

传统受控词表的SKOS描述规范及应用指南^{*}

曾新红

(深圳大学图书馆NKOS研究室, 深圳 518060)

摘要: 依据我国情报检索语言的权威理论和构建标准, 参考中国分类主题词表、汉语主题词表、中图法、军用电子分类表的编制说明或规范, 以及已有的国内外研究成果, 以SKOS标准文本的定义为准绳, 将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, 从易于共享和实现的角度, 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, 并通过示例介绍了该规范的具体应用方法。

关键词: 受控词表; 分类法; 叙词表; 形式化表示; 语义描述; SKOS; CNKOS; 知识组织系统

中图分类号: G254; TP18

DOI: 10.3772/j.issn.1673—2286.2014.04.002

1 引言

国家数字图书馆工程“知识组织标准规范”项目(项目编号D009, 合同编号GC-HD090453)的总目标是设计中文NKOS的建设和共享共建平台, 作为国内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”, 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7份报告一起, 共同组成了“国家图书馆知识组织标准规范”^[4]。完整的“受控表语义描述规范”请见国家图书馆出版社近期出版的“网络环境下的知识组织规范和应用指南”一书第二篇^[5], 或登录“<http://nkos.lib.szu.edu.cn/2011/08/cnkos/>”下载最新版“通用CNKOS语义描述规范”, 供机器使用的RDF Schema版本的URI为: <http://nkos.lib.szu.edu.cn/2011/12/ckos/>。

“受控表语义描述规范”(CNKOS)实际包含两个版本: 纯SKOS版和CNKOS扩展版。CNKOS扩展版

的主要内容已在文献[6]中进行了介绍。鉴于目前国内受控词表的语义描述研究和实践重点还在SKOS描述阶段, 并且因为理解上的不同产生了一些不同的具体描述方法, 为了方便本规范的使用者更准确地把握SKOS在中文受控词表语义描述中的使用方法, 本文将专门介绍纯SKOS版, 包括在其设计过程中的具体考量、具体的描述示例以及使用中需要注意的问题。

2 国内外研究现状

简单知识组织系统SKOS^[7,8]及其在国内外的一般研究现状在文献[6]中已作了详细介绍, 本文不再赘述。简言之, 对于中文叙词表(主题词表)而言, 国外已有研究可以直接借鉴, 其特殊性主要在于拼音、英译名、族首词、附表等的描述以及URI的确定。笔者对已有的国内外相关研究进行了比较深入的分析, 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、注释等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

^{*} 本研究得到国家数字图书馆工程项目“知识组织标准规范”(编号: D009)、广东省哲学社会科学“十一五”规划项目“中文知识组织系统形式化语义描述标准体系研究”(编号: GD10CTS02)和国家社会科学基金项目“中文知识组织系统形式化语义描述标准体系研究”(编号: 12BTQ045)的资助。

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等的语义化描述,由于SKOS的“简单”定位以及研究者的不同理解,要么没有涉及,要么做了一些展望或不尽相同的建议,没有直接可利用的成功经验。

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。SKOS主要是针对叙词表标准的,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,与这部分的描述相关的研究非常稀少。

笔者依据我国情报检索语言的权威理论^[9]和构建标准^[10,11],参考中国分类主题词表^[12]、汉语主题词表、中图法^[13]、军用电子分类表^[14]的编制说明或规范,以及已有的国内外研究成果,以SKOS标准文本的定义为准绳,将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,找到它们之间的契合点,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。

3 CNKOS纯SKOS版的总体介绍

3.1 命名域和词汇表

纯SKOS版中所采用的SKOS/RDF/RDFS/OWL词汇使用其原有的命名域及其缩写,详见表1。

表1 命名域URI 缩写

URI	缩写
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:

纯SKOS版原则上包含SKOS Core的所有词汇,规范中列出了其中在传统中文KOS语义描述中可能会经常使用的词汇。详见CNKOS规范全文^[1,5]中的表2。规范中还包含了这些SKOS词汇的说明及定义的中译文,并对某些比较抽象的定义在其后用“()”添加了自然语言注释。

3.2 CNKOS纯SKOS版的使用方法

CNKOS参考了W3C发布的SKOS Reference 20090818^[7]的目录结构,引用了SKOS Core的核心内容(译为中文)。为便于使用者阅读理解,CNKOS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CNKOS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及它们的作用,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。

