 都是 owl: AnnotationProperty 的实例。
ckos: altClassNote	交替类目注释。以“宜入 xxxx”显示,例如:宜入 X17。(CLC) 扩展 ckos: altClassEntryOf 揭示其中隐含的链接关系(见 D009-3 第 7 章)。	定义 C12: ckos: crossClassNote ,ckos: altClassNote ,ckos: relatedClassNote ,ckos: referenceNote ,ckos: classGuideNote 都是 skos: scopeNote 的子属性(sub-property)。
ckos: relatedClassNote	类目参照注释。以“参见 xxxx”显示,例 1:参见 E11 战役指挥 ,F11 战斗指挥 ,G31 作战指挥。(军用电子分类表) 例 2:参见 G420。(CLC) 使用 skos: related 揭示其中隐含的链接关系(见 D009-3 第 9 章)。	ckos: stopNote 是 skos: historyNote 的子属性。 定义 C13: ckos: combineNote 和 ckos: imitateNote 都是 skos: note 的子属性(sub-property)。
ckos: referenceNote	参考类目注释(关于“类目注释参见关系”的注释)。以“见 xxxx 注”显示,例 1:见 T1121.71 注。(军用电子分类表) 例 2:见 A16 注。(CLC) 扩展 ckos: referenceClassEntry 揭示其中隐含的链接关系。	例:《中图法》第四版“B 哲学、宗教”类目概念的注释相关描述 <2WorldRegionTabel > rdf: type skos: ConceptScheme; ckos: facetIdentity (“)”。(注:区分标识符,对附表中所有类目的类号适用。若附表中的类号已加此标识符,则不必声明,如总论复分表的区分符“-”)
ckos: classGuideNote	相关类目注释。以“xxxx 入 xxxx”显示,“入”字左面为类名或主题,右面为分类号,例 1:总论防空武器入此;专论入有关各类,例如:高射炮入 V2311.21 防空导弹入 V5311-15 地空导弹。(军用电子分类表) 例 2:学习心理学入 G442。(CLC) 可以使用 SKOS 的映射属性(如 skos: exactMatch ,skos: broadMatch)揭示主题与类目之间的映射关系,以方便自动分类等应用需求。参见 D009-3 第 11 章。也可使用 skos: related 建立当前类目与这些相关类目之间的链接关系。	<1GeneralTable > rdf: type skos: ConceptScheme. <B 哲学、宗教 > rdf: type skos: Concept; skos: prefLabel “哲学、宗教”; skos: scopeNote “总论哲学及兼论哲学与宗教的著作入此。”; ckos: classGuideNote “宗教入 B9。”; skos: related <B9 宗教 >; skos: scopeNote “专门科学的哲学理论入有关各类。例:教育哲学入 G40-02;历史哲学入 K01。”; skos: related <G40-02 教育哲学 >; skos: related <K01 史学的哲学基础 >;
ckos: combineNote	复分注释。例 1:依“军兵种复分表”组配复分。例如:步兵运用为 E8281 <21 >。(军用电子分类表) 例 2:依世界地区表分,如有必要,再仿 F249.2 分。(CLC) 扩展 ckos: combineFrom 揭示类目概念与复分表(组配类目所属附表或概念体系)之间的链接关系。	ckos: combineNote “依总论复分表分,-0 理论与方法论所属类目入 B0。”; skos: related <B0 哲学理论 >; ckos: combineFrom <1GeneralTable > . <B9 宗教 > rdf: type skos: Concept; skos: prefLabel “宗教”;

