

传统知识组织系统的关联数据化发展^{*}

贾君枝^{1,2} 李衍¹

(1. 山西大学经济与管理学院, 太原 030006; 2. 中国人民大学信息资源管理学院, 北京 100872)

摘要: 以受控词表为核心的传统知识组织系统在资源描述与表示中发挥着重要作用。本文在对大量研究文献调查研究的基础上, 概述国内外知识组织系统的关联数据化发展研究情况, 并通过对相关研究内容的总结提炼, 明确传统知识组织系统语义化、关联化的实施需经过SKOS模型化、RDF序列化和关联数据发布3个主要步骤, 并对各部分进行详细描述, 以期为其他传统知识组织系统实现关联数据化提供参考。此外, 本文总结了传统知识系统关联化应用主要体现在提高受控词表的访问量、有效地实现对数据集的描述表示, 以及促进数据集成实现。

关键词: 知识组织系统; 关联数据; SKOS

中图分类号: G254; TP18

DOI: 10.3772/j.issn.1673-2286.2020.03.005

数量庞大的数据资源由于缺乏形式化描述、缺少明确的模型和框架描述, 而无法被有效利用和关联, 致使这些人类生产的海量信息成为“信息孤岛”。关联数据的出现, 使得世界上数以万计海量资源形式化、语义化、关联化及共享化成为可能。关联数据作为一种用来发布和联接各类数据、信息和知识的规范, 提供了机器可理解的表述方式, 达到了消除歧义的效果, 可以实现资源的真正共享, 以其独有的优势获得了相关研究者的大量关注。自关联数据提出以来, 国内外学者积极实践, 不仅将资源发布为关联数据, 也先后将以受控词表为核心的传统知识组织系统发布为关联数据, 旨在实现对资源的描述, 并以此为核心实现多个关联数据集之间的相互映射, 达到知识资源融合、序化的效果。近年来关联的开放数据呈几何级飞速增长, 截至2019年12月, 关联开放数据 (Linked Open Data, LOD) 云中简单知识组织系统 (Simple Knowledge Organization System, SKOS) 描述的数据集达152个^[1], 在关联开放词表 (Linked Open Vocabularies, LOV) 中与SKOS关联的词汇表达264个^[2]。

本文在对国内外各类型受控词表研究成果及实践调查的基础上, 梳理从受控词表到关联数据的语义迁

移路径, 分析受控词表的关联数据实现步骤, 对其在网络环境下应用进行系统分析。

1 传统知识组织系统关联数据化研究进展

传统知识组织系统旨在清晰地揭示知识的含义, 充分表达知识系统之间的语义关系, 为信息组织和检索提供科学有效的途径。各种知识组织系统从卡片式、书本式向机读化转变, 但由于格式所限, 各个词表之间无法互相兼容, 不能被万维网所通用。关联数据技术可实现对知识资源的立体化组织, 方便用户检索利用、机器理解与共享, 因此将传统知识组织系统发布为关联数据并建立互联是传统知识组织系统在语义网环境下的必然发展趋势和要求。语义网环境下知识组织系统关联化改造的关键是使用计算机可以理解的语言, 形式化表示其结构特征和内容特征, 构建概念模型和语义关系, 因此选择合适的资源描述语言非常重要。针对受控词表语义化描述, W3C提出的SKOS描述作为表示知识组织系统的标准规范, 可将叙词表、分类表、标题表等各种传统的知识组织系统以RDF方式描述, 使之

^{*}本研究得到国家社会科学基金项目“数据开放环境中的词表重用问题研究” (编号: 19BTQ023) 资助。

具有“机器可读”能力,为语义网环境下的知识组织提供描述和转化机制,解决数据资源间的语义互操作问题。SKOS发布以来,研究人员从不同角度开展了叙词表、分类法、主题词表等的SKOS转化研究,促进了叙词表的进一步利用和发展。综合来看,相关的研究主要集中在对词表转换规则和转换方法的理论研究、实践及关联发布3个方面。

