

●贾君枝

简单知识组织系统与汉语主题词表^{*}

摘 要 简单知识组织系统(SKOS)是W3C组织语义网下的工作组织发布并维护,为知识组织系统的表达提供强有力且简单的机器理解方式。它的建立为叙词表的形式化表达提供了概念、模型,使各叙词表之间、叙词表与本体之间的映射成为可能。如何利用SKOS,使叙词表在词义检索中得到更好地应用,实现用户检索请求的语义本体化,将是叙词表研究者更大的挑战。图1。参考文献5。

关键词 简单知识组织系统 SKOS 语义网 汉语主题词表

分类号 G354.4

ABSTRACT In this paper, the author explains SKO (simple knowledge organization), which provides a powerful and simple way for the machine understanding of representation of knowledge organization. It provides a conceptual model for the formalized representation of thesauri, and makes possible the mapping between thesauri and notologies. It will be a great challenge to thesaurus researchers to fully utilize SKOS, to better apply thesauri in word meaning searches and realize semantic web and notology application in user search queries. 1 fig. 5 refs.

KEY WORDS Simple knowledge organization system. CKOS. Semantic web. Classified Chinese Thesaurus.

CLASS NUMBER G354.4

本体的构建是一项复杂工程,尤其是领域知识的获取更是一件难事,需要领域专家的参与。当前的叙词表已集专家知识成果,如能被本体直接引用,对于本体构建将是一件非常有意义的事情。汉语主题词表作为一种规范化词表,在信息结构化检索中起到准确匹配的作用。但同现有的本体相比,叙词表更多地面向人而不是面向计算机构建,没有采用计算机形式化语言表述,因而限制其发展。许多叙词表或者采用形式化语言进行描述,或者被一些本体加以利用以构建新的本体,以寻求语义网环境下的新发展^[1]。

在叙词表形式化表示的过程中,存在着这样的问题:如果采用RDF或者OWL进行描述,不同人认识不同,可能形成不同的模式(schemes),这给词表之间的共享及互操作带来一定困难。传统叙词表都是基于图书情报领域的国际标准制定的,采用ISO2788、ISO5964、ISO5963三个标准定义术语之间的关系,保证了各叙词表间彼此的结构及其语义关系含义一致。既然叙词表构建过程都遵循一定的标准规范,如果定义一个中间语言,即定义一个标准的模式语言,大家遵循这样的标准对叙词表进行形式化描述,可保证形

式化的叙词表具有共享及易于映射的特点。简单知识组织系统(Simple Knowledge Organization System, SKOS)因此产生,它是W3C组织语义网下的工作组织发布并维护的,强调为知识组织系统的表达提供强有力且简单的机器理解方式。目前已有一些叙词表尝试采用SKOS描述,如英国档案叙词表等。本文以汉语主题词表为核心,对简单知识组织系统的内容及结构作了较明确描述,并结合现有语义网环境特点,对SKOS特点做进一步剖析。

1 SKOS 内容及结构

SKOS Core提供表达概念模型(concept schemes)的基本结构及内容^[2]。概念模型表示概念及其概念之间的语义关系。SKOS Core词汇是由一系列RDF属性及其RDF类构成,并用RDF图式来表达概念模型的内容及结构。如图1所示,左边是选自汉语主题词表中“海洋法”术语及其他术语之间关系的显示^[3],右边则采用RDF图形式对它们进行形式化描述。