续表

URI	说明	定义
ckos: imitateNote	仿分注释。例 1: 仿“E6121 进攻战役”分。(军用电子分类表) 例 2: 如需细分时,可仿 F301/306 分。例: 林业企业经营管理为 F307.26。(CLC) 扩展 ckos: imitateClassEntry 揭示类目概念与仿分类目之间的链接。或析出仿分表,使用 ckos: combineFrom 揭示类目与仿分表之间的链接。	skos: broader <B 哲学、宗教>; ckos: combineNote “依总论复分表分。”; ckos: combineFrom <1 GeneralTable>; skos: historyNote “<3 版类名: 无神论、宗教>”; ckos: previousMatch <3 版/B 无神论、宗教>。 (注: 此时 3 版中的类目概念也必须可以通过 URI 访问,否则不作此前映射描述)
ckos: stopNote	停用注释。例如: <停用; 4 版改入 D815.6>(CLC) 扩展 ckos: useClassEntry 揭示停用类目与改入类目之间的链接。	
ckos: combineFrom	有复分表(属性) 建立类目概念与复分表或仿分表(即: 组配类目所属集合、附表或概念体系)之间的链接。	定义 C14: ckos: combineFrom 是 owl: ObjectProperty 的一个实例。 定义 C15: ckos: combineFrom 的 rdfs: domain 是 skos: Concept rdfs: range 是 skos: Collection, skos: ConceptScheme 和 ckos: Auxiliary 的并集(union)。
ckos: facetIdentity	分面标识(属性) 用于描述组配类号的分面标识符(如通用复分表的区分标识,号码加“0”等)。不同于 Node Label 的描述,参见 D009-3 第 12.3 节。	定义 C16: ckos: facetIdentity 是 owl: DatatypeProperty 的一个实例。其 rdfs: domain 是 skos: Collection, skos: ConceptScheme 和 ckos: Auxiliary 的并集。
ckos: crossClassEntry	有交叉类目(属性) 揭示交叉类目注释中隐含的链接关系。	定义 C17: ckos: crossClassEntry 是 owl: ObjectProperty 的一个实例。其 rdfs: domain 和 rdfs: range 都是 skos: Concept。它是 skos: related 的子属性。 定义 C18: ckos: crossClassEntry 是 owl: SymmetricProperty 的一个实例(即对称属性)。
ckos: referenceClassEntry	有参考类目(属性) 揭示参考类目注释中隐含的链接关系。也可用来进一步揭示指示性类目类号范围内包含的各个类目与该指示性类目的关系,可使分类款目更加完整,不丢失必要的注释信息。参见 D009-3 第 12.1 节。 和类目参照注释“参见 xxxx”类目的进一步揭示(用 skos: related)不同,这个属性表示的是要参考客体所指出类目的注释进行分类。	定义 C19: ckos: referenceClassEntry 是 owl: ObjectProperty 的一个实例。其 rdfs: domain 和 rdfs: range 都是 skos: Concept。它是 skos: related 的子属性,但不继承其对称性。

