

中文高受控词表的 OntoThesaurus 描述规范*

■ 曾新红

[摘 要] 针对中文叙词表等高受控词表的描述需求以及向细粒度本体的演化趋势,在已有中文叙词表本体研究成果的基础上,基于 OWL DL 和 OWL2 研究制定高受控词表的 OntoThesaurus 描述规范,确定独立的命名域及其缩写以及一整套严格的形式化定义。根据语义 Web 的应用环境,编写 OntoThesaurus 的 RDF Schema,以 RDF/XML 语法的方式向机器用户提供其中的类和属性的使用。说明 OntoThesaurus 不同于 SKOS 和 CNKOS 的特殊定义,给出描述示例以及三者之间的词汇对应关系。最后简单介绍 OntoThesaurus 支持系统的功能。

[关键词] 中文叙词表本体 高受控词表 OWL DL OWL2 形式化表示 语义描述 知识组织系统

[分类号] G254 TP18

DOI: 10.7536/j.issn.0252-3116.2013.17.021

1 引言

笔者对于中文叙词表本体的研究始于 2004 年。一次偶然的机会,笔者了解到“本体”的概念与叙词表是如此相似,通过查找资料发现了刚刚发布为 W3C 推荐标准的 OWL^[1-5]以及国外在词表转换方面的研究。凭着多年编目工作中积累下来的对《中国分类主题词表》(CCT)的了解,笔者研究了将 CCT 转换为 OWL 的可能性^[6],随后在国家自然科学基金支持下,将这一设想进一步细化为形式化表示的中文叙词表本体 (OntoThesaurus)^[7],并实现了其支持系统“中文叙词表本体共建共享系统”(OTCSS)^[7-9]。通过多年的努力,目前的 OTCSS 系统已经具备了一致性检测^[10]、共建共享^[8,11]、Web Service 服务^[12]、Linked Data 服务^[13-14]、可视化^[15]、自动构建等多项功能,成为一个比较成熟的系统。

本研究进一步将 OntoThesaurus 提升为符合语义 Web 规范的描述语言形式,拥有独立的命名域、缩写和严格的形式化定义,以便更好地融入语义 Web 应用环境,并根据 CNKOS^[16]研究过程中发现的新问题,对少许类和属性进行了调整。

2 国内外研究现状

之前的“中文叙词表本体”系列成果论文已对本

研究领域的国内外现状做了比较详细的介绍,在此不再重复。需要强调的是,在 SKOS^[17-18]与 OWL 之间缺乏一种适合高受控词表使用的描述语言。SKOS 过于松散,无法表示某些控制语义,实现严格的一致性检测。而 OWL 一般面向领域本体,每一个概念都表示为 owl:Class 的实例,每一个 OWL 文件都需要专门的支持 (TBox 级别),不直接提供像 skos:Concept 的实例这种 (ABox) 级别的统一支持和处理。OntoThesaurus 处在这两者之间,高受控词表中的每一个概念都描述为 ont:Concept 的实例,每一个语义元素都有严格的形式化定义,基于此可以提供面向所有高受控词表的统一的支持系统,并实现严格的一致性检测等服务。

3 OntoThesaurus 总体介绍

OntoThesaurus 的命名域 (namespace) 缩写为“ont”,所采用的 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://nkos.lib.szu.edu.cn/2010/10/ont#	ont:
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:

* 本文系国家自然科学基金项目“中文知识组织系统的形式化语义描述标准体系研究”(项目编号 12BTQ045)和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”(项目编号:GD10CTS02)研究成果之一。

[作者简介] 曾新红,深圳大学图书馆研究馆员,NKOS 研究室主任,硕士,E-mail: zengxh@szu.edu.cn。

收稿日期:2013-07-15 修回日期:2013-08-20 本文起止页码:114-120 本文责任编辑:王善军

OntoThesaurus 词汇表以及每个词汇的说明和格式的形式化定义请见 OntoThesaurus 的 RDF “prose”形式的形式化定义详见表 2,完整的 RDF/XML Schema^[19]。