该版本适用于国内传统的叙词表(主题词表)和分类法的SKOS描述,也可应用于中文规范文档和其他KOS类型的SKOS描述。

4 中文受控词表的描述需求及具体的SKOS描述规范

4.1 受控词表整体的描述

作为整体的中文KOS,如中文叙词表(主题词表)、分类法等国内受控词表,可以视为一个概念体系(concept scheme),用skos:ConceptScheme及其相应的属性进行描述。

根据SKOS的定义,一个SKOS概念体系可以被看作是一个或多个SKOS概念(concept)的集合(aggregation)。这些概念之间的语义关系(链接)也可以被看作是一个概念体系的一部分(此句仅为建议而非约束)。一个独立的SKOS概念体系大致对应于一个独立的叙词表、分类法、主题标题表或其他知识组织系统。

当一个CNKOS文件(RDF/XML序列格式)中含有多个概念体系时,概念与概念体系之间的关系可以用skos:inScheme来描述。从利于实现的角度,本规范建议尽量不要将两个以上独立的概念体系放在一个CNKOS文件中进行描述,尤其当它们规模庞大或分别属于不同的NKOS类型时。除非这些概念体系不可分割,如将分类

法的附表也描述为概念体系，参见4.2节。

概念体系与其包含的顶级概念之间的关系可用skos:topConceptOf或skos:hasTopConcept来描述。这两个属性是互逆的，建议可选择其中之一（如skos:hasTopConcept）进行描述。需要注意的是，用skos:topConceptOf或skos:hasTopConcept描述的是概念体系与其包含的顶级概念之间的关系。这与传统中文叙词表叙词款目中的族首词关系在语义上有所不同。

叙词款目中的族首词表达的是叙词与其所属词族的顶级概念之间的关系，即SKOS概念与SKOS概念之间的关系，而非SKOS概念体系与SKOS概念之间的关系，SKOS未提供相应的描述词汇。如需声明叙词款目中的族首词关系，可采用扩展的ckos:TopConcept进行描述^[5,6]。

如例1中将中图法（CLC）描述为一个概念体系，并描述了它与其一级大类（顶级概念）的关系。

例 1
<pre><CLC> rdf:type skos:ConceptScheme; skos:hasTopConcept <B 哲学、宗教>. <B 哲学、宗教> rdf:type skos:Concept.</pre>

注：RDF/XML格式的SKOS描述代码可登录CLC Linked Data服务^[15]检索下载。

例2中将汉语主题词表（CT）描述为一个概念体系，并描述了它与其顶级概念（词族的族首词）的关系。

例 2
<pre><CT> rdf:type skos:ConceptScheme; skos:hasTopConcept <植物> . <植物> rdf:type skos:Concept.</pre>

注：RDF/XML格式的SKOS描述代码可登录CCTI Linked Data服务^[16]检索下载。

4.2 附表的描述

采用CNKOS扩展版时叙词表中的附表（特种概念集合）可以直接处理为skos:Concept的子类。当选择用纯SKOS版本描述叙词表时，因只能采用skos:Concept来描述所有的叙词概念，如想保留某些特种概念的特殊性，可采用skos:ConceptScheme来描述这些附表或特种概念集合（利于独立使用和共享）。这样在需要时可以根据指定的skos:ConceptScheme的URI来识别其所包含的特种概念，也可再转换为相应的子概念类型。

对于分类法中不能独立使用的附表（通用复分表，专类复分表，被仿分类目范围），有以下三个备选方案：

•skos:Collection

适于表示简单的、无等级关系和注释的附表。对于有等级关系和注释的复杂附表，建议可以采用以下两种表示方式：一是采用嵌套的skos:Collection表示；二是只列出附表中的一级概念，需要时再由支持系统推出

所有下位概念和注释等。

需要注意的是，分类法不同附表中同一类号可能代表不同的类目，如中图法“二、世界地区表”中的“1 世界”和“三、中国地区表”中的“1 北京市”，类号都是“1”。根据SKOS的定义，skos:Collection与其成员之间的关系用skos:member属性表示，这个属性的rdfs:domain是skos:Collection，rdfs:range是skos:Concept或skos:Collection。也就是说，如果用skos:Collection来描述分类法中的附表，那么这些附表中的类目概念也是同一概念体系中的概念。而SKOS认为，按照常规同一概念体系中的两个不同的概念不应拥有同一个符号（notation），否则就不可能用这个符号来唯一地指引一个概念。因此，如果用skos:Collection来表示分类法中的附表，可能会违反SKOS中一个符号唯一表示一个概念的约定。