续表

URI	说明	定义
ckos: imitateClassEntry	<p>有仿分类目(属性)</p> <p>建立类目概念与仿分类目之间的链接,隐含的意思是:仿该类目的下位类目复分。</p> <p>如果是根据独立的仿分表仿分(若仿分的是一组类目,也建议析出为独立的仿分表)则不使用这个扩展属性描述,而是使用 ckos: combineFrom 描述其与仿分表之间的链接关系。</p>	<p>定义 C20: ckos: imitateClassEntry 和 ckos: useClassEntry 都是 owl: ObjectProperty 的实例。其 rdfs: domain 和 rdfs: range 都是 skos: Concept。它们都是 skos: related 的子属性,但不继承其对称性。</p>
ckos: useClassEntry	<p>改入类目(属性)</p> <p>揭示停用类目与改入类目之间的链接关系。</p>	
ckos: nonCloseMatch	<p>非近似映射(属性)</p> <p>相对于 skos: closeMatch,用于揭示 skos: closeMatch (和 skos: exactMatch) 之外的任何一种非近似映射关系。</p>	<p>定义 C21: ckos: nonCloseMatch 和 ckos: previousMatch 都是 owl: ObjectProperty 的实例。</p> <p>定义 C22: ckos: nonCloseMatch 是 skos: mappingRelation 的子属性(sub-property)。</p> <p>定义 C23: ckos: previousMatch 是 skos: relatedMatch 的子属性(sub-property)。但不继承其对称性。</p>
ckos: previousMatch	<p>前概念映射(属性)</p> <p>揭示修改过 URI 的概念与其前一版本概念之间的链接。</p>	
ckos: classEntryType	<p>类目类型(属性)</p>	<p>定义 C24: ckos: classEntryType 是 owl: DataTypeProperty 的一个实例。其 rdfs: domain 是 skos: Concept。</p>
ckos: coordinationOf	<p>由…组配而成(属性)</p> <p>用于揭示复合概念与其组配概念之间的链接关系。</p>	<p>定义 C25: ckos: coordinationOf 是 owl: ObjectProperty 的一个实例。</p> <p>定义 C26: ckos: coordinationOf 的 rdfs: domain 是 ckos: CompoundConcept,其 rdfs: range 是 skos: Concept。</p>
ckos: notationSpan	<p>类号范围(属性)</p> <p>仅排序用,不直接用来标引,可区别于 skos: notation 表示的类号。</p>	<p>定义 C27:</p>
ckos: notationCommon	<p>共有类号(属性)</p> <p>类号范围中的公共部分,用作组配的基础号码。</p>	<p>ckos: notationSpan, ckos: notationCommon, ckos: notationBegin 和 ckos: notationEnd 都是 owl: DataTypeProperty 的实例。</p> <p>定义 C28:</p>
ckos: notationBegin	<p>起始类号(属性)</p> <p>指明一组类目的起始类号。</p>	<p>ckos: notationSpan, ckos: notationCommon, ckos: notationBegin 和 ckos: notationEnd 都是 skos: notation 的子属性。</p>
ckos: notationEnd	<p>终止类号(属性)</p> <p>指明一组类目的终止类号。</p>	

3.2 CNKOS 的体例和结构说明

在编写体例上, CNKOS 规范参考了 W3C 发布的 SKOS Reference 20090818^[5] 的目录结构, 引用了 SKOS Core 的核心内容(译为中文), 并将国内受控词表的描述需求、扩展 CKOS 词汇的定义等融入其中, 以便于规范文本的网络发布和国际交流。同时也遵循了“GB/T 1.1-2009 标准化工作导则——标准的结构和编写”的基本要求。

为便于使用者阅读理解, 本规范在引言部分给出了详细的阅读说明, 包括形式化定义、URI 引用和构成建议、描述示例的格式说明等。另外, 在 D009-8 受控表语义描述之调研报告^[3]中, 给出了详细的参考资料要点, 说明采用 SKOS 词汇或扩展 CKOS 词汇进行描述的参考依据, 并对语义 Web 建模语言(RDF/RDFS/OWL)的基础知识进行了介绍, 以利于快速了解语义 Web 的基本原则、建模的意义、建模语言的种类和基本结构和它们的作用, 以及该领域常用术语的含义, 并站在语义 Web 的大背景下来看待 SKOS 的作用以及对 SKOS 进行的扩展^[28]。

4 中文受控词表的描述需求及对策

4.1 中文受控词表中与其他 KOS 类型共性部分的描述

采用 SKOS 词汇进行描述。这一步的关键在于将传统情报检索语言的理论和构建规范与 SKOS 的建模规范对应起来, 找到它们的契合点。例如: SKOS 的概念体系、概念、首选标签、候选标签、概念的标记符号、上位/下位/相关概念、一般的注释类型/映射关系等等, 可以与中文叙词表和分类法中的相应语义元素直接对应上。叙词或类名的汉语拼音、英译名采用相应语种的 `skos: altLabel` 进行描述。因篇幅关系, 本文不详细介绍这部分内容, 详情请查阅 D009-3 报告^[1]。

中文受控词表中的族首词、附表的描述, 采用 SKOS 词汇进行描述在语义上会略有损失或变化。因此, 对族首词、附表的描述除建议使用 SKOS 词汇外, 还提出了能够表达更准确语义的扩展建议(`ckos: topConcept` 和 `ckos: Auxiliary`, 详