1.1 转换规则和转换方法研究

这类研究视角较多存在于早期的相关研究中。国外学者Harper^[3]探讨运用XSLT (Xtensible Stylesheet Language Transformations)对《美国国会图书馆标题表》(LCSH)的特点进行深入分析,并完成SKOS描述;欧洲语义网高级开发项目推出了GEMET的SKOS格式^[4],Assem等^[5]《公共部门主题词表》(IPSV)、《荷兰视听档案通用词汇表》(GTAA)及《医学主题词表》(MeSH)叙词表为例,讨论了从叙词表到SKOS的转换方法。国内学者范炜^[6]对所选词表中的Librarianship类进行了SKOS化描述,对叙词语义关系可视化表示以及网页进行主题标引;贾君枝^[7]对《汉语主题词表》的部分叙词及其关系用SKOS Core进行了转换描述;曾新红^[8]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,针对现有中文词汇表中SKOS能够描述的部分,研究制定了用SKOS描述的具体规范;段荣婷^[9]以《中国档案主题词表》的语义置标为例,详细阐述了SKOS系统化与规范化控制的具体实现;张士男等^[10]论述了《科图法》的SKOS化描述的技术方法;喻菲^[11]提出了《中国分类主题词表》“主题词一分类法对应表”部分的“主题词规范数据格式”与SKOS词汇的大致对应关系;白华^[12]以《中国图书馆分类法》为例,以SKOS语言为工具,探讨了如何建构分类本体,并研究如何描述分类本体中的复杂关系,为中文分类法建立一个轻量级本体奠定了方法基础;施国良^[13]针对分面分类的特性、功能和优势,对SKOS的原有核心进行扩展,尝试做出一个实例模型使其能够表征分面分类表的结构和标引方法。

1.2 特定词表的转换实践研究

学者针对特定词表的转换实践也开展了相关研究。Assem等^[5]提出了一种基于SKOS的叙词表转换

RDF/OWL模式的方法,且将这种转换方法应用于IPSV、GTAA和MeSH,并通过案例研究评估了转换方法和对SKOS叙词表的适用性,其带动了大量叙词表转换成果的出现,如《德国经济学叙词表》(STW)、LCSH、《艺术与建筑叙词表》(AAT)、《多语种农业叙词表》(AGROVOC)、《通用多语种环境叙词表》(GEMET)、《美国农业图书馆叙词表》(NALAT)、《欧洲联盟多语种叙词表》(EUROVOC)等。其中,Harper^[3]和Summers等^[14]以LCSH为研究对象,分别进行了MARC格式的LCSH到SKOS的转换实验;Morshed等^[15]将AGROVOC进行了SKOS转化实验;Smedt^[16]针对EUROVOC进行了SKOS转化实验。刘丽斌等^[17]利用Java语言写入SKOS描述,将《中国分类主题词表》SKOS化;刘春艳等^[18]应用SKOS Core词汇表完成《英国档案叙词表》(UKAT)节选段落到本体的转换;刘华梅^[19]对《中国分类主题词表》主题词进行了SKOS化描述及自动转换研究,提出了主题词从MARC格式到SKOS格式的转换方案;杨雪莉^[20]以《中国图书馆分类法》为例,采用SKOS方法构建分类法的本体模型;贾君枝等^[21]讨论《汉语主题词表》的SKOS转换过程,明确了从数据库向SKOS语言表示的自动转换思路,并且编写Java代码实现了该转换功能。

1.3 关联数据发布研究

关联数据发布方面,许多图书馆机构将自己的书目数据转换成关联数据发布,如瑞典皇家图书馆将书目数据发布为关联数据并提供了与DBPedia的连结。书目数据的发布带动了知识组织系统的关联数据实现。《杜威十进分类法》(DDC)进行SKOS语义化描述,并在2009年将其前3级类目发布成为关联数据形式。美国国会图书馆将《美国国会图书馆分类法》(LCC)在2012年以SKOS形式发布为关联数据。此外,各种叙词表、主题词表也陆续被国外研究人员发布为关联数据,图书馆关联数据孵化小组(Library Linked Data Incubator Group)的报告显示,主题词表方面有LCSH、《法国国立图书馆主题词表》(RAMEAU)、《德国国立图书馆受控词表》(SWD)、《日本国立国会图书馆主题词表》(NDLSH);叙词表方面有STW、AGROVOC、EUROVOC、《图形材料叙词表》(TGM)、《地球环境叙词表》(EARTH)等。其中Neubert^[22]和Albertoni等^[23]分别利用SKOS来描述STW和EARTH,