•skos:ConceptScheme

将附表视为独立的概念体系，可描述任意复杂度

的附表,并解决不同附表中同一类号代表不同类目的问题。但对于不能独立使用的、专用于某个类目的专类复分表(仿分表),将其描述为独立的概念体系有些勉强。

此方案不违背SKOS对概念体系的定义,即“一个SKOS概念体系可以被看作是一个或多个SKOS概念的集合”,但不太符合“一个独立的SKOS概念体系大致对应于一个独立的叙词表、分类法、主题标题表或其他知识组织系统”这句约定。并且,如果分类法整体与其附表都采用skos:ConceptScheme进行描述,它们之间就形成了一种并列关系,原有的包含和隶属关系语义会丢失。

•ckos:Auxiliary

扩展的skos:ConceptScheme的子类,可用于表示所有附属于某一独立概念体系的、不能独立使用的子概念体系(如《中图法》中的附表)。

本规范建议:可用skos:Collection表示简单的、无等级关系和注释的、类号与主表类号及其他采用skos:Collection表示的附表类号不重复的附表;用skos:ConceptScheme描述相对独立的通用复分表;用ckos:Auxiliary描述其他类号重复又无法独立使用(即必须与主表特定类目联合使用才有意义)的附表(如中图法中的专类复分表)。当采用纯SKOS版时,只能选用前两种方案。

如例3中将中图法(CLC)中的“二、世界地区表”描述为一个独立的概念体系,并描述了它与其一级大类(顶级概念)的关系,以及附表中的类目概念。

4.3 概念及概念标签的描述

4.3.1 概念的描述

中文叙词表中的叙词(我国多称为正式主题词)是

取自自然语言并经过规范化处理的、以基本概念为基础的表达文献主题的词或词组^[9]。

传统分类法中的类目是一个个表达文献、信息内容的概念。每个类目都代表具有某种共同属性的文献、信息集合。分类法的每一个类目都是一个特定的主题概念(或主题概念集合)。类目的含义在一个由上位概念、同位概念、下位概念、相关概念和类目注释构成的语义空间中进行限定。类目应当是稳定的^[9]。

在SKOS标准中,SKOS概念(SKOS concept)是一种观念(idea)或想法(notion),一个思想(thought)单位。因此,叙词表中的叙词和分类法中的类目都可以视为SKOS概念,描述为skos:Concept的实例。

作为RDF资源的SKOS概念必须用URI来表示。当描述后的CNKOS用作Linked Data目的时,一般用完整URI来表示概念;当描述后的CNKOS用作交换格式、一个CNKOS文件只包含一个概念体系、且所有的概念都拥有同样的base URI(基准URI)时,可以采用相对的URI来表示概念。

若叙词概念本身或类目概念中的类号(或类名)发生变化,且概念的内涵或外延发生了变化,则URI也应该变,此时不应视为URI不稳定。可以通过自动生成历史注释(skos:historyNote)和扩展一个映射属性(skos:relatedMatch的子属性ckos:previousMatch)链接前URI来解决^[5,6]。

SKOS本身没有为skos:Concept定义子类,当采用纯SKOS词汇进行描述时,叙词表或分类法中的特种概念也只能用skos:Concept来描述,其特殊性会丢失。如想保留,可采用上一节中建议的方式进行描述。

描述示例见例1至例3,其三元组格式形如“<概念URI> rdf:type skos:Concept.”。

例 3

```
<2WorldRegionTable> rdf:type skos:ConceptScheme; skos:prefLabel "二、世界地区表";
skos:hasTopConcept <1 世界>, <2 中国>, ……; <7 美洲>.
<1 世界> rdf:type skos:Concept; skos:notation "1";
skos:inScheme <2WorldRegionTable>.
……
```

4.3.2 概念的语言标签的描述

SKOS将概念（或其他资源）的语言形式视为语言标签，即一个UNICODE字符串。可用skos:prefLabel和skos:altLabel分别描述概念的首选标签和可选标签。对于叙词表而言，概念的首选标签就是叙词本身，其拼音、英译名和入口词（非叙词）都可以描述为相应语种的

的skos:altLabel值；对于分类法而言，概念的首选标签是类名，而类名的拼音及其他语种类名可以描述为相应语种的skos:altLabel值。对于容易写错的叙词或类名，其错误形式可以用skos:hiddenLabel进行描述，以起到一定的入口作用。
例4展示了中图法（CLC）中类目概念的语言标签（类名）的描述方法。

例 4
<B 哲学、宗教> rdf:type skos:Concept; skos:prefLabel "哲学、宗教".