表 2 OntoThesaurus 词汇

URI	说 明	定 义
ont: Concept	概念(类): 高受控词表中所有概念(如主题词表中的正式主题词)都是这个类的 individual(个体)。	定义 O1: ont: Concept 是 owl: Class 的实例。它刚好拥有一个 ont: pinyin 值。ont: Concept 与 ont: NTerm 不相交(disjoint)。
ont: CompoundConcept	复合概念(类): 是 ont: Concept 类的子类,用于描述两个以上 ont: Concept 实例的组配(如主题词表中的主题词串)。	定义 O2: ont: CompoundConcept, ont: GeneralConcept, ont: PersonConcept, ont: LocationConcept, ont: OrganizationConcept, ont: TimeConcept, ont: NationalityConcept 是 ont: Concept 的子类。 定义 O3: ont: WorldRegionConcept, ont: ChinaRegionConcept 是 ont: RegionConcept 的子类。 定义 O4: ont: WorldEraConcept, ont: ChinaEraConcept 是 ont: TimeConcept 的子类。 定义 O5: ont: ChinaNationalityConcept 是 ont: NationalityConcept 的子类。
ont: GeneralConcept	一般通用概念(类)	
ont: PersonConcept	人物概念(类)	
ont: LocationConcept	地名概念(类)	
ont: WorldRegionConcept	世界地名概念(类)	
ont: ChinaRegionConcept	中国地名概念(类)	
ont: OrganizationConcept	机构概念(类)	
ont: TimeConcept	时间概念(类)	
ont: WorldEraConcept	世界时代概念(类)	
ont: ChinaEraConcept	中国时代概念(类)	
ont: NationalityConcept	民族概念(类)	
ont: ChinaNationalityConcept	中国民族概念(类)	
ont: NTerm	非正式主题词(非叙词)(类)	定义 O6: ont: NTerm 是 owl: Class 的实例。
ont: hasNTerm	入口词(对象属性)	定义 O7: ont: hasNTerm 是 owl: ObjectProperty 的实例。其 rdfs: domain 是 ont: Concept, rdfs: range 是 ont: NTerm。
ont: broader	上位词(对象属性)	定义 O8: ont: broader 是 owl: ObjectProperty 的实例。其 rdfs: domain 和 rdfs: range 均为 ont: Concept。它具有传递性、非自反性和非对称性。它是 ont: narrower 的逆属性。
ont: broaderGeneric	类属关系上位词(对象属性)	定义 O9: ont: broaderGeneric, ont: broaderInstance 和 ont: broaderPart 都是 ont: broader 的子属性(subproperty)。ont: broaderGeneric 是 ont: narrowerGeneric 的逆属性, ont: broaderInstance 是 ont: narrowerInstance 的逆属性, ont: broaderPart 是 ont: narrowerPart 的逆属性。
ont: broaderInstance	实例关系上位词(对象属性)	
ont: broaderPart	整体/部分上位词(对象属性)	
ont: narrower	下位词(对象属性)	定义 O10: ont: narrower 是 owl: ObjectProperty 的实例。其 rdfs: domain 和 rdfs: range 均为 ont: Concept。它具有传递性、非自反性和非对称性。
ont: narrowerGeneric	类属关系下位词(对象属性)	定义 O11: ont: narrowerGeneric, ont: narrowerInstance 和 ont: narrowerPart 都是 ont: narrower 的子属性(subproperty)。
ont: narrowerInstance	实例关系下位词(对象属性)	
ont: narrowerPart	整体/部分下位词(对象属性)	
ont: topConcept	族首词(对象属性)	定义 O12: ont: topConcept 是 owl: ObjectProperty 的实例。其 rdfs: domain 和 rdfs: range 均为 ont: Concept。它具有非自反性和非对称性。
ont: related	参见(对象属性)	定义 O13: ont: related 是 owl: ObjectProperty 的实例。其 rdfs: domain 和 rdfs: range 均为 ont: Concept。具有对称性和非自反性。
ont: cause_Effect	原因/结果相关词(对象属性)	定义 O14: ont: cause_Effect 是 ont: related 的子属性(subproperty),但不继承其对称性。它是 ont: effect_Cause 的逆属性。具有非自反性和非对称性。
ont: effect_Cause	结果/原因相关词(对象属性)	定义 O15: ont: effect_Cause 是 ont: related 的子属性,但不继承其对称性。它具有非自反性和非对称性。
ont: process_Agent	处理/工具相关词(对象属性)	定义 O16: ont: process_Agent 是 ont: related 的子属性,但不继承其对称性。它是 ont: agent_Process 的逆属性。具有非自反性和非对称性。
ont: agent_Process	工具/处理相关词(对象属性)	定义 O17: ont: agent_Process 是 ont: related 的子属性,但不继承其对称性。它具有非自反性和非对称性。
ont: process_CounterAgent	处理/反工具相关词(对象属性)	定义 O18: ont: process_CounterAgent 是 ont: related 的子属性,但不继承其对称性。它是 ont: counterAgent_Process 的逆属性。具有非自反性和非对称性。
ont: counterAgent_Process	反工具/处理相关词(对象属性)	定义 O19: ont: counterAgent_Process 是 ont: related 的子属性,但不继承其对称性。它具有非自反性和非对称性。
ont: action_Product	行为/产品相关词(对象属性)	定义 O20: ont: action_Product 是 ont: related 的子属性,但不继承其对称性。它是 ont: product_Action 的逆属性。具有非自反性和非对称性。
ont: product_Action	产品/行为相关词(对象属性)	定义 O21: ont: product_Action 是 ont: related 的子属性,但不继承其对称性。它具有非自反性和非对称性。