并对SKOS数据集进行了关联发布,实现了与其他外部资源(如DBpedia)的映射连接。Morshed等^[24]对不同领域叙词表间的自动匹配方法进行比较评估,通过skos:exactMatch和skos:closeMatch术语精确匹配,实现AGROVOC与The Soz的映射。北卡罗来纳州立大学和NESCent共同开发的跨学科词汇辅助工程(HIVE),可以基于SKOS格式的受控词表集合动态地集成和吸取跨学科的主题词汇,创建自动化的元数据描述方式,实现术语系统之间的交互^[25]。欧石燕^[26]以《汉语主题词表》为例,将其进行了SKOS转化实验,利用Jena、Fuseki、Pubby的组合,将SKOS/RDF的词表数据发布为关联数据;周峰^[27]选取《地质学汉语叙词表》,设计了叙词表的SKOS映射关系,将其转换为RDF数据,最后利用Jena和Pubby实现了关联数据的发布;贾君枝等^[28]以DDC为例,探讨了DDC关联数据从SKOS化到关联发布的具体实现;鲜国建等^[29]基于SKOS及SKOS-XL规范和模型,将《农业科学叙词表》(CAT)中的叙词及词间语义关系进行规范表达和关联描述,成功转化为CAT/SKOS,并与AGROVOC、LCSH等知识组织体系在概念实例层面建立语义关联,用Virtuoso构建了CAT关联数据发布系统;黄华军等^[30]以《中国图书馆分类法》(第四版)为例,采用CNKOS对分类法进行语义化描述,使用Lucene全文检索引擎、ExtJs插件、Axis SOAP引擎等关键技术实现了分类法共享服务(CLSS)原型系统,发布了《中国分类主题词表》的关联数据服务。

LOD数据云中出版类数据集(目前包含150个数据集)中包含许多已经发布为关联数据的知识组织系统,可将其归为知识组织工具数据集类,其中包括叙词表、分类表、元数据、规范文档等知识组织工具,知识组织工具实现关联发布的有30个(占比20%),使用SKOS描述的分类表和叙词表有12个(占比8%),可看出叙词表的数量远超过了分类表的数量,原因在于与叙词表通过使用参照系统和索引系统揭示知识关联不同,分类法主要通过隶属、并列和设置交替类目、多重列类、类目参照等方式揭示知识关联,类目体系与叙词表相比更加明确清晰和严格^[31],因此在SKOS模型化过程中更容易受到“语义折扣化”影响,转换过程通常会受到其本身结构的影响而无法使分类法语义完整表达,阻碍了分类表关联化的进程。与此同时,众多已经SKOS模型化和关联发布的传统知识组织系统并未在LOD平台进行发布,且没有中文传统知识组织系统发

布,说明目前传统知识组织系统关联化水平依旧有待进一步提高。

2 传统知识组织系统关联化的具体实现

通过对国内相关文献的分析,知识组织系统关联化步骤基本分为三步(见图1)。第一步为SKOS模型化,利用SKOS提供的一整套概念体系,包括概念、语义关系描述、注释、映射关系等标签,表示传统知识组织系统中的语义元素和概念体系。由于关联数据要以RDF格式编码和关联,SKOS核心词表是RDF的应用,可以比较完备地将传统KOS的内容结构RDF序列化,为实现关联化提供必要的基础。第二步为RDF关联化,包含URI命名、关联数据链接和RDF表示,即先给传统知识组织系统中的每个概念进行唯一URI标识,用RDF语言来描述实体间的关系。RDF语言通过引用命名空间中的属性来对数据进行语义描述,以实现数据的互联,以RDF三元组的形式清晰地表示知识组织系统的等级关系。第三步为关联数据发布,包含外部关联、建立SPARQL查询、注册有关平台等。

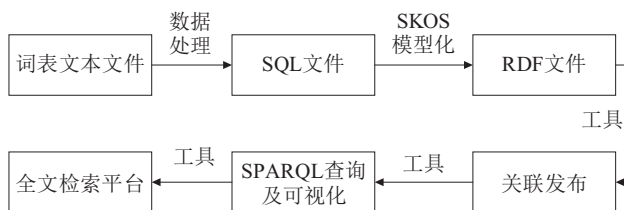


图1 词表发布为关联数据的步骤

2.1 SKOS模型化

SKOS为传统知识组织系统(如叙词表、分类法、主题词表、术语表等)提供了一套语义环境下简单灵活、机器可读的描述和转换机制,使叙词表之间、叙词表与本体之间的共享成为可能。W3C的SKOS数据模型提供了用于表达概念基本结构和内容的规范化标准。辞典、主题词表和分类法等通过SKOS规范中定义的类和属性表示为RDF模型图,从而与语义网技术兼容。尽管语义更丰富的RDFS/OWL本体可以执行更广泛的逻辑推理功能,但在一般情况下,SKOS格式的叙词表足以满足日常需求^[32]。然而,知识组织系统中包含大量复杂的语义信息,如组配概念、主题索引、非正式主题词之间的映射、术语元数据(创建时间、最后修改

日期、版本等)等^[33],难以用SKOS直接描述,除使用SKOS-XL外,针对无法满足需求的解决办法主要有:自定义SKOS标签;将SKOS与RDF、OWL等其他建模原语结合。