例5展示了军用电子分类表中类目概念的不同语种类名的描述方法。

例 5
< K1541 后勤训练> rdf:type skos:Concept; skos:prefLabel "后勤训练"; skos:altLabel "hou qin xun lian" @zh-pinyin; skos:altLabel " Logistical training" @en.

例6展示了汉语主题词表（CT）中叙词概念的不同语种语言标签（首选标签，入口词，首选标签的汉语拼音，首选标签的英译名）的描述方法。

例 6
<宗教> rdf:type skos:Concept; skos:prefLabel "宗教"; skos:altLabel "zong jiao" @zh-pinyin; skos:altLabel "Religion" @en.

建议不描述入口词的拼音，否则将无法分辨若干拼音形式的语言标签与中文形式的语言标签之间的对应关系。唯一的语种为汉语拼音的skos:altLabel值默认对应skos:prefLabel中的中文语种标签。
出于对类名叙词化发展趋势（如军用电子分类表中类名的描述需求）以及支持系统实现便利方面的考虑，本规范对叙词概念和类目概念的语言标签采用了统一的描述方式。但严格来讲，一个类目的完整显示应该是类号+类名，即skos:notation+空格+skos:prefLabel，它们的完整出现才代表一个唯一的类目。一

般情况下，单独的类名不能代表一个类目。因此，需要通过dc:type（词表类型，参见文献[5]第一篇表1-7-1）来区分不同的KOS类型，以使支持系统可以选择不同的显示方案（对于不同的KOS类型，同样的类和属性在显示时的语言标签也可能不同）。

4.4 概念的标记符号的描述

在分类法中，类号是类目的标记符号或代号，它用号码表示类目的含义，决定类目在分类体系中的排列

位置,表达类目之间的关系。在分类法主表中,类号具有唯一性,即一个类号只能代表一个类目。附表(复分表,仿分表)中的类号需与主表类号组配使用,不能独立使用。同一分类法所包含的不同附表中的类号可能重复,即同一个类号在不同附表中代表不同的类目。

SKOS规定,标记符号(notation)是一个字符串(如“T58.5”或“303.4833”),用于唯一地标识一个给定的概念体系范围内的一个概念。一个概念可以有0个、1个或多个标记符号(来自同一个或不同的符号系统)。但来自同一个概念体系中的两个概念不能拥有相同的标记符号,否则将不可能用这个标记符号唯一地指向一个概念(即,标记符号将有多义性)。

显然,分类法主表中的类号可以用skos:notation来描述。如果将附表中的类目视为与主表类目同属于一个概念体系,那么不同附表中的重复类号将会违反SKOS的“标记符号唯一地标识一个给定的概念体系范围内的一个概念”的原则。因此,在不同附表中的类目概念存在类号重复的情况时,需要将附表视为独立的概念体系

(如准备独立使用的通用复分表)或子概念体系(如不能独立使用的通用复分表、专类复分表和仿分表,采用CNKOS扩展版时可用)。此时,用skos:notation来描述附表中的类号将不再违反SKOS的原则。

叙词表中叙词概念对应的分类号或范畴号,本身并不是叙词概念的唯一标记符号,而是与某一分类体系中的类目概念的对应关系,如中图法类号对应的是中图法中的相应类目,范畴号对应的是叙词表本身带有的分类索引(范畴表)中的类目。而且,在叙词表中,同一个分类号或范畴号一般对应于多个叙词概念,并不具有唯一性。因此,本规范建议:叙词表中叙词概念对应的分类号或范畴号,不用skos:notation来描述,而采用SKOS的映射属性进行描述。此时,对应的分类法或分类索引(范畴表)应描述为独立的概念体系。同时,可以采用国际上已有的分类号元数据来描述一些常用的分类号,如DC Terms的UDC、LCC、DDC元数据。

例7中用skos:notation描述了CLC主表中的类目的类号。

例 7
<pre><B 哲学、宗教> rdf:type skos:Concept; skos:prefLabel "哲学、宗教"; skos:notation "B".</pre>

例8展示了CLC中附表(通用复分表“二、世界地区表”)中类目的类号的描述方法(参见例3)。

例 8
<pre><2WorldRegionTable> rdf:type skos:ConceptScheme. <1 世界> rdf:type skos:LocationConcept; skos:inScheme <2WorldRegionTable>; skos:notation "1". <2 中国> rdf:type skos:LocationConcept; skos:inScheme <2WorldRegionTable>; skos:notation "2".</pre>