(续表 2)

URI	说 明	定 义
ont: action_Property	行为/属性相关词(对象属性)	定义 O22: ont: action_Property 是 ont: related 的子属性,但不继承其对称性。它是 ont: property_Action 的逆属性。具有非自反性和非对称性。
ont: property_Action	属性/行为相关词(对象属性)	定义 O23: ont: property_Action 是 ont: related 的子属性,但不继承其对称性。它具有非自反性和非对称性。
ont: action_Target	行为/目标相关词(对象属性)	定义 O24: ont: action_Target 是 ont: related 的子属性,但不继承其对称性。它是 ont: target_Action 的逆属性。具有非自反性和非对称性。
ont: target_Action	目标/行为相关词(对象属性)	定义 O25: ont: target_Action 是 ont: related 的子属性,但不继承其对称性。它具有非自反性和非对称性。
ont: conOrObj_Property	概念或物体/性质相关词(对象属性)	定义 O26: ont: conOrObj_Property 是 ont: related 的子属性,但不继承其对称性。它是 ont: property_ConOrObj 的逆属性。具有非自反性和非对称性。
ont: property_ConOrObj	性质/概念或物体相关词(对象属性)	定义 O27: ont: property_ConOrObj 是 ont: related 的子属性,但不继承其对称性。它具有非自反性和非对称性。
ont: conOrObj_Origins	概念或物体/来源相关词(对象属性)	定义 O28: ont: conOrObj_Origins 是 ont: related 的子属性,但不继承其对称性。它是 ont: origins_ConOrObj 的逆属性。具有非自反性和非对称性。
ont: origins_ConOrObj	来源/概念或物体相关词(对象属性)	定义 O29: ont: origins_ConOrObj 是 ont: related 的子属性,但不继承其对称性。它具有非自反性和非对称性。
ont: conOrObj_Measure	概念或物体/度量单位或机制相关词(对象属性)	定义 O30: ont: conOrObj_Measure 是 ont: related 的子属性,但不继承其对称性。它是 ont: measure_ConOrObj 的逆属性。具有非自反性和非对称性。
ont: measure_ConOrObj	度量单位或机制/概念或物体相关词(对象属性)	定义 O31: ont: measure_ConOrObj 是 ont: related 的子属性,但不继承其对称性。它具有非自反性和非对称性。
ont: rMaterial_Product	原材料/产品相关词(对象属性)	定义 O32: ont: rMaterial_Product 是 ont: related 的子属性,但不继承其对称性。它是 ont: product_RMaterial 的逆属性。具有非自反性和非对称性。
ont: product_RMaterial	产品/原材料相关词(对象属性)	定义 O33: ont: product_RMaterial 是 ont: related 的子属性,但不继承其对称性。它具有非自反性和非对称性。
ont: discOrField_ObjOrPrac	学科或领域/对象或从业者相关词(对象属性)	定义 O34: ont: discOrField_ObjOrPrac 是 ont: related 的子属性,但不继承其对称性。它是 ont: objOrPrac_DiscOrField 的逆属性。具有非自反性和非对称性。
ont: objOrPrac_DiscOrField	对象或从业者/学科或领域相关词(对象属性)	定义 O35: ont: objOrPrac_DiscOrField 是 ont: related 的子属性,但不继承其对称性。它具有非自反性和非对称性。
		定义 O36: ont: hasNTerm, ont: broader, ont: broaderGeneric, ont: broaderInstance, ont: broaderPart, ont: narrower, ont: narrowerGeneric, ont: narrowerInstance, ont: narrowerPart, ont: related, ont: cause_Effect, ont: effect_Cause, ont: process_Agent, ont: agent_Process, ont: process_CounterAgent, ont: counterAgent_Process, ont: action_Product, ont: product_Action, ont: action_Property, ont: property_Action, ont: action_Target, ont: target_Action, ont: conOrObj_Property, ont: property_ConOrObj, ont: conOrObj_Origins, ont: origins_ConOrObj, ont: conOrObj_Measure, ont: measure_ConOrObj, ont: rMaterial_Product, ont: product_RMaterial, ont: discOrField_ObjOrPrac, ont: objOrPrac_DiscOrField 两两不相交。
ont: CLC	《中图法》分类号(数据类型属性)	定义 O37: ont: CLC 是 owl: DataTypeProperty 的实例。
ont: LCCAS	《科图法》分类号(数据类型属性)	定义 O38: ont: LCCAS 是 owl: DataTypeProperty 的实例。
ont: pinYin	汉语拼音(数据类型属性)	定义 O39: ont: pinYin 是 owl: DataTypeProperty 的实例。
ont: engCounterpart	英译名(数据类型属性)	定义 O40: ont: engCounterpart 是 owl: DataTypeProperty 的实例。
ont: scopeNote	范畴注释(数据类型属性)	定义 O41: ont: scopeNote 是 owl: DataTypeProperty 的实例。