传统知识组织系统的SKOS化基本分为4个步骤。

①确定概念体系,利用SKOS核心词汇表完整地表达出知识组织系统的概念体系。每个知识组织系统都可以看作一个概念框架,即知识组织系统内所有概念及概念之间的语义关系的集合,一个概念框架下可以有多个顶级概念,采用skos:ConceptScheme表示。因此,通过skos:ConceptScheme和skos:has-TopConcept对整个知识组织体系进行概括说明,对每个概念通过skos:inScheme声明其所属关系。②定义概念,提取知识组织系统的概念、定义、注释、标签等关键项,利用SKOS中的类和属性标签分别对其进行语义描述。每个概念均可用skos:Concept来进行标注。skos:preflabel用于描述正式的概念名称,skos:altLabel是用于说明概念类的交替词汇标签,当使用多个词汇标签表示同一概念时,可采用此词汇标签。分别用skos:snote和skos:notation对应于叙词的注释和所属范畴号。③描述概念间关系,SKOS定义了相应的属性标签来描述概念之间的语义关系,skos:broader和skos:narrower仅用来表示概念与其直接上下位概念的链接关系,不具有传递性。但为了兼容各方需求,SKOS又定义了两个具有传递性的属性(skos:broaderTransitive和skos:narrowerTransitive)来表示任意级别的上下位链接关系,这两个属性不能用作断言声明,仅用于推理或查询扩展等应用。skos:related用来声明具有关联但又不是层级关系的类目的描述相关关系等^[34],其具有对称性,但不遗传给其子关系,且SKOS较为松散,允许skos:related等属性具有自反性(概念与其自身之间的关系)。

2.2 RDF关联化

2.2.1 URI命名

关联数据是通过URI识别实体的,明确实体之后,要对实体进行URI命名。通过URI定位,能够识别知识组织系统中的所有实体。传统的知识组织系统是基于语词的概念表达和组织的符号系统,每一个款项单元都是一个语词,而SKOS的建模方法是通过SKOS

规范中定义的类和属性表示为RDF模型图,从而与语义网技术兼容。将传统的知识组织系统转换成机器可理解的形式时,要在模型上进行修改,每个款项不再是一个语词,而是一个概念,有其唯一的URI。网络中的任何事物通过数据源的URI名称都是可辨别的,为知识组织系统的每个概念分配唯一的URI标识,便于之后进行RDF数据互联,是知识组织系统关联化至关重要的基础。SKOS中的URI识别机制解决了概念歧义的问题。通过机器分配给每个实体特定的URI,除需自动定义词表中扩展词汇的命名域外,所采用的SKOS/RDF/RDFS/OWL词汇使用其原有的命名域及缩写(见表2)。

表2 词表命名域URI缩写

URI	缩写
对象词表URI命名域(自定义)	自定义
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

针对词表的URI命名,首先以自定义命名域为基本地址,该地址即为该关联数据的发布平台。URI要求满足唯一性,鉴于词表中类号唯一的特点,采用分类号或叙词名作为URI内容。

2.2.2 RDF表示

经过URI命名和SKOS模型化之后,便可以运用RDF语言来描述实体间的关系。类及类间关系的RDF描述是整个RDF描述的主体,RDF多种词汇表可以用来描述资源之间的关系。语义化词表的关联数据发布也可基于SKOS语言,其能够把叙词表内容转换为机器可读可理解的RDF三元组格式,并且对不同词表提供统一的语义互操作,而且能够将词表数据以关联数据形式在网络上发布并与其他资源(如语义化书目元数据)建立关联关系。

该过程分为两部分,一是利用SKOS术语表示受控词表之间的映射关系,二是将SKOS转化为RDF描述。两个词表之间的映射关系可以分为等同、等级及相关3种关系,分别有SKOS语义关系标签定义这些关系,如skos:broadMatch、skos:narrowMatch表示

两个资源或知识组织系统之间具有上下位匹配关系, `skos:exactMatch`、`skos:closeMatch`表明具有精确或近似匹配关系, `skos:relatedMatch`表明具有相关关系, 依此表示不同词表的两个概念之间的对应关系。SKOS是简单的结构化语言, 将各部分内容与SKOS词表的标签一一对应, 写入相应标签, 利用程序来对其进行词表向SKOS的自动转化, 形成RDF格式的文档。此阶段还要进行数据SKOS语义化描述的质量管理。在对传统知识组织系统进行SKOS转化过程中, 难免会出现语法、字符、编码等格式问题, 从而造成不能够满足SKOS的基本语法规则。因此, 需要质量管理来确保知识组织系统语义转化的准确性。qSKOS是基于Java的SKOS文件质