见表 2)。

4.2 中文受控词表中组配概念及特种概念的描述

SKOS 还无法满足中文受控词表中组配概念(如主题词串、组配类号)的描述。本规范扩展 `skos: Concept` 的子类 `ckos: CompoundConcept` 进行描述。

组配概念与其组配成分概念之间的组配关系(LCSH/SKOS 也有此需求^[29]), 参考 SKOS primer(20090818 版)的建议, 本规范扩展定义了一个 OWL 对象属性 `ckos: coordinationOf` 对其进行描述, 未来可进一步扩展更具体的指明组配分面的子属性^[7]。

中文受控词表中特种概念(如附表中的地名概念等)的描述, SKOS 还无法满足(LCSH/SKOS 也有此需求^[29])。本规范扩展了 `skos: Concept` 的 9 个子类(如: `ckos: LocationConcept`)进行描述。这些子类的扩展主要来自国家社科基金项目 05CTQ001 的成果 *OntoThesaurus*^[8-10], 部分类名作了修改, 以便拥有更广泛的适应性。通过扩展子类来提高描述能力是建模语言常用的一种手段, 也是让 CNKOS 未来继续向细粒度本体发展的重要一步。

4.3 传统分类法中特殊语义元素的描述

分类法中有大量不同于叙词表的特性, 如果仅采用 SKOS 词汇进行描述, 将会牺牲掉其原有的丰富语义。因此, 本规范针对交替类目、类目注释的具体类型及其隐含语义、类号范围等问题的描述提出了扩展建议。

(1) 交替类目。扩展 `ckos: altClassEntryOf` 描述交替类目与其使用类目之间的关系, 这是两个类目概念之间的关系(详见表 2 及 D009-3/D009-4 第 1 部分第 7 章以及 D009-8 第 1.3.1 中的说明)。

(2) 类目注释的具体类型及其隐含语义。SKOS 现有的注释属性无法区分分类法中的一些具体注释类型。为了揭示这些注释类型中隐含的链接关系并实现注释文本的自动生成, 本规范提出了一些扩展建议(详见表 2)。

(3) 类号范围的描述。起止类(如《中图

法》中的“I3/7 各国文学”)是中外体系分类法中共有的描述问题,它用一个类号范围(起止类号)形式代表一个上位类目概念,实质上包含这组类号的公共部分及其复分方式,需进一步揭示才能让机器理解,以满足自动分类等应用需求。

另外,《中图法》中的指示性类目,其类号也是一个类号范围,但它表示的是与该指示性类目中注释相关的一组已存在类目的起止类号,一般采用缩略形式(如 B302/305),机器不易理解。如明确揭示其起止类号,可支持系统将将该指示性类目的注释和组配依据应用于起止类号之间所有类号对应的类目。

本规范扩展 `ckos: notationSpan`、`ckos: notationCommon`、`ckos: notationBegin` 和 `ckos: notationEnd` 等四个属性分别描述类号范围、共有类号、起始类号和终止类号,它们都是 `skos: notation` 的子属性。类号范围不能直接用于分类索引,与一般的类号有区别。

(4) 类目类型的描述和指示性类目的进一步揭示。本规范建议扩展 `ckos: classEntryType` 描述除使用类目(首选类目)之外的特殊类目类型;使用扩展的 `ckos: notationSpan`、`ckos: notationBegin` 和 `ckos: notationEnd` 描述指示性类目中的类号范围及其起止类号;使用扩展的 `ckos: referenceClassEntry`,进一步揭示指示性类目类号范围内包含的各个类目与该指示性类目的关系,可使这些类目语义更加完整,在独立使用时不会丢失必要的注释信息(详见 D009-3/D009-4 第 1 部分第 12.1 节)。

4.4 不同概念体系中概念之间映射关系的描述

本规范基本采用 SKOS 已有的映射属性进行描述,对中文叙词表和分类法中存在的类号与主题词之间的各种映射类型(如直接对应、间接对应)选择哪种 SKOS 映射属性进行描述提出了具体建议。