注:表中《中图法》为《中国图书馆分类法》的简称,《科图法》为《中国科学院图书馆分类法》的简称。

本规范适用于需要严格控制的中文叙词表(主题词表)、规范档(名称、地名、机构)等,并利于它们向细粒度本体的演化。

4 高受控词表的描述需求及对策

本研究根据中文高受控词表的描述需求,采用 OWL DL 及其升级版 OWL2 的建模结构制定了具体的描述规范。限于篇幅,本文着重列出其中不同于 SKOS 和 CNKOS^[16]的部分。

4.1 整体的描述

一个受控词表对应于一个本体,因此直接采用 OWL 的相应注释、本体头、导入和版本信息等属性描

述其名称及相关信息。例如:

```
<owl:Ontology rdf:about="">
  <owl:versionInfo>...</owl:versionInfo>
  <rdfs:comment>...</rdfs:comment>
  <owl:imports rdf:resource="..." />
</owl:Ontology>
```

4.2 概念的描述

高受控词表中的概念采用 ont: Concept 进行描述,附表中的专类概念采用 ont: Concept 的子类进行描述。这种方式更符合语义 Web 界的概念层次建模方式(rdfs:subClassOf),有利于广泛的集成问题的解决。以

这种方式(胜过非形式化的规范说明)为关系建模的好处是,多个规范说明的交互可以被理解甚至被自动处理^[20]。

4.3 入口词的描述

与 SKOS 将入口词描述为可选标签 (skos:altLabel) 的方式不同,OntoThesaurus 为入口词定义了专门的类 ont:NTerm 和相应的属性 ont:hasNTerm。在此基础上进一步规定了 ont:Concept 与 ont:NTerm 不相交,从而杜绝了入口词与概念同形等错误的发生。

4.4 拼音和英译名的描述

ont:pinYin 和 ont:engCounterpart 都是 OWL 数据类型属性,其 rdfs:Domain 都是 ont:Concept。对其可以进行基数等限制,如规定 ont:Concept 有且只有一个拼音属性值。

4.5 分类号的描述

最初的 OntoThesaurus 为《中图法》分类号、《科图法》分类号、UDC、DDC 和 LCC 各建立了一个数据类型属性。笔者在 CNKOS^[16]的研究过程中,发现国际上比较通用的元数据方案 DCMI Metadata Terms^[21]中已有对应的 dcterms:UDC、dcterms:DDC 和 dcterms:LCC 元数据,因此本方案删除了这几个属性,并将国内常用的《中图法》分类号和《科图法》分类号调整定义为 ont:CLC 和 ont:LCCAS,以便与国际接轨。

4.6 等级关系的描述及扩展

高受控词表中的等级关系具有传递性。OntoThesaurus 明确定义 ont:broader 和 ont:narrower 为传递属性,并参考 ANSI/NISO Z39.19-2005^[22],分别为它们定义了三个子属性:ont:broaderGeneric(类属关系上位词)、ont:broaderInstance(实例关系上位词)、ont:broaderPart(整体/部分关系上位词)以及 ont:narrowerGeneric(类属关系下位词)、ont:narrowerInstance(实例关系下位词)、ont:narrowerPart(整体/部分关系下位词)。

4.7 相关关系的描述及扩展

参考 ANSI/NISO Z39.19-2005^[22],OntoThesaurus 为相关关系定义了 11 对共 22 个比较通用的子关系属性,每对子关系属性为互逆关系。为便于理解和判断,在属性名称的设置上尽量保留了关系双方的名称信息。这些子关系的具体含义和示例可参见 ANSI/NISO Z39.19-2005^[22]。