量检测工具, 支持对SKOS中的卷标和文件、结构、关联数据、SKOS等4个方面的一致性检验, 包括“语言卷标遗漏或无效”“链接缺失”“坏连结”“资源未定义”等具体问题, 进行质量验证和错误纠正以后, 可为关联数据发布奠定一定的基础^[34]。

以《中国分类主题词表》主题词“鲁迅”为例(见图2), 将其各部分SKOS模型化之后, 对叙词的属性进行映射, 用关系标签进行标注, 使得概念之间通过等同关系、等级关系以及相关关系相互连接起来, 并形成SKOS形式的RDF文档, 将词表各部分通过这种方式构成一个完整的语义网络。

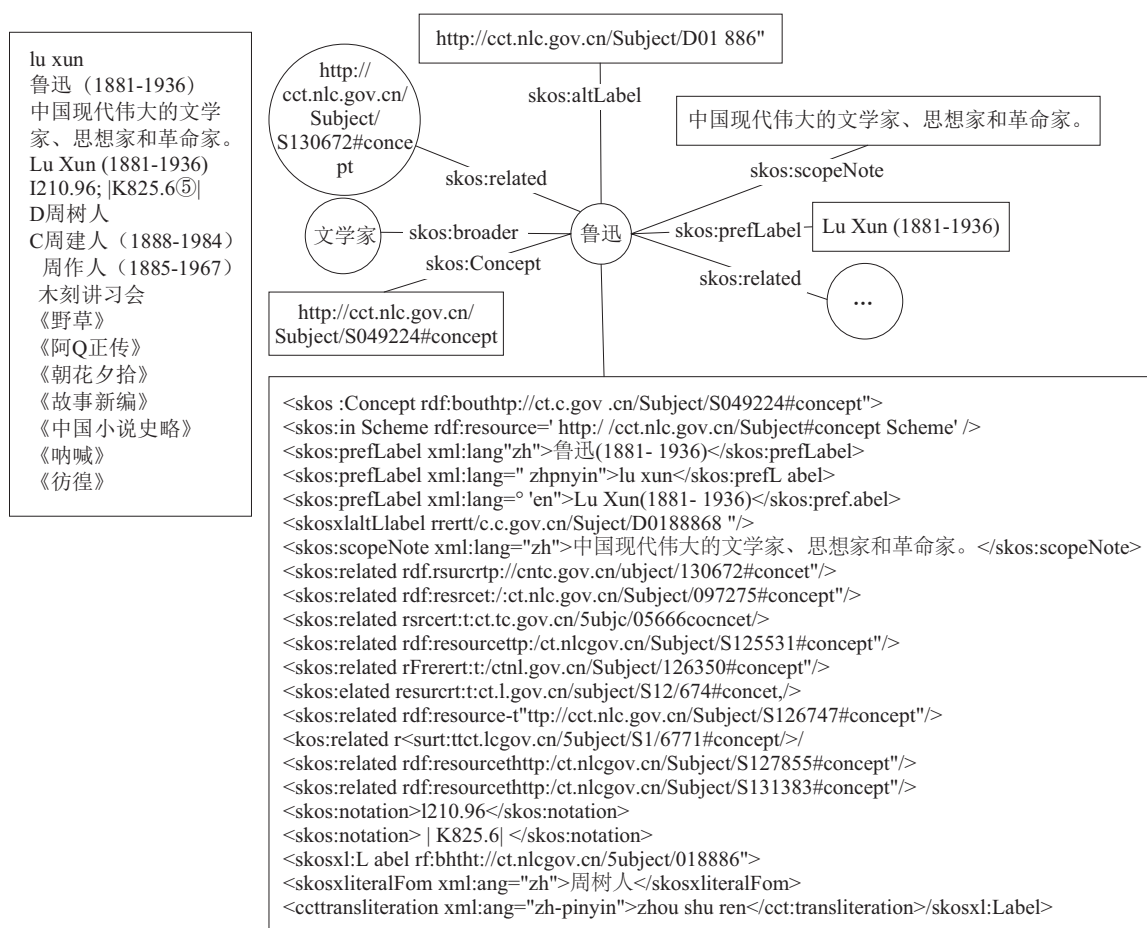


图2 《中国分类主题词表》主题词“鲁迅”SKOS描述及映射过程

2.3 关联数据发布

关联数据有利于在数据网络环境中的资源导航, 提高资源可获取性, 通过与更大的数据集建立链接, 可以促进相关知识的发现。将知识组织系统发布为关联数

据, 不仅提供概念体系的可视化浏览, 还提供SPARQL端点供用户下载数据, 以较好地嵌入外部应用。作为RDF数据集的一种类型, 各类RDF数据集的关联发布方法均适用于词表。从关系数据库的输出角度来看, 包括文本、数据库及RDF文件(包括rdf/xml、owl、skos