针对间接对应转换中需要大量人工判断才能确定上位/下位/相关映射的情况,本规范建议扩展一个与 `skos: closeMatch` 对应的 `ckos: nonCloseMatch` 属性,用于描述 `skos: closeMatch`

(及 `skos: exactMatch`)之外的上位/下位/相关映射等关系。

当修改造成叙词概念或类目概念的内涵或外延发生变化时,其 URI 也应随之变化。本规范建议扩展 `skos: relatedMatch` 的子属性 `ckos: previousMatch`,用于描述不同版本概念体系中新旧概念之间的链接关系,以方便应用程序作出相应的反应。

特例:《中国分类主题词表》(CCT)内分类法和主题词表的映射机制

CCT 由分类法和主题词表两部分组成,二者相互对照、相互融合、互为索引^[25]。这两个组成部分规模庞大,并且一个属于分类法类型,一个属于叙词表类型,对支持系统有不同的处理要求,因此,本规范建议将 CCT 描述为两个独立的概念体系: CCT_CLC(分类号—主题词对应表)和 CCT_CT(主题词—分类号对应表),并分别用两个 RDF/XML 文件保存和提供服务。

CCT_CLC 的 MARC 记录中包含了各类目对应的主题词,而 CCT_CT 的 MARC 记录中也包含了主题词对应的分类号(类目),因此,只需依照本规范中不同概念体系中概念之间映射关系的描述方法,在各自的描述文件中,将对应关系转换为与另一概念体系中相应概念的映射关系即可。

同样,其他分类法中的类—主题或类—索引术语对应关系,以及叙词表(主题词表)中的叙词—范畴号对应关系,均可采用以上方式进行描述。

4.5 分面分析及节点标记(Node Label)的描述

传统叙词表中存在一种比较特殊的结构:由节点标记引导的一组分面。ISO25964-1^[30]认为叙词表中的分面包含两种类型:①当节点标记显示上位术语的区分特征时,跟随的是真正的下位词;②当节点标记引进新分面时,跟随的往往不是下位关系词。而节点标记本身,并不是叙词表中的术语,它们只用作系统显示目的。为了给这种概念集合结构正确建模,SKOS 引入了 `skos: Collection` 类。SKOS primer(20090818 版)^[6]对分面分析及节点标记(Node

Label) 的描述提供了两种建议。

本规范建议采用其中的第二种建议进行改造和描述(详见 D009-3 第 12.3 节)。

5 中文知识组织系统 CNKOS 语义描述版本的应用前景

通过注册中心或建立服务网站,中文知识组织系统 CNKOS 语义描述版本可提供以下面向全球的服务:

① Linked Data 服务(参考 LCSH^[13] 以及 OTCSS Linked Data 服务示例^[31])。

② 术语服务。通过检索界面向网络用户(人)提供服务(如 OntoThesaurus-TS 的做法^[32]);通过 Web Service API 向应用程序(机器)提供服务(如 OntoThesaurus-API 的做法^[33])。

③ 网络共建服务,即在使用过程中广泛收集用户建议,允许网上修订,实现及时更新(如 AGROVOC^[12] 和 OTCSS^[32] 的做法)。

④ 可添加与其他广泛应用的词表(如 LCSH、UDC、DDC) 之间的映射,以实现更广泛的互操作和国际间合作。

⑤ 为一些有需要的领域扩展细化新的关系种类,形成领域本体,提供 OWL 版本的下载和服务(如 AGROVOC^[12] 的做法)。

6 结语

本规范保留了 SKOS Core 的完整数据模型,因此对于 SKOS 适用的主题标题表、分类表、民间分类表和其他类似的受控词表同样适用。比分类法和叙词表结构更简单的中文规范档、术语列表(规范代码表)等类型,也可以简化使用本规范进行描述。在今后的研究中,将在规范中增加对这些 KOS 类型的显性支持,也将继续开发针对 CNKOS 描述规范的支持系统,使其在实践中得到验证和改进。