这些子关系介于粗略的“参见”关系与各领域内概念之间最专指的具体关系之间,对于综合性的高受控词表中的各个领域的概念间关系具有通用性,因此

便于进行统一的支持处理。如果需要具体到某一领域的最具体、最专指的关系,则应该采用领域本体的描述方式,直接采用 OWL 进行描述(将在本系列研究的后续论文具体介绍)。

4.8 描述示例

以下为 CCT1 中“传声器”叙词概念的 OntoThesaurus 描述,其中的 ont:broader、ont:narrower 和 ont:related 在必要时都可以改为用更具体的子关系属性进行表示。登录 CCT1_OTCSS 的 Linked Data 服务^[14]可以获得更多示例的不同格式描述代码。

```
<? xml version = "1.0" encoding = "GBK" ? >
< rdf: RDF
  xmlns: rdf = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns: xsd = "http://www.w3.org/2001/XMLSchema#"
  xmlns: rdfs = "http://www.w3.org/2000/01/rdf-schema#"
  xmlns: owl = "http://www.w3.org/2002/07/owl#"
  xmlns: ont = "http://nkos.lib.szu.edu.cn:8080/2010/10/ont#"
  xml: base = "http://nkos.lib.szu.edu.cn:8080/CCT_CT_V1.0#" >
  < ont: Concept rdf: ID = "传声器" >
  < ont: CLC > TN641⑦ </ont: CLC >
  < ont: CLC > TU112.2+6⑧ </ont: CLC >
  < ont: pinYin > chuanshengqi </ont: pinYin >
  < ont: topConcept rdf: resource = "http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#换能器" / >
  < ont: topConcept rdf: resource = "http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#音频设备" / >
  < ont: topConcept rdf: resource = "http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#声学测试仪器" / >
  < ont: related rdf: resource = "http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#扩声系统" / >
  < ont: related rdf: resource = "http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#立体声录音机" / >
  < ont: related rdf: resource = "http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#振膜" / >
  < ont: narrower rdf: resource = "http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#电磁传声器" / >
  < ont: narrower rdf: resource = "http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#测试传声器" / >
```

```
< ont: narrower rdf: resource = " http://nkos. lib. szu. edu. cn/CCT_CT_V1.0#抗噪声送话器" / >
< ont: narrower rdf: resource = " http://nkos. lib. szu. edu. cn/CCT_CT_V1.0#压电传声器" / >
< ont: narrower rdf: resource = " http://nkos. lib. szu. edu. cn/CCT_CT_V1.0#电容传声器" / >
< ont: narrower rdf: resource = " http://nkos. lib. szu. edu. cn/CCT_CT_V1.0#动圈传声器" / >
< ont: narrower rdf: resource = " http://nkos. lib. szu. edu. cn/CCT_CT_V1.0#单向传声器" / >
< ont: narrower rdf: resource = " http://nkos. lib. szu. edu. cn/CCT_CT_V1.0#送话器" / >
< ont: narrower rdf: resource = " http://nkos. lib. szu. edu. cn/CCT_CT_V1.0#无线传声器" / >
< ont: broader rdf: resource = " http://nkos. lib. szu. edu. cn/CCT_CT_V1.0#声级计" / >
< ont: broader rdf: resource = " http://nkos. lib. szu. edu. cn/CCT_CT_V1.0#声换能器" / >
< ont: broader rdf: resource = " http://nkos. lib. szu. edu. cn/CCT_CT_V1.0#电声器件" / >
< ont: hasNTerm rdf: resource = " http://nkos. lib. szu. edu. cn/CCT_CT_V1.0#话筒" / >
< ont: hasNTerm rdf: resource = " http://nkos. lib. szu. edu. cn/CCT_CT_V1.0#微音器" / >
< ont: hasNTerm rdf: resource = " http://nkos. lib. szu. edu. cn/CCT_CT_V1.0#麦克风" / >
< /ont: Concept >
< /rdf: RDF >
```

5 OntoThesaurus 的严格形式化定义与一致性检测

严格的形式化定义使 OntoThesaurus 能够实现严格的一致性检测^[10] ,不仅可以保留高受控词表原有的控制成果 ,也有利于其向细粒度的本体演化。

除充分利用 OWL DL 的各种结构(如定义域、值域、基数、传递性、对称性等)进行定义之外 ,OntoThesaurus 还采用了 OWL 的升级版本 OWL2^[23-25] 的一些结构——两两不相交(pairwise disjoint properties)、非自反性(irreflexive property)以及非对称性(asymmetric property)来进行更严格的形式化定义(2009年10月27日 OWL2 成为 W3C 推荐标准 ,深圳大学图书馆 NKOS 研究室已将其中的三个标准文本翻译成中文 ,并发布在 W3C Translations 网站(<http://www.w3.org/2005/11/Translations/Lists/ListLang-zh-hans.html>)上 ,欢迎参考使用)。