等)。目前已有多种关联数据发布工具被开发出来,这些工具均支持URI到RDF描述的逆向引用,其中一些工具还为数据集提供SPARQL访问并支持RDF集的发布。这些工具包括D2R Server服务器、OpenLink Virtuoso、Virtuoso RDF视图、Tail Platform、Vapour、Pubby、Triplify、SparqPlug、OAI2LOD Server、SIOC Exporters等,其中D2R Server用于将关系型数据库转换为关联数据,Virtuoso Universal Server提供RDF模式的关联数据的存储与检索服务,这些数据可以直接存储到Virtuoso服务器中,也可以存储到关系型数据库中,然后映射为关联数据。成熟的SKOS词表关联数据检索系统Skosmos由芬兰国家图书馆负责开发,能完成RDF资料访问和SPARQL检索,目前已经被用于包括EUROVOC、LCSH、STW、AGROVOC等在内的13种通用或者专业叙词表的SKOS检索服务,实现词表知识概念的中英文浏览、查询和检索、可视化等应用服务。

目前所提供的关联数据服务平台主要有3种类型。词表提供者服务平台,如美国国会图书馆关联数据服务、MESH关联数据服务、Getty LOD Vocab服务; AAT、TGN、ULAN、垂直领域词表服务平台,如VIAF、文化遗产关联数据词表、斯坦福大学计算机及生物医学信息研究所的生物医学本体;词表注册服务平台,如LOV。

3 传统知识组织系统关联数据化应用发展

传统知识组织系统关联应用发展可概括为:提高受控词表的访问量,提供对数据集的描述表示,促进数据集成实现。

3.1 提高受控词表访问量

传统知识组织系统发布为关联数据之后,通过自身数据的开放与外部数据的互联,增加了对受控词表的访问入口,服务对象由原来的机构内部扩展至网络全体用户,扩大了用户对词表的访问量,可以被越来越多的资源所应用,方便用户使用,促进了知识组织系统的共享,未来可产生更多价值。目前有多个传统知识组织系统发布了自己的浏览器,如MeSH Browser、STW Browse、AAT Brower等;也有美国国会图书馆在其网页发布了ID.LOC.GOV关联数据服务平台,将LCSH、

LCC、LCMPT、TGM等多个知识组织系统进行集成,方便常用本体、受控词汇表和其他书目描述列表的交互式机器访问。将知识组织系统与大规模百科数据库如Wikidata等建立互联,扩展了知识组织系统的访问入口,可以提高词表数据的开放性、可获取性和可视化。

3.2 提供对数据集的描述表示

数据集是结构化数据的集合,数据集中的每一个资源由一个实体元素(如类或属性)和该元素的一个“值”组成。使用的元素通常从元数据元素集中的词表进行选择,这些词表重点在于为数据集提供类和属性,如DC(都柏林核心元数据)词表、FOAF词表等。元素的“值”取自值词汇表,值词汇表定义了用作元数据记录中元素值的资源(如主题实例、艺术风格或作者)。值词汇表包括叙词表、分类法、规范文档、人名地名录等类型的知识组织系统。因此,知识组织系统可以作为值词汇表来提供对数据集的描述表示,可以填充数据集元数据记录的“构建块”,许多数据集要求用特定的值词汇表来作为特定的元数据元素选择值,因此,值词汇表代表一个元素允许值的“受控列表”,LCSH定义了文学作品的主题,如旅行等;《艺术和建筑词典》(AAT)定义了作品的艺术风格,如印象派等;虚拟名称规范档(VIAF)定义了规范名称,如马克·吐温等。通过共享词表术语可以促进词表重用,一定程度上可以提高数据集之间的互操作能力,避免表述的歧义和冲突,提升数据集的质量,促进了知识组织系统的再利用,促进了知识共享,为进一步的应用发展提供了广阔空间。

3.3 促进数据集成实现

知识组织系统关联数据化可以促进数据集成实现,在此背景下,大规模的基于云平台的数字资源整合项目如Europeana、Hathitrust、SaaS、LoCloud等发布使用。这些平台要求提供者使用一个统一的元数据数据模型来描述数字资源,并且为了将对象集合与外部资源(如维基百科)之间和上下文信息之间建立链接,通过使用相似概念或相同的知识组织系统来促进数据重用和数据共享。以Europeana为例,欧盟数字图书馆云平台是欧盟图书馆开展的一个合作项目,整合了欧洲的博物馆、图书馆与档案馆的文化遗产数据,目前已