注:采用纯SKOS版时,上例中的skos:LocationConcept应替换为skos:Concept。

规范代码表,即每个术语(概念)都拥有唯一规范代码的术语列表,也可以仿照分类法的类目描述方式进行描述:每一条规范记录视为一个SKOS概念(类似于类目概念),用skos:Concept表示;规范代码类似于类

号,用skos:notation表示;相对应的术语类似于类名,用skos:prefLabel表示。

目前国内外元数据标准中还没有中图法分类号等国内常用分类号的元数据元素,建议相关部门在制定或

引进元数据标准(如DC Terms)时,增加CLC等国内常用分类号的元数据元素,使它们能够像UDC、LCC、DDC那样,广泛参与国际间交流。OntoThesaurus中为国内常用的两种分类号定义了相应的属性: ont:CLC和 ont:LCCAS^[17],如有需要也可以使用。

4.5 概念间语义关系的描述

中文叙词表中叙词概念之间的语义关系包括属(S,即上位词)、分(F,即下位词)、族(Z,即族首词)、参(C,即相关词)关系。用代关系在SKOS中被处理为叙词概念与两个语言标签之间的关系,详见第4.3节。在叙词表中,叙词概念之间的等级关系(即属、分关系)是经过严格控制的,一般具有传递性。但在叙词款目中,一般只包含直接上下位关系词。

传统分类法(体系分类法)中类目概念之间的语义关系主要包括上位类、下位类和相关类关系。交替类目与正式类目之间的关系可视为类目概念之间的一种等同关系(需扩展SKOS进行描述,参见文献[5][6])。在分类法中,同位类也是一种比较重要的类目间关系,指与某一类目具有同一个上位类的其他类目,它们能够通过上(下)位类关系推理得出,可以不作明确的描述声明。

分类法中类目概念的含义在一个由上位概念、同位概念、下位概念、相关概念和类目注释构成的语义空间中进行限定。分类法中的类目注释种类繁多,其中隐含了大量的语义关系,包括相关关系,CNKOS扩展版扩展了一些关系属性可将其明确揭示出来,参见文献[5][6]。

SKOS的语义关系(semantic relation)是SKOS概念之间的链接(link),这种链接是两个被链接概念的含义中固有的。SKOS区分两种基本的语义关系种类:等级(hierarchical)和相关(associative)。两个概念之间的等级链接指明一个概念(broader)在某一方面比另一个概念(narrower)更全面(或概括、广义,general)。两个概念之间的相关链接指明两个概念是内在“相关的”(related),但其中一个并不比另一个更全面。

SKOS用skos:broader和skos:narrower来声明两个SKOS概念之间的直接等级链接。三元组“<A> skos:broader ”断言:这个三元组的客体,是三元组的主体<A>的上位概念。类似地,三元组“<C> skos:narrower <D>”断言:这个三元组的客体<D>,是三元组的主体<C>的下位概念。

按照常规,skos:broader和skos:narrower只用来声明两个SKOS概念之间直接的等级链接。这给了应用程序一种方便而可靠的方式来访问任意一个给定概念的直接上位或下位链接。这两个属性没有被声明为传递属性。

一些应用程序需要使用概念之间的直接和间接等级链接,例如通过查询扩展提高搜索的查全率(recall)。为此目的,SKOS提供了属性skos:broaderTransitive和skos:narrowerTransitive。三元组“<A> skos:broaderTransitive ”表示了一个直接或间接的等级链接,其中是<A>的一个上位“祖先”(ancestor)。类似地,三元组“<C> skos:narrowerTransitive <D>”也表示了一个直接或间接的等级链接,其中<D>是<C>的一个下位“后代”(descendant)。

但按照SKOS的规定,skos:broaderTransitive和skos:narrowerTransitive不用来进行声明(断言)。这两个属性用来推理等级链接的传递闭包(transitive closure),该传递闭包可以用来访问概念之间的直接或间接等级链接。