具体如下:

- 所有的对象属性(ont: topConcept 除外 ,它由支持系统根据等级关系自动推出)两两不相交(见表2定义 O36) ,即同一概念的不同关系属性不允许有相同的属性值(例如: < A > ont: broader < B > 同时 < A > ont: related < B > 是不允许的)。

- 所有的对象属性都具有非自反性(即不允许自身与自身的关系存在 ,例如: < A > ont: broader < A >)。

- 所有的对象属性(ont: related 除外)都具有非对称性(即不允许对称 ,例如: < A > ont: broader < B > 同时 < B > ont: broader < A > 是不允许的)。

以上形式化定义可有效扼制逻辑错误的出现 ,以及本体文件的无意义膨胀 ,减少系统负担和网络传输负担 ,对于大型高受控词表的形式化语义描述尤其重要。

6 OntoThesaurus 与 CNKOS、SKOS 之间的对应关系

OntoThesaurus 与 CNKOS、SKOS 之间在一定条件下可以相互转换(见表3) ,但可能会造成语义的丢失。高受控词表应首选用 OntoThesaurus 进行描述 ,以最大程度地保留原有的词表控制成果和预留扩展空间 ,并获得配套系统 OTCSS 的支持。在需要时也可以牺牲一些语义 ,转换成其他两种格式 ,参与 CNKOS 或 SKOS 层次的关联数据交流和共享。

表3 OntoThesaurus 词汇与 CNKOS 词汇、SKOS 词汇的对应关系

原 TBox 定义	OntoThesaurus 词汇	CNKOS 词汇	SKOS 词汇
Concept	ont: Concept	skos: Concept skos: prefLabel @ zh	skos: Concept skos: prefLabel @ zh
CompoundConcept	ont: CompoundConcept	ckos: CompoundConcept skos: prefLabel @ zh	skos: Concept skos: prefLabel @ zh
GeneralConcept	ont: GeneralConcept	ckos: GeneralConcept skos: prefLabel @ zh	skos: Concept skos: prefLabel @ zh
PersonConcept	ont: PersonConcept	ckos: PersonConcept skos: prefLabel @ zh	skos: Concept skos: prefLabel @ zh
RegionConcept	ont: LocationConcept	ckos: LocationConcept skos: prefLabel @ zh	skos: Concept skos: prefLabel @ zh
WorldRegionConcept	ont: WorldRegionConcept	ckos: LocationConcept skos: prefLabel @ zh	skos: Concept skos: prefLabel @ zh
ChinaRegionConcept	ont: ChinaRegionConcept	ckos: LocationConcept skos: prefLabel @ zh	skos: Concept skos: prefLabel @ zh
InstituteConcept	ont: OrganizationConcept	ckos: OrganizationConcept skos: prefLabel @ zh	skos: Concept skos: prefLabel @ zh
EraConcept	ont: TimeConcept	ckos: TimeConcept skos: prefLabel @ zh	skos: Concept skos: prefLabel @ zh
WorldEraConcept	ont: WorldEraConcept	ckos: TimeConcept skos: prefLabel @ zh	skos: Concept skos: prefLabel @ zh
ChinaEraConcept	ont: ChinaEraConcept	ckos: TimeConcept skos: prefLabel @ zh	skos: Concept skos: prefLabel @ zh

(续表 3)