获得来自超过2 300个机构的超过3 000万的数字化对象,并允许第三方利用其汇集的数据开发工具和服务,该平台都采用RDF数据格式描述数据资源,将对象描述链接到知识组织系统的核心词汇表,当描述对象和其他数据资源相关时,通过资源间建立链接来获得更多的相关数据,并将多个知识组织系统链接起来丰富了对对象描述元数据,用知识组织系统术语来丰富了上下文信息,采用适当的搜索机制(如语义搜索)利用对象描述和上下文数据,使用户能够检索出需求匹配程度更高的对象,提高检索效率,扩大知识共享范围,目前在Europeana平台中使用的知识组织系统有VIAF、Geonames、TGN、AAT、DDC、UDC、ULAN等。

4 总结

传统知识组织系统仍是信息组织和服务的基础,其通过与各种类型信息资源之间建立关联,达到对信息资源的集成管理、深度挖掘和高效利用,对于更好地建立信息资源索引、浏览和展示信息资源等方面将具有重要的作用。知识组织系统的关联化可以用于资源发现、词表的构建与维护(重用)及自动标引等知识组织处理过程中,所形成的词表网络,可以有效地提高词表的价值。传统知识组织系统SKOS化,使其成为用于定义数据描述的潜在语义结构,作为值词表或属性词表,为数据集描述提供属性、属性值取值来源,作为语义知识地图,方便人或机器浏览领域内知识体系。SKOS的一致性与良好兼容性为其实现提供了有利条件。通过在不同知识组织系统之间建立明确的对应关系,从而提高了互访效率,有助于实现跨库、跨机构的检索,从而满足用户一站式浏览和检索信息的需求。目前已有成熟的理论与方法支持受控词表的关联数据化,但是就当前应用看,如何充分利用以关联数据表达的知识组织系统资源,还需进一步在理论与实践上深入探讨。

参考文献

- [1] Planet Data. State of the LOD Cloud 2019 [EB/OL]. [2019-12-01]. <http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/>.
- [2] VANDENBUSSCHE P Y. Linked open vocabularies [EB/OL]. [2019-12-01]. <http://lov.okfn.org/dataset/lov/details/vocabulary/skos.html>.
- [3] HARPER C A. Encoding library of congress subject headings in SKOS: authority control for the semantic web [C] // Proceedings of the 2006 International Conference on Dublin Core and Metadata Applications: Meadata for Knowledge and Learning. Dublin Core Metadata Initiative. 2006: 89-94.
- [4] SWAD-Europe Thesaurus Activity. RDF thesaurus prototype [EB/OL]. [2020-03-20]. <http://www.w3.org/2001/sw/Europe/reports/thes/8.7>.
- [5] ASSEM M V, VÉRONIGUE M, MILES A, et al. A method to convert thesauri to SKOS [EB/OL]. [2020-02-01]. http://link.springer.com/chapter/10.1007/978-3-540-30475-3_3.
- [6] 范炜. 语义网环境中的叙词表实例研究——利用SKOS构造机器可理解的知识组织体系 [J]. 情报科学, 2006 (7): 1073-1077.
- [7] 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.
- [8] 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, 38 (3): 57-68.
- [9] 段荣婷. 基于简约知识组织系统的主题词表语义网络化研究——以《中国档案主题词表》为例 [J]. 中国图书馆学报, 2011, 37 (3): 54-65.
- [10] 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.
- [11] 喻菲. 《中国分类主题词表》网络化研究: 从MARC到SKOS [C] // 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.
- [12] 白华. 基于SKOS方法的分类法本体描述研究 [J]. 图书情报工作, 2012, 56 (1): 120-124, 147.
- [13] 施国良. 基于分面分类应用的SKOS扩展模型 [J]. 图书情报工作, 2009, 53 (10): 95-97.
- [14] SUMMERS E, ISAAC A, REDDING C, et al. LCSH, SKOS 和关联数据 [J]. 数据分析与知识发现, 2009, 3 (3): 8-14.
- [15] MORSHED A, KEIZER J, JOHANNSEN G, et al. From AGROVOCOWL Model towards AGROVOC SKOS Model [EB/OL]. [2019-07-10]. <http://www.fao.org/docrep/012/al300e/al300e00.pdf>.
- [16] SMEDT J D. SKOS extensions for the EUROVOC thesaurus [EB/OL]. [2019-07-04]. <http://www.athenaeurope.org/getFile.php?id=663>.
- [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, 35 (6): 56-60.