属性skos:related用来声明两个SKOS概念之间的相关链接,它是对称属性。

中文叙词表和分类法中的等级关系和相关关系显然可以采用以上SKOS的语义关系属性进行描述。

在传统的中文叙词表(主题词表)中,叙词的族首词指的是叙词概念与其所属词族等级的顶级概念之间的关系,即应描述为两个SKOS概念之间的关系。SKOS中的skos:hasTopConcept和skos:topConceptOf属性揭示的是概念体系与其包含的顶级概念之间的关系(参见4.1节),因此如果用SKOS的这两个属性来描述叙词表中的族首词关系,语义上会有所改变。CNKOS规范提供两种选择:一是不明确声明叙词款目中叙词概念的族首词关系,需要时通过上下位关系属性推理得出;二是使用扩展的owl:ObjectProperty的实例ckos:topConcept来描述传统叙词表叙词款目中的族首词关系,以方便从任一叙词出发直接获取其族首词。这个扩展属性可与skos:hasTopConcept和skos:topConceptOf同时使用,表达相应的语义。鉴于查看某一叙词款目时需要经常通过族首词显示整个词族等级,使用第二种方案,即明确揭示叙词概念与其族首词之间的关系,可以大大减少推理的负担。采用纯SKOS版时则只能选用第一种方案。

例9展示了CLC中类目概念的上位类的描述方法。

例 9
<pre><B 哲学、宗教> rdf:type skos:Concept; skos:prefLabel "哲学、宗教". <B9 宗教> rdf:type skos:Concept; skos:prefLabel "宗教"; skos:broader <B 哲学、宗教>.</pre>

例10展示了《中国分类主题词表》的主题词表部分 (CCT_CT) 中叙词概念的语义关系的描述方法。

例 10
<pre><马克思主义哲学> rdf:type skos:Concept; skos:broader <马克思主义三个组成部分>, <哲学>; skos:narrower <辩证唯物主义>, <历史唯物主义>, <唯物辩证法>; ckos:topConcept <马克思主义三个组成部分>, <哲学>; (采用纯SKOS版时略) skos:related <辩证逻辑>, <自然辩证法>.</pre>

未来需要的进一步扩展及思考: 如果要支持ISO 25964-1中扩展的种属 (generic)、实例 (instance) 和部分-整体 (part-whole) 这三种子等级关系的语义描述, 则需要扩展新的子属性。请参见OntoThesaurus的相应定义^[17,18]。

CNKOS规范是基于SKOS的扩展版本, 因此尽量选择使用SKOS的原有定义。SKOS中定义的以上四种等级关系属性对于现有的传统中文KOS中的语义关系描述是基本适用的。有扩展描述需求的高受控词表宜采用OntoThesaurus进行描述。

4.6 注释的描述

中文叙词表中的注释一般分为含义注释、用法注释和历史注释, 注释种类相对较少, 表现形式也比较统一, 一般可以找到对应的SKOS注释属性进行描述。ISO25964-1^[19]中的note种类与SKOS的注释属性种类基本对应。

传统分类法中的注释虽然一般分为含义范围注释、使用说明注释和沿革注释, 与叙词表中的注释类型基本对应, 但实际上可细分的注释种类很多, 表现形式各

异, 而且还隐含了很多类目与类目之间、类目与附表之间、类目与主题之间的链接关系。这些注释对类目的正确使用至关重要, 是类目概念款目中不可缺少的组成部分。如果只采用SKOS现有的注释属性进行描述, 有些注释在语义上会有所缺失, 而且隐含的语义关系也只能供人阅读, 机器无法理解。因此, CNKOS规范建议进行一些必要的扩展, 以明确揭示不同注释的语义和隐含的链接关系^[5,6]。

对于对注释没有细分要求 (即对各种注释的处理无区别, 只显示给人看) 的KOS, 如中文叙词表, 建议选用skos:note (一般注释)、skos:definition (定义注释)、skos:scopeNote (含义范围注释)、skos:historyNote (历史注释) 这几个属性描述相应的注释类型。其中最常用的是skos:scopeNote。

对于对注释有细分要求的KOS (即需要支持系统对不同的注释类型区别对待, 如CLC, 军用电子分类表等), 则建议根据需要选用skos:note及其相应子属性, 甚至扩展skos:note (或其子属性) 的子属性进行描述。

当采用纯SKOS版时, 只能选用SKOS已有的注释属性描述分类法中的注释, 具体对应关系请参见

CNKOS规范^[5]中的表3。

例11展示了CLC中类目注释的描述方法。

例 11
<pre><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。", "宗教入B9。", "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:note "依总论复分表分，-0理论与方法论所属类目入B0。"; <B9 宗教> rdf:type skos:Concept; skos:prefLabel "宗教"; skos:broader <B 哲学、宗教>; skos:note "依总论复分表分。"; skos:historyNote "<3版类名：无神论、宗教>:。</pre>

4.7 不同概念体系中概念之间映射关系的描述

SKOS采用映射属性（mapping property）来声明不同概念体系中SKOS概念之间的映射（对齐）链接，这些链接是被链接概念的含义中内在固有的。

SKOS的映射属性包括skos:mappingRelation及其子属性：skos:closeMatch（包含子属性skos:exactMatch）、skos:broadMatch、skos:narrowMatch和skos:relatedMatch。