原 TBox 定义	OntoThesaurus 词汇	CNKOS 词汇	SKOS 词汇
	ont: NationalityConcept	ckos: NationalityConcept skos: prefLabel @ zh	skos: Concept skos: prefLabel @ zh
ChinaNationality Concept	ont: ChinaNationality Concept	ckos: NationalityConcept skos: prefLabel @ zh	skos: Concept skos: prefLabel @ zh
NTerm	ont: NTerm	无对应类 视为语言标签	无对应类 视为语言标签
HasNTerm	ont: hasNTerm	skos: altLabel @ zh	skos: altLabel @ zh
Broader	ont: broader	skos: broader	skos: broader
BroaderGeneric	ont: broaderGeneric	skos: broader	skos: broader
BroaderInstance	ont: broaderInstance	skos: broader	skos: broader
BroaderPart	ont: broaderPart	skos: broader	skos: broader
Narrower	ont: narrower	skos: narrower	skos: narrower
NarrowerGeneric	ont: narrowerGeneric	skos: narrower	skos: narrower
NarrowerInstance	ont: narrowerInstance	skos: narrower	skos: narrower
NarrowerPart	ont: narrowerPart	skos: narrower	skos: narrower
TopConcept	ont: topConcept	ckos: topConcept	
Related	ont: related	skos: related	skos: related
Cause_Effect	ont: cause_Effect	skos: related	skos: related
Effect_Cause	ont: effect_Cause	skos: related	skos: related
.....			
CLCCode	ont: CLC 同右 增加指向相应分类 法类目的 SKOS 映射 属性	ont: CLC 增加指向相应分类法类 目的 SKOS 映射属性 如 skos: closeMatch	skos: closeMatch 根据左方的数据类型属 性值和 URI 生成模板生 成相应类目的 URI
LCCASCode	ont: LCCAS 同右 增加 SKOS 映射 属性	ont: LCCAS 增加指向相应分类法类 目的 SKOS 映射属性	同上
UDCCode	采用国际通用的词汇 determs: UDC ^[21] 同右	采用国际通用的词汇 determs: UDC ^[21] 增加指 向相应分类法类目的 SKOS 映射属性	同上
DDCCode	采用国际通用的词汇 determs: DDC ^[21] 同右	采用国际通用的词汇 determs: DDC ^[21] 增加指 向相应分类法类目的 SKOS 映射属性	同上
LCCCode	采用国际通用的词汇 determs: LCC ^[21] 同右	采用国际通用的词汇 determs: LCC ^[21] 增加指 向相应分类法类目的 SKOS 映射属性	同上
PmYin	ont: pinYin	skos: altLabel @ zh-pinyin	skos: altLabel @ zh-pinyin
EngCounterpart	ont: engCounterpart	skos: altLabel @ en	skos: altLabel @ en
ScopeNote	ont: scopeNote	skos: scopeNote	skos: scopeNote

注: OntoThesaurus 词汇→CNKOS 词汇,或 OntoThesaurus 词汇→SKOS 词汇,或 CNKOS 词汇→SKOS 词汇。反向转换也可以,但没有太大意义。语义丰富度: OntoThesaurus > CNKOS > SKOS。ont: topConcept 与 skos: hasTopConcept 之间不能直接转换^[16-26]。

7 结 语

OntoThesaurus 是适用于中文高受控词表的语义描述方案,采用其进行语义描述的中文叙词表、规范档等可以获得 OntoThesaurus 的支持系统——中文叙词表本体共建共享系统 OTCSS 的一整套服务功能: 格式转换、共建共享、一致性检测、术语服务、Web Service 服务、Linked Data 服务、可视化、自动构建等。欢迎登录深圳大学图书馆 NKOS 研究室网站(<http://nkos.lib.szu.edu.cn>) 了解其各项功能,示范系统: CCT1 _

OTCSS。

基于云计算模型的中文知识组织系统注册与服务平台正在开发中,目标是为中文知识组织系统的拥有者提供自动上传和管理服务,并为所有上传的 NKOS 自动配置成套服务功能,从而推动中文 NKOS 的开放服务进程。

中文知识组织系统形式化语义描述标准体系包含纯 SKOS 版^[16]、CNKOS 扩展版^[16-26]、高受控 OntoThesaurus 版及领域本体 OWL 应用规范 4 个部分。深圳大学图书馆 NKOS 研究室目前提供 CCT 第一版全部数据^[13-14]和 CLC 第四版 B 大类数据^[27]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、CNKOS 和 OntoThesaurus 格式语义描述代码(可选择 RDF/XML、NTriple 和 JSON 格式),便于研究者对这三种不同的描述规范进行更直观的比较分析和应用。欢迎使用并多提宝贵意见。

参考文献:

- [1] W3C. OWL Web Ontology Language Overview[EB/OL]. [2004 - 04 - 14]. <http://www.w3.org/TR/2004/REC-owl-features-20040210/>.
- [2] W3C. OWL Web Ontology Language Reference[EB/OL]. [2004 - 02 - 12]. <http://www.w3.org/TR/2004/REC-owl-ref-20040210/>.
- [3] W3C. OWL Web Ontology Language Guide[EB/OL]. [2004 - 02 - 12]. <http://www.w3.org/TR/2004/REC-owl-guide-20040210/>.
- [4] W3C. OWL Web Ontology Language Semantics and Abstract Syntax[EB/OL]. [2004 - 02 - 12]. <http://www.w3.org/TR/2004/REC-owl-semantics-20040210/>.
- [5] W3C. OWL Web Ontology Language Use Cases and Requirements[EB/OL]. [2004 - 02 - 12]. <http://www.w3.org/TR/2004/REC-webont-req-20040210/>.
- [6] 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.
- [7] 曾新红. 中文叙词表本体——叙词表与本体的融合[J]. 现代图书情报技术 2009(1): 34 - 43.
- [8] 曾新红, 明仲, 蒋颖, 等. 中文叙词表本体共建共享系统研究[J]. 情报学报 2008(3): 386 - 394.
- [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.
- [10] 曾新红, 林伟明, 明仲. 中文叙词表本体一致性检测机制研究与实现[J]. 现代图书情报技术 2008(5): 1 - 9.
- [11] 曾新红, 林伟明, 明仲. 中文叙词表本体的检索实现及其术语学服务研究[J]. 现代图书情报技术 2008(2): 8 - 13.
- [12] 林伟明, 曾新红. OntoThesaurus Web Service API 及其应用研究