- [18] 刘春艳, 陈淑萍, 伍玉成. 基于SKOS的叙词表到本体的转换研究[J]. 现代图书情报技术, 2007(5): 32-35.
- [19] 刘华梅. 《中国分类主题词表》主题词SKOS化描述及自动转换研究[J]. 图书馆建设, 2014(8): 29-32, 36.
- [20] 杨雪莉. 《中国图书馆分类法》的SKOS化描述研究[J]. 图书馆论坛, 2015, 35(10): 43-48.
- [21] 贾君枝, 杨洁, 卫荣娟. 《汉语主题词表》简单知识组织系统表示的自动转换设计[J]. 情报理论与实践, 2011, 34(5): 14, 54-57.
- [22] NEUBERT R. Bringing the "Thesaurus for Economics" on to the Web of Linked Data [EB/OL]. [2020-03-01]. doi:http://dx.doi.org/.
- [23] ALBERTONI R, DE MARTINO M, DI FRANCO S, et al. EARTH: An environmental application reference thesaurus in the linked open data cloud [J]. Semantic Web, 2014, 5(2): 165-171.
- [24] MORSHED A, ZAPILKO B, JOHANNSEN G, et al. Evaluating approaches to automatically match thesauri from different domains for linked open data [EB/OL]. [2019-11-01]. https://at-web1.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2011/presentations/NKOS2011_Zapilko_Mayr_et_al.pdf.
- [25] GREENBERG J, LOSEE R, AGÜERA J R P, et al. HIVE: Helping inter-disciplinary vocabulary engineering [J]. Bulletin of the American Society for Information Science and Technology, 2011, 37(4): 23-26.
- [26] 欧石燕. 中文叙词表的语义化转换[J]. 图书情报工作, 2015, 59(16): 110-118.
- [27] 周峰. 地学文献资源的语义关联数据构建研究[D]. 北京: 中国地质大学(北京), 2016.
- [28] 贾君枝, 赵洁. DDC关联资料实现研究[J]. 中国图书馆学报, 2014, 40(4): 76-82.
- [29] 鲜国建, 赵瑞雪, 朱亮, 等. 农业科学叙词表的SKOS转化及其应用研究[J]. 现代图书情报技术, 2012(10): 16-20.
- [30] 黄华军, 曾新红, 林伟明, 等. 中文知识组织系统形式化语义描述标准体系研究(二)——分类法共享服务系统CLSS研究与实现[J]. 中国图书馆学报, 2015(2): 17-28.
- [31] 王知津, 赵梦菊. 论知识组织系统中的语义关系(上)[J]. 图书馆工作与研究, 2014(8): 65-69.
- [32] TUOMINEN J, FROSTERUS M, VILJANEN K, et al. ONKI SKOS Server for Publishing and Utilizing SKOS Vocabularies and Ontologies as Services [EB/OL]. [2020-03-01]. DOI: 10.1007/978-3-642-02121-3_56.
- [33] BAKER T, BECHHOFFER S, ISAAC A, et al. Key choices in the design of Simple Knowledge Organization System (SKOS) [J]. Journal of web semantics, 2013, 20(5): 35-49.
- [34] 石泽顺, 肖明. 基于网络叙词表的图情学科SKOS构建与可视化研究[J]. 情报学报, 2018, 37(3): 274-284.

作者简介

贾君枝, 女, 1972年生, 博士, 教授, 研究方向: 信息组织与信息检索、关联数据, E-mail: junzhij@163.com。
李衍, 女, 1995年生, 硕士研究生, 研究方向: 信息组织与信息检索、关联数据。

The Development of Linked Data of Traditional Knowledge Organization System

JIA JunZhi^{1,2} LI Kan¹

(1. School of Economics and Management, Shanxi University, Taiyuan 030006, China; 2. School of Information Resources Management in Renmin University of China, Beijing 100872, China)

Abstract: The traditional knowledge organization system with controlled word list as the core plays an important role in the description and presentation of resources. This article through to a large number of research literature on the basis of investigation and research, summarizes the domestic and foreign correlation digital development related research of knowledge organization system, and through summarizing the content of the related research, puts forward the traditional semantic relevance knowledge organization system implementation steps, and clear the traditional knowledge organization systems association takes SKOS model, RDF serialization and associated data released three main steps, and carries on the detailed description of the parts, so as to provide reference for other related traditional knowledge organization system digital. The paper summarizes that the application of traditional knowledge system association is mainly reflected in improving the page view of the controlled word list, effectively realizing the description and representation of the data set, and promoting the realization of data integration.

Keywords: Knowledge Organization System; Linked Data; SKOS

(收稿日期: 2020-02-15)