中文知识组织系统形式化语义描述标准体系拟包含纯 SKOS 版、CNKOS 扩展版、高受控 KOS 版及 OWL 应用规范四个部分。本规范的研究已基本完成前两个部分的主要内容,今后

还需扩展对其他 KOS 类型的支持。第三部分将在中文叙词表本体(OntoThesaurus)的基础上进行扩展和改进,目标是提供一个高形式化程度的描述规范,拥有严格的一致性检测功能和共建共享配套功能。第四部分直接使用 OWL(及 OWL2) 制订中文 KOS 应用规范并补充相应的中文 KOS 描述示例。

本课题所有的研究成果将陆续在深圳大学图书馆 NKOS 研究室网站(<http://nkos.lib.szu.edu.cn>) 上发布,敬请关注并提出宝贵意见,也希望与各方有识之士建立良好的合作关系,共同推进我国 NKOS 研究的发展。

参考文献:

- [1] 曾新红. 受控表语义描述规范(D009-3) [R]. 国家图书馆, 2011. 8. (Zeng Xinhong. Semantic description specification for Chinese controlled vocabularies(D009-3) [R]. National Library of China 2011. 8.)
- [2] 曾新红, 王军. “中国分类主题词表”语义描述规范及应用指南(D009-4) [R]. 国家图书馆, 2011. 8. (Zeng Xinhong, Wang Jun. Semantic description specification for Chinese Classified Thesaurus and application guide(D009-4) [R]. National Library of China 2011. 8.)
- [3] 曾新红. 受控表语义描述规范之调研报告(D009-8) [R]. 国家图书馆, 2011. 8. (Zeng Xinhong. Research report of semantic description specification for Chinese controlled vocabularies(D009-8) [R]. National Library of China, 2011. 8.)
- [4] 王军, 欧石燕, 曾新红. 国家图书馆知识组织标准规范 [R]. 国家图书馆, 2011. 8. (Wang Jun, Ou Shiyuan, Zeng Xinhong. Knowledge organization standards of National Library of China [R]. National Library of China 2011. 8.)
- [5] W3C. SKOS Simple Knowledge Organization System reference: W3C recommendation 18 August 2009 [OL]. [2010-02-23]. <http://www.w3.org/TR/2009/REC-skos-reference-20090818/>.
- [6] W3C. SKOS Simple Knowledge Organization System primer: W3C working group note 18 August 2009 [OL]. [2010-02-23]. <http://www.w3.org/TR/2009/NOTE-skos-primer-20090818/>.
- [7] 曾新红. 《中国分类主题词表》的 OWL 表示及