- [J]. 图书情报工作 2010(2):119-122,139.
- [13] 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.
- [14] 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.
- [15] 曾新红,蔡庆河,曾汉龙,等. 中文叙词表本体可视化群组布局算法研究与实现[J]. 现代图书情报技术 2012(10):8-14.
- [16] 王军,卜书庆. 网络环境下的知识组织规范和应用指南[M]. 北京:国家图书馆出版社 2012.
- [17] W3C. SKOS Simple Knowledge Organization System Reference[EB/OL]. [2010-02-23]. <http://www.w3.org/TR/2009/REC-skos-reference-20090818/>.
- [18] W3C. SKOS Simple Knowledge Organization System Primer[EB/OL]. [2010-02-23]. <http://www.w3.org/TR/2009/NOTE-skos-primer-20090818/>.
- [19] 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/>.
- [20] Allemang D, Hendler J. 实用语义网:RDFS 与 OWL 高效建模(英文版)[M]. 北京:人民邮电出版社 2009.
- [21] DCMI. DCMI Metadata Terms[EB/OL]. [2010-03-11]. <http://purl.org/dc/terms/>.
- [22] Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies[EB/OL]. [2005-12-28]. <http://www.niso.org/standards/resources/Z39-19.html>.
- [23] W3C. OWL2 Web Ontology Language Document Overview[EB/OL]. [2011-04-13]. <http://www.w3.org/TR/2009/REC-owl2-overview-20091027/>.
- [24] W3C. OWL2 Web Ontology Language Quick Reference Guide[EB/OL]. [2011-04-13]. <http://www.w3.org/TR/2009/REC-owl2-quick-reference-20091027/>.
- [25] W3C. OWL2 Web Ontology Language New Features and Rationale[EB/OL]. [2011-04-13]. <http://www.w3.org/TR/2009/REC-owl2-new-features-20091027/>.
- [26] 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.
- [27] 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.

OntoThesaurus Specializations for High-controlled Chinese Vocabularies

Zeng Xinhong

The NKOS Research Office, Shenzhen University Library, Shenzhen 518060

[Abstract] Aiming to the description requirements and the evolution trend to fine-grained ontology of high-controlled Chinese vocabulary, the specifications for high-controlled Chinese vocabularies are studied based on the existing research of OntoThesaurus including the unique namespace URI and its abbreviation, and a set of strict formal definitions. The RDF Schema of OntoThesaurus in RDF/XML syntax based on OWL DL and OWL 2 is created according to the application environment of Semantic Web, to provide Classes and Properties for machine usage. The paper describes the special definitions of the OntoThesaurus which are different from the SKOS and the CNKOS, and displays the example, as well as the correspondence among the three vocabularies. Finally, the functions of supporting system for OntoThesaurus are introduced briefly.

[Keywords] OntoThesaurus high-controlled vocabulary OWL DL OWL 2 formal representation semantic description KOS

科研教育开放信息创新应用大赛即将拉开帷幕

2013年9月,由中国科学院大学(简称“国科大”)和中国科学院国家科学图书馆(简称“国科图”)联合举办的“科研教育开放信息创新应用大赛”启动仪式将在国科图举行。

大赛平台将提供大量的开放数据集、开放工具、相应的咨询服务以及场地等基础性支持条件,参赛者可以以各种方法灵活运用大赛提供的以及参赛者自己拥有或发现的开放数据资源,结合科研、教育、管理以及社会生活等各方面的需求,通过数据嵌入、混搭、集成、整合、分析、计算、可视化等方法,创造出新颖实用的应用、方法及工具。

据悉,这是国内首次由高校或图书馆组织的开放信息创新应用大赛,也是国际首个由图书馆大规模组织的类似大赛,是国科大和国科图适应数字科研时代支持科技创新、教育创新和文献情报服务开放创新的积极尝试。

详细信息请登录大赛官方网站 <http://www.openinnovation.cn>。