属性skos:broadMatch和skos:narrowMatch用来声明两个概念之间的一个等级映射链接。属性skos:relatedMatch则用来声明两个概念之间的一个相关映射链接。

属性skos:closeMatch用来链接两个足够相似的概念，它们可以在某些信息检索应用程序中交换使用。为了避免当组合跨越两个以上概念体系的映射时出现“复合错误”（compound errors）的可能性，skos:closeMatch没有被声明为是一个传递属性。

属性skos:exactMatch用来链接两个概念，表明了一种高度的信心：这两个概念可以在很大范围的信息检索应用程序之间交换使用。它是一个传递属性，而且是skos:closeMatch的子属性。

在国内的受控词表中，不同分类法的类目概念之间，不同叙词表的叙词（正式主题词）概念之间，以及分类法的类目概念与叙词表的叙词概念之间，都存在概念含义中固有的映射关系。我们可以采用以上所述

的SKOS的映射属性对它们进行描述。具体使用哪一个属性需要根据具体情况而定。

例如，在中国分类主题词表（CCT）中，类目与主题词之间的对应关系分为三种类型：（1）直接对应；（2）间接对应；（3）非主要类目对应。

第一种类型是直接的对应，建议统一使用skos:closeMatch来描述。如果选用skos:exactMatch，需要人工进一步确认其可交换使用的程度，因为即使类名与主题词字面上完全一致，类目的涵盖范围也未必和主题词完全相同（类目的含义是在一个由上位概念、同位概念、下位概念、相关概念和类目注释构成的语义空间中进行限定的）。

第二种类型是一种间接的对应，可能是除skos:closeMatch（和skos:exactMatch）之外的任何一种映射关系（skos:broadMatch、skos:narrowMatch和skos:relatedMatch）。如果要统一描述，现有的SKOS映射属性似乎没有完全适用的。若统一使用skos:mappingRelation来描述这种间接对应关系，又体现不出主次对应关系的区别。CNKOS规范建议，扩展一个与skos:closeMatch对应的映射属性skos:nonCloseMatch，它是skos:mappingRelation的子属性，用于描述skos:closeMatch（和skos:exactMatch）之外的所有关系^[5,6]。

第三种类型是用竖线标识的与非主要类目对应的主题词，建议选用skos:relatedMatch进行描述。

CCT_CT（主题词—分类号对应表）将主题词对应的分类号区分为等同对应类号、主要类号、次要类

号、交替等同对应类号、交替类号。其中主、次、交替是从类号的使用角度进行区分的,主要类号用作排架类号,次要类号不用于排架,但可用于检索,交替类号则作为专业机构的一种选择。一个类号可以同时是等同对应类号和主类号。因此建议第一步仍按直接(`skos:closeMatch`)、间接(`ckos:nonCloseMatch`)和非主要类目对应(`skos:relatedMatch`)的方式对它们分别进行描述,以便与上述类目-主题词对应关系的描述双向统一。例如:等同对应类号、交替等同对应类号、主要类号(直接对应)和交替类号(直接对应)均用`skos:closeMatch`描述,间接对应的主要类号和交替类号用`ckos:nonCloseMatch`描述,次要类号用`skos:relatedMatch`描述。映射的类目概念是否是交替类目,在该类目概念本身的描述中会进行声明。

当采用纯SKOS版时,上面所说的间接对应需明确区分`skos:broadMatch`、`skos:narrowMatch`和`skos:relatedMatch`,或暂时先采用`skos:mappingRelation`统一描述,之后再逐步细化描述。

5 结语

因篇幅所限,本文不能提供示例的SKOS详细代码。深圳大学图书馆NKOS研究室已经实现了OTCSS Linked Data服务,目前提供CCT一版全部数据^[16,20]和CLC四版全部数据^[15]的Linked Data服务,可在线检索下载所需叙词概念和类目概念的SKOS和CNKOS格式语义描述代码(可选择RDF/XML、N-Triples和JSON格式)。欢迎使用并多提宝贵意见。