- 其语义深层揭示研究[J]. 情报学报, 2005(2): 151 - 160. (Zeng Xinhong. Research on representation of Chinese Classified Thesaurus in OWL and Its implied semantic reveal [J]. Journal of the China Society for Scientific and Technical Information 2005(2): 151 - 160.)
- [8] 曾新红, 明仲, 蒋颖, 等. 中文叙词表本体共建共享系统 OTCSS 的研究与实现[R]. 全国哲学社会科学规划办公室, 2008. 8. [2010 - 10 - 20]. <http://nkos.lib.szu.edu.cn/OntoThesaurus/05CTQ001Report.pdf>. (Zeng Xinhong, Ming Zhong, Jiang Ying, et al. Research and implementation of OTCSS(Onto thesaurus co-construction and sharing system) [R]. National Planning Office of Philosophy and Social Sciences 2008. 8. [2010 - 10 - 20]. <http://nkos.lib.szu.edu.cn/OntoThesaurus/05CTQ001Report.pdf>.)
- [9] 曾新红. 中文叙词表本体的形式化表示与 SKOS 的比较研究——以及对建立中文知识组织系统形式化表示标准体系的建议[J]. 中国图书馆学报, 2010(2): 99 - 106. (Zeng Xinhong. A comparative study of the formal representation of OntoThesaurus and SKOS: The suggestion for the establishment of a formal representation standard system for Chinese KOS [J]. Journal of Library Science in China 2010(2): 99 - 106.)
- [10] 曾新红. 中文叙词表本体 OntoThesaurus 词汇表 [EB/OL]. [2010 - 11 - 11]. <http://nkos.lib.szu.edu.cn/2010/10/ont/>. (Zeng Xinhong. OntoThesaurus vocabulary [OL]. [2010 - 11 - 11]. <http://nkos.lib.szu.edu.cn/2010/10/ont/>.)
- [11] W3C. OWL Web ontology language reference: W3C recommendation 10 February 2004 [OL]. [2004 - 04 - 14]. <http://www.w3.org/TR/2004/REC-owl-ref-20040210/>.
- [12] FAO. AGROVOC Thesaurus [OL]. [2010 - 08 - 11]. <http://aims.fao.org/website/AGROVOC-Thesaurus/sub>.
- [13] Library of Congress. Library of congress subject headings [OL]. [2011 - 09 - 15]. <http://id.loc.gov/authorities/subjects.html>.
- [14] Marcia Lei Zeng, Jian Qin. CCT2 senariors [OL]. [2010 - 03 - 04]. http://www.metadataetc.org/wiki/index.php?title=CCT2_senariors.
- [15] 喻菲. 《中国分类主题词表》网络化研究: 从 MARC 到 SKOS [G] // 戴维民, 赵建华, 汪东波, 等. 网络环境下信息组织的创新与发展: 全国第五次情报检索语言发展方向研讨会论文集. 北京: 国家图书馆出版社, 2009: 139 - 147. (Yu Fei. A study on converting the MARC data of Chinese Classified Thesaurus into SKOS format [G] // Dai Weimin, Zhao Jianhua, Wang Dongbo, et al. Innovation and development of information organization in a networked environment: Proceedings of the 5th national symposium on information retrieval language development direction. Beijing: National Library of China Publishing House 2009: 139 - 147.)
- [16] 贾君枝. 简单知识组织系统与汉语主题词表 [J]. 中国图书馆学报, 2008(1): 75 - 78, 84. (Jia Junzhi. Simple Knowledge Organization System and Classified Chinese Thesaurus [J]. Journal of Library Science in China 2008(1): 75 - 78, 84.)
- [17] 刘丽斌, 张寿华, 濮德敏, 等. 《中国分类主题词表》的 SKOS 描述自动转换研究 [J]. 中国图书馆学报, 2009(6): 56 - 60. (Liu Libin, Zhang Shouhua, Pu Demin, et al. Automatic transformation of *Classified Chinese Thesaurus*(CCT) description with SKOS [J]. Journal of Library Science in China 2009(6): 56 - 60.)
- [18] 张士男, 宋文. 《科图法》SKOS 描述方案设计 [J]. 现代图书情报技术, 2010(6): 7 - 11. (Zhang Shinan, Song Wen. Description scheme of LASC in SKOS [J]. New Technology of Library and Information 2010(6): 7 - 11.)
- [19] 段荣婷. 基于简约知识组织系统的《中国档案主题词表》语义网络化应用研究 [J]. 现代图书情报技术, 2010(10): 33 - 42. (Duan Rongting. Study on the Chinese Archival Thesaurus application in semantic Web based on SKOS [J]. New Technology of Library and Information 2010(10): 33 - 42.)
- [20] Michael Panzer, Marcia Lei Zeng. Modeling classification systems in SKOS: Some challenges and best-practice recommendations [C]. Proc. Int'l Conf. on Dublin Core and Metadata Applications. DCMI 2009: 3 - 14.
- [21] OCLC. Dewey. info [OL]. [2010 - 03 - 04]. <http://www.oclc.org/dewey/webservices/default.htm>.
- [22] 戴维民. 信息组织 [M]. 第二版. 北京: 高等教育出版社, 2009. (Dai Weimin. Information organ-