^{*} 本文为国家社会科学基金项目“汉语框架网络知识本体构建研究”(06CTQ004)的成果。

海洋法

D 海上法
S 国际公法
Z 法(法律)
C 哥伦伯斯, C. J.
联合国海洋会议
领海

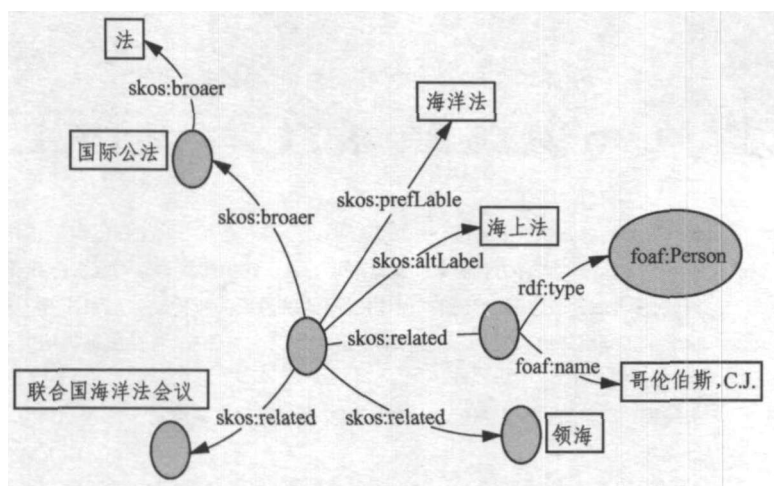


图 1 SKOS RDF 图示汉语主题词表
</rdf:RDF>

1.1 概念类定义

叙词表中每一术语词汇都可定义为概念类。如下描述,词汇“盗窃”被定义为概念类。

```
< rdf:RDF.
  xmlns:rdf = " http://www. w3. org/1999/
    02/22-rdf-syntax-ns#"
  xmlns:skos = " http://www. w3. org/
    2004/02/skos/core#" >
  < skos: Concept rdf: about = " http://
    www. example. com/concepts#盗窃" / >
```

1.2 属性定义

属性表示概念之间的关系,用来表示叙词表中“用、代、属、分、参”语义关系。SKOS Core 的属性分为以下类型:标签化属性、文档属性、语义关系属性、主题索引属性、集合属性。

标签化属性描述叙词表中术语的“用代”关系,采用 skos:prefLabel、skos:altLabel 定义。如下所示,左面代表叙词表中的用代关系,右面采用相应的属性描述这种关系。

徒刑 D 无期徒刑 有期徒刑	<pre>< rdf:RDF xmlns:rdf = " http://www. w3. org/1999/02/22-rdf-syntax-ns#" xmlns:skos = " http://www. w3. Org/2004/02/skos/core#" > < skos: Concept rdf: about = " http://www. example. com/concepts#徒刑" > < skos: prefLabel > 徒刑 </ skos: prefLabel > < skos: altLabel > 无期徒刑 creatures </ skos: altLabel > < skos: altLabel > 有期徒刑 </ skos: altLabel > </ skos: Concept > </ rdf: RDF ></pre>
----------------------	--

文档属性用以描述叙词表中术语的注释说明。包含 7 个属性,有概念的定义、使用范围、实例、修订说明、管理信息、数据修改信息等,分别用 skos:definition、skos:scopeNote、Skos: example、skos: historyNote、Skos: editorialNote、skos: changeNote 定义。如下所示,叙词表对“死缓”作进一步限定说明。

死缓 注 判处死刑、缓期二年执行,强迫劳动,以观后效	<pre> <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:skos="http://www.w3.org/2004/02/skos/core#" > <skos:Concept rdf:about="http://www.example.org/concepts#死缓"> <skos:definition>判处死刑、缓期二年执行, 强迫劳动,以观后效. </skos:definition> </skos:Concept> </rdf:RDF> </pre>
-------------------------------	--

语义关系属性主要表示叙词表中的属分关系和传递性,相关关系属性具有对称性特征。下面以叙词表中“诉讼程序”为例,采用上述语义关系属性对其进行描述。

诉讼程序 F 起诉 上诉 审理 申诉 执行 C 诉讼	<pre> <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:skos="http://www.w3.org/2004/02/skos/core#" > <skos:Concept rdf:about="http://www.example.com/concepts#诉讼程序"> <skos:narrowerrdf:resource="http://www.example.com/concepts#起诉"/> <skos:narrowerrdf:resource="http://www.example.com/concepts#上诉"/> <skos:narrowerrdf:resource="http://www.example.com/concepts#审理"/> <skos:narrowerrdf:resource="http://www.example.com/concepts#申诉"/> <skos:narrowerrdf:resource="http://www.example.com/concepts#执行"/> <skos:relatedrdf:resource="http://www.example.com/concepts#诉讼"/> </skos:Concept> </rdf:RDF> </pre>
--	---

主题索引属性声明该概念是网络信息资源的主题,即用叙词表中词汇表示网络信息资源的主题。包含属性 skos:subject、skos:isSubjectOf、skos:primarySubject、skos:isPrimarySubjectOf。这些属性用于标引网络信息资源的主题,如果信息资源涉及多个主题,则采用 skos:primarySubject 声明其关键主题。其中 skos:subject 与 skos:isSubjectOf 互为逆属性,skos:primarySubject 与 skos:isPrimarySubjectOf 互为逆属性。下面采用叙词表中的“法律”术语表示北大法律信息网网页的主题。

<pre> <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:skos="http://www.w3.org/2004/02/skos/core#" xmlns:foaf="http://xmlns.com/foaf/0.1/" > <foaf:Document rdf:about="http://www.chinalawinfo.com/index.htm"> <skos:Concept rdf:about="http://www.example.org/concepts#法律"> <skos:isSubjectOf rdf:resource="http://www.chinalawinfo.com/index.htm"/> </skos:Concept> </rdf:RDF> </pre>