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。

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OntoThesaurus版及OWL应用规范四个部分。敬请关注本系列论文的后续部分以及深圳大学图书馆NKOS研究室网站(<http://nkos.lib.szu.edu.cn>)。本研究室致力于中文NKOS的公益性研究和服务,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.
- [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.
- [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.
- [4] 王军,曾新红,欧石燕.国家图书馆知识组织标准规范[R].国家图书馆,2011.
- [5] 王军,卜书庆.网络环境下的知识组织规范和应用指南[M].国家图书馆出版社,2012.
- [6] 曾新红.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[J].中国图书馆学报,2012(3):57-68.
- [7] W3C. SKOS Simple Knowledge Organization System Reference: W3C Recommendation 18 August 2009 [EB/OL]. [2010-02-23]. <http://www.w3.org/TR/2009/REC-skosreference-20090818/>.
- [8] W3C. SKOS Simple Knowledge Organization System Primer: W3C Working Group Note 18 August 2009 [EB/OL]. [2010-02-23]. <http://www.w3.org/TR/2009/NOTE-skosprimer-20090818/>.
- [9] 戴维民.信息组织[M].2版.北京:高等教育出版社,2009.
- [10] 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.
- [11] 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].国家质量技术监督检验检疫总局.
- [12] 国家图书馆《中国图书馆分类法》编辑委员会.中国分类主题词表:第二版[M].北京:图书馆出版社,2005.
- [13] 中国图书馆分类法编辑委员会.中国图书馆分类法[M].4版.北京:图书馆出版社,1999.
- [14] 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[M].中国人民解放军总装备部.
- [15] 深圳大学图书馆NKOS研究室. CLC Linked Data服务[EB/OL]. [2012-12-25]. http://nkos.lib.szu.edu.cn/CCT_CLC_V4.0.
- [16] 深圳大学图书馆NKOS研究室. CCT1 Linked Data服务[EB/OL]. [2011-01-16]. http://nkos.lib.szu.edu.cn/CCT_CT_V1.0.
- [17] 曾新红.中文叙词表本体OntoThesaurus词汇总表[EB/OL]. [2010-11-11]. <http://nkos.lib.szu.edu.cn/2010/10/ont/>.
- [18] 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.
- [19] ISO. ISO/CD 25964-1, Information and documentation -- Thesauri

and interoperability with other vocabularies -- Part 1: Thesauri for information retrieval [S]. ISO, 2008.

[20] 黄华军,曾新红,林伟明.OTCSS关联数据服务的研究与实现[J].现代图书情报技术,2012(7/8):40-47.

作者简介

曾新红,女,1968年生,硕士,深圳大学图书馆NKOS研究室研究馆员。E-mail: zengxh@szu.edu.cn。

SKOS Specializations and Guides for Chinese Controlled Vocabularies

ZENG XinHong

(NKOS Research Office, Shenzhen University Library, Shenzhen 518060, China)

Abstract: Based on the authoritative theory and building standards for information retrieval languages (such as thesauri and Classification Schemes) in China, and the preparation instructions for Chinese Classification Thesaurus, Chinese Thesaurus, Chinese Library Classification, and the Guidelines for establishment of military electronic classification schemes, as well as the existing research at home and abroad, the SKOS specifications for Chinese Controlled Vocabularies are studied and formulated with the SKOS standards as the criterion, from the point of view of the easy sharing and implementation. The application guides are also given with examples.

Keywords: Controlled vocabulary; Classification scheme; Thesaurus; Formal representation; Semantic description; SKOS; CNKOS; KOS

(收稿日期: 2014-04-04)

■ 书 讯 ■

《国家学位论文资源管理与共享系统研究》

我国学位论文法定收藏工作已正式确立30周年。为进一步推进学位论文资源建设与共享服务,中国科学技术信息研究所所长贺德方研究员牵头组织赵嘉朱、姜爱蓉、陈传夫、张建勇、曾建勋五位专家学者对近十年来相关联合研究课题的成果进行归纳、凝练、整理和补充,形成了《国家学位论文资源管理与共享系统研究》这部学术专著。

本书总结了我国学位论文的管理历程,凝练了我国学位论文的管理建议,归纳了学位论文资源的整合方案,提出了学位论文资源的共享模式,展望了学位论文的深层次服务目标。对于全社会整体把握学位论文收藏格局,系统构建学位论文管理体系有指导作用;对于高校和研究机构的学位管理人员强化学位论文质量管理和过程监控,做好学位论文存交,推进学位论文管理和机构知识库建设有参考作用;对于图书信息机构做好学位论文管理和收藏,推进学位论文资源组织和保障利用有引导作用;也可作为高校信息管理等专业研究生学习的参考书。

《国家学位论文资源管理与共享系统研究》于2014年3月由科学技术文献出版社出版,全书27万字,定价68.00元。