- ization [M]. Second Edition. Beijing: Higher Education Press 2009.)
- [23] 中华人民共和国国家标准. GB 13190 - 91 汉语叙词表编制规则 [S]. 国家技术监督局, 1991. (National Standard of the People's Republic of China. GB 13190 - 91 ,Guidelines for establishment and development of Chinese thesauri [S]. China State Bureau of Technical Supervision , 1991.)
- [24] 中华人民共和国国家标准. GB/T 3860 - 2008 , 文献主题标引规则(报批稿) [S]. 国家质量监督检验检疫总局. (National Standard of the People's Republic of China. GB/T 3860 - 2008 , Guidelines for determining the subjects and selecting the descriptors(approval draft) [S]. General Administration of Quality Supervision , Inspection and Quarantine of the People's Republic of China.)
- [25] 《中国图书馆分类法》编辑委员会. 中国分类主题词表[M]. 第二版. 北京: 北京图书馆出版社, 2005. (*Chinese Library Classification* Editorial Board. Chinese Classified Thesaurus [M]. Second Edition. Beijing: Beijing Library Press 2005.)
- [26] 《中国图书馆分类法》编辑委员会. 中国图书馆分类法[M]. 第四版. 北京: 北京图书馆出版社, 1999. (*Chinese Library Classification* Editorial Board. Chinese Library Classification [M]. Fourth edition. Beijing: Beijing Library Press ,1999.)
- [27] 中华人民共和国国家军用标准. GJB6793 - 2009 军用电子分类表编制规则 [S]. 中国人民解放军总装备部. (National Military Standard of the People's Republic of China. GJB6793 - 2009 , Guidelines for establishment of military electronic classification schemes [S]. PLA General Armament Department.)
- [28] Dean Allemang James Hendler. 实用语义网: RDFS 与 OWL 高效建模[M]. 北京: 人民邮电出版社, 2009. (Dean Allemang James Hendler. Semantic Web for the working ontologist: Effective modeling in RDFS and OWL [M]. Beijing: People's Posts and Telecommunications Press 2009.)
- [29] Ed Summers ,Antoine Isaac ,Clay Redding 等. LCSH , SKOS 和关联数据[J]. 现代图书情报技术 2009(3): 8 - 14. (Ed Summers ,Antoine Isaac ,Clay Redding et al. LCSH ,SKOS and linked data [J]. New Technology of Library and Information 2009(3): 8 - 14.)
- [30] ISO. ISO/CD 25964 - 1 Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval [S]. ISO 2008.
- [31] 深圳大学图书馆 NKOS 研究室. OTCSS linked data [EB/OL]. [2011 - 06 - 13]. http://nkos.lib.szu.edu.cn/level_2/service/linkeddadata.jsp. (NKOS Research Office of Shenzhen University Library. OTCSS linked data [EB/OL]. [2011 - 06 - 13]. http://nkos.lib.szu.edu.cn/level_2/service/linkeddadata.jsp.)
- [32] 深圳大学图书馆 NKOS 研究室. OntoThesaurus-TS [EB/OL]. [2011 - 06 - 13]. http://nkos.lib.szu.edu.cn/level_2/communication.jsp. (NKOS Research Office of Shenzhen University Library. OntoThesaurus-TS [EB/OL]. [2011 - 06 - 13]. http://nkos.lib.szu.edu.cn/level_2/communication.jsp.)
- [33] 深圳大学图书馆 NKOS 研究室. OntoThesaurus-API [EB/OL]. [2011 - 06 - 13]. http://nkos.lib.szu.edu.cn/level_2/service/ontothesaurusapi.jsp. (NKOS Research Office of Shenzhen University Library. OntoThesaurus-API [EB/OL]. [2011 - 06 - 13]. http://nkos.lib.szu.edu.cn/level_2/service/ontothesaurusapi.jsp.)

曾新红 深圳大学图书馆研究馆员 ,NKOS 研究室主任。通讯地址: 广东省深圳市南山区深圳大学图书馆。邮编: 518060。

(收稿日期: 2011-09-15; 修回日期: 2011-11-14)