集合属性用以定义概念集合,实际上主要表达叙词表中的属分关系。采用 skos:Collection 及 skos:member 定义。另外可以用 skos:OrderedCollection、skos:memberList 描述有顺序的集合。一般可包含在语义关系属性中定义,并可进行嵌套定义。下面是叙词表词族索引“法”的相关子类。

法(法律)	< rdf:RDF
.. 国际法	xmlns:rdf = "http://www. w3. org/1999/02/22-rdf-syntax-ns#"
... 国际公法	xmlns:skos = "http://www. w3. org/2004/02/skos/core#"
... 捕获法	xmlns:rdfs = "http://www. w3. org/2000/01/rdf-schema#" >
... 国际习惯法	< skos:Concept rdf:about = "http://www. example. com/concepts#法" >
... 海洋法	< rdfs:label > 法律 </rdfs:label >
.. 国际私法	< skos:narrower >
.. 国际刑法	< skos:Collection >
	< skos:member rdf:resource = "http://www. example. com/concepts#国际法" / >
	< skos:member >
	< skos:Collection >
	< skos:member rdf:resource = "http://www. example. com/concepts#国际公法" / >
	< skos:member >
	< skos:Collection >
	< skos:member rdf:resource = "http://www. example. com/concepts#捕获法" / >
	< skos:member rdf:resource = "http://www. example. com/concepts#国际习惯法" / >
	< skos:member rdf:resource = "http://www. example. com/concepts#海洋法" / >
	</skos:Collection >
	</skos:member >
	< skos:member rdf:resource = "http://www. example. com/concepts#国际私法" / >
	< skos:member rdf:resource = "http://www. example. com/concepts#国际刑法" / >
	</skos:Collection >
	</skos:member >
	</skos:Collection >
	</skos:narrower >
	</skos:Concept >
	</rdf:RDF >

2 SKOS 的特点

2.1 SKOS 的开放性

由于 SKOS 词汇的有限性,需要借助于其他的 RDF 词汇对它进行扩展。当前主要结合 FOAF、DC 元数据的使用来扩展其功能,为其成为受控词汇的潜在的标准描述框架提供可能。FOAF (Friend of a Friend) 描述人、机构、公司等组织所特有的姓名、地址、邮编、主页等信息,并通过参照指针将多个文档建立联接,以 RDF/OWL 语言建立计算机可读的信息。SKOS 能够将叙词表中的术语通过 FOAF 词汇建立与其他文档的联接。同时 SKOS 通过相互子类及子属性定义方式建立与 DC 元数据之间的映射。比如 SKOS:concept 作为 dc:subject 属性的描述对象,skos:subject 可定义为 dc:subject 的子属性。这样既可以完整地描述网络资源创建者、出版者等机构或人的详细信息,弥补 DC 元数据 dc:creator 属性描述不足的缺陷,又可采用叙词表的规范术语对资源的主题进行准确描述,并建立与其他文档的联系。如叙词表中的术语 <http://www. example. com/concepts#henry8> 是由

名字叫 Alistair Miles 的人提出,并修订于 2005 年 2 月 6 日,描述于下:

```
< rdf:RDF
  xmlns:rdf = "http://www. w3. org/1999/02/22-rdf-syntax-ns#"
  xmlns:skos = "http://www. w3. org/2004/02/skos/core#"
  xmlns:dcterms = "http://purl. org/dc/terms/"
  xmlns:foaf = "http://xmlns. com/foaf/0.1/" >
  < skos:Concept rdf:about = "http://www. example. com/concepts#henry8" >
    < skos:prefLabel > King Henry VIII </skos:prefLabel >
    < dct:modified > 2005-02-06 </dct:modified >
    < dc:creator >
      < foaf:Person >
        < foaf:name > Alistair Miles </foaf:name >
      </foaf:Person >
    </dc:creator >
  </skos:Concept >
</rdf:RDF >
```

(下转第 84 页)

检错效果影响很大。虽然 Racer 是推理能力较为强大的商业化推理工具,但是其局限性也显而易见。由于 Racer 是基于描述逻辑 Tableau 算法实现,所以对 TBox 概念层次的推理有较好支持,而对 ABox 实例推理支持不太好。而且对于 OWL 的一些属性,如 sameAs、基数约束集等也支持不太好。这就意味着使用 Racer 进行本体库检错推理时并不能完全检测出本体库中的知识描述错误;而且 Racer 是个商业化推理工具,不提供开源服务,我们无法进行二次开发和算法优化,不能为系统提供更好的个性化支持。

参考文献

- 1 R. Moller, V. Haarslev, The Description Logic Systems. Cambridge University Press, 2003: 289 ~ 310
- 2 Franz Baader. The Description Logic Handbook: Theory, Implementation and Applications, Cambridge University Press, 2003. 47 ~ 56
- 3 Schmidt-Schaup M, Smolka G, Attributive concept descriptions with complements. Artificial Intelligence, 1991. 48(1)
- 4 文斌. 基于描述逻辑的语义网知识推理研究. [学位论文]

文], 云南师范大学, 2005: 123 ~ 133

- 5 Baader F, Sattler U, An Overview of Tableau Algorithms for Description Logics. Studia Logica, 2001(69)
- 6 徐国虎. 基于本体的国共合作领域知识推理研究. [博士学位论文]. 武汉大学, 2006
- 7 I Horrocks, S Tobies. Reasoning with Axioms: Theory and Practice. Proc. of the 7th Int. Conf. on the Principles of Knowledge Representation and Reasoning (KR' 2000), 2000: 286 ~ 294
- 8 Volker Haarslev, Ralf Möller. Description of the Racer System and its Applications, Workshop on Description Logics (DL - 2001), Stanford, USA, August 2001

徐国虎 中南财经政法大学工商管理学院教师。
通讯地址: 武汉中南财经政法大学(南湖校区)工商
管理学院物流教研室。邮编 430064。

许芳 中南民族大学管理学院教师。通讯地
址: 武汉中南民族大学管理学院。邮编 430074。

董 慧 武汉大学信息管理学院教授、博士生导师。
通讯地址: 武汉大学。邮编 430072。

(来稿时间: 2007-03-06)

(上接第78页)

2.2 SKOS 的局限性

SKOS 采用 RDF 图形式描述, 与本体语言相比有一定的局限, 它并不规定语义, 只为每一个资源描述体系提供一个能够描述其特定需求的语义结构的能力, 缺乏对属性值加以限制, 不能像本体那样采用描述逻辑进行推理, 无法进行一致性检验或者对新类进行归类。因此采用 SKOS 描述的叙词表只是叙词表形式化表示的过渡阶段, 揭示的语义关系有待进一步明确, 离真正转换为本体还需一定的路程, 需要进一步的人工参与, 尤其在语义关系上作进一步调整, 需建立与其他叙词表、本体的映射, 尽可能利用已构建成功的本体的语义关系, 利用各词表及本体的优点, 通过相互补充相互作用发挥本体在语义网中的作用。

3 结束语

W3C 组织明确提出传统叙词表的存在为发展语义网中的本体提供了起点。如果不借助当前叙词表的成果, 语义网的发展将会变得缓慢^[4]。语义网的关键元素是使用本体来定义概念及关系, 通过本体提供数据的上下文环境, 信息检索及搜索引擎能够利用这些上下文信息来执行基于概念的语义搜索, 同时本体所获得的丰富语义可将简单事实合并, 推理出新事实, 并推导出本体所隐含的知识。如何在语义层面共享或重用叙词表, 并使之与本体融合适应当前语义检索的要求, 成为目前叙词表在语义网环境下发展所面临的主要问题。W3C 组织适时发布并维护的简单知

识组织系统为叙词表向形式化表述迈进了一大步, 为叙词表的形式化表达提供了概念模型, 使各叙词表之间、叙词表与本体之间的映射成为可能^[5]。但这并不是叙词表发展的最终目标, 如何利用 SKOS, 采用 Web 上表达机器可理解内容更强的语言描述叙词表, 叙词表如何在语义检索中得到更好应用, 从语义层面组织信息、检索信息, 实现用户检索请求的语义本体化, 将是叙词表研究者面临的更大挑战。

参考文献

- 1 贾君枝. FrameNet 叙词表与传统叙词表语义关系比较研究. 情报理论与实践, 2006(5)
- 2 W3C Working Draft 2. SKOS Core Guide. [2007-01-12] <http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20051102/> SKOS.
- 3 中国科学技术情报研究所, 北京图书馆. 汉语主题词表. 北京: 科学技术文献出版社, 1980
- 4 Salvador Sanchez-alonso, Elena Garcia-barriocanal. Making Use of Upper Ontologies to Foster Interoperability Between Skos Concept Schemes. Online Information Review, 2006, 30(3)
- 5 B. J. Wielinga A. Th. Schreiber J. Wielemaier J. A. C. Sandberg From Thesaurus to Ontology. [2006-07-14]. <http://www.cs.vu.nl/guus/papers/Wielinga01a.pdf>.

贾君枝 山西大学管理学院副教授, 博士。通讯地址:
太原山西大学管理学院。邮编 030006。

(来稿时间: 2007-03-26)