Digital Library Forum 数字图书馆论坛
http: // www.dlf.net.cn

知识组织的工具及其语义互操作方法体系

□ 王景侠 / 南京政治学院上海校区军事信息管理系 上海 200433

摘要：网络环境下，知识组织所用的工具主要是各类知识组织系统，同时知识组织系统之间的互操作已成为知识组织研究和应用中的热点问题。文章在分析知识组织工具的基础上重点阐述了传统知识组织工具以及本体为代表的现代知识组织工具的语义互操作方法，以期为图书馆等信息机构在进行数字资源整合、资源共享和知识服务时提供参考。

关键词：知识组织，知识组织系统，KOS，互操作
DOI：10.3772/j.issn.1673—2286.2013.05.007

网络环境下，图书馆等信息机构除了要对实体资源进行知识组织外，还要对越来越多的馆藏数字资源尤其是网络资源进行知识组织。资源的知识组织是通过分析揭示知识之间的内容特征，并利用各种工具将知识之间各种关系构成语义网络，进而将知识组织成一个相互联系的知识体系。目前，知识组织工具的互操作已成为图书情报学界知识组织、数字资源集成、资源共享和知识服务等领域理论和应用研究的一大热点，本文试对知识组织的工具以及知识组织工具之间的互操作及其语义互操作方法进行探讨。

1 知识组织的工具

“知识组织”（Knowledge Organization）一词最初由美国图书馆学家、分类学家布利斯（H. E. Bliss）于1929年提出。1989年，国际知识组织学会（ISKO）成立，并于1993年将《国际分类法》（IC）更名为《知识组织》（KO）作为其会刊出版。知识组织是指按照知识的内在逻辑联系，运用一定的组织工具、方法和标准对知识对象进行整理、加工、表示、控制等一系列的序化、系统化的活动[1]。无论是传统图书馆还是现代数字图书馆，知识组织一直是图书馆的核心工作。网络环境下，伴随着数字化、网络化资源大量出现，图书馆等机构的知识组织工具也在快速发展、不断丰富。为适应数字资源的组织需求，一方面分类法、叙词表等传统知识组织工具不断改造，另一方面基于网络的概念地图、语义网络、本体和大众分类法等新型语义组织工具不断涌现，并催生了能囊括所有工具的术语——“知识组织系统”（Knowledge Organization System，简称KOS）。

1.1 传统知识组织工具

传统知识组织工具既包括分类法（如DDC、LCC、UDC、CLC等），也包括叙词表、标题表等受控词表（如LCSH、MeSH、《汉语主题词表》、《中国分类主题词表》等）。前者主要从学科体系的角度组织和揭示资源的内容特征，而后者则从内容角度标引和组织序化信息资源。均在纸本环境下产生并发展起来的分类法和受控词表，对实体资源的组织和控制可谓达到了炉火纯青的地步。网络环境下，经过改造后的分类法和受控词表虽然也能对网络资源进行知识组织，但其功能仅限于网站或知识库的导航与浏览[2]。

1.2 现代知识组织系统

诞生于网络环境下的知识组织系统，或称为知识组织体系，是对各种人类知识结构进行表达和有组织阐述的语义工具的统称，包括分类表、叙词表、语义网络、本体以及更泛指的情报检索语言、标引语言[3]。知识组织系统这一术语虽然最早由网络环境中的知识组织系统/服务研究小组在ACM（美国计算机协会）1998年的数字图书馆会议上提出，但事实上知识组织系统在图书馆界早就存在，例如传统的文献分类法、主题词表、术语表等都是一种知识组织系统。因此，知识组织系统这一术语实际上是各类知识组织工具的统称，不仅包括分类表、标题表等传统知识组织工具，也包括网络环境下产生的以本体为代表的新型知识组织工具。

2 知识组织系统的互操作

在开放、动态、分布式的网络环境下，由于其所组织的资源对象也存在结构、功能、表示、领域应用等方面的差异，所以也导致了不同的知识组织系统之间越来越严重的异构性，而随之带来了资源集成和共享的困难，所有这些均需互操作（interoperability）的技术方法来解决。目前，互操作已成为国际知识组织领域的研究热点，同时也是网络环境下知识组织必须解决的问题[4]。

2.1 知识组织系统互操作的概念

所谓知识组织系统的互操作，就是指不同知识组织系统之间的兼容互换，即在不同的分类表、叙词表、本体等知识组织工具中实现兼容互换。当前，各种数字图书馆、主题网关（学科信息门户）以及搜索引擎等成为业界研究的热点和建设的重点，而如何通过异构知识组织系统间的互操作，为用户提供跨库、跨系统、跨语言的浏览与检索是图书情报界正在重点关注的一个热点课题[5]。知识组织系统的互操作能够解决知识组织系统应用中面临的多语言、异构性和跨领域三个问题，即多类型、多语言的知识组织系统之间的互操作，不仅是实现分布式信息系统交叉浏览和集成检索的有效方法，也是目前知识组织系统建设的主要内容。国外学术界以及信息机构对知识组织系统的互操作一直十分重视，开展了大量的研究计划，有许多研究成果也已得到应用，例如欧盟的DESIREII和Renardus、英国的HILT，以及美国的UMLS等。进入21世纪，我国也有许多机构和研究项目正在积极开展知识组织系统的互操作及其应用研究。例如国家图书馆的《NKOS的构建与应用规范研究》项目，其主要目的就是使得国家数字图书馆具备和国内外相关机构进行知识组织工具互操作的能力，同时具备发布关联数据、进行语义数据整合并提供基于内容的智能服务的能力。再如中国科学技术信息研究所开展的“《汉语科技词系统》建设与应用工程”项目重点研究网络环境下叙词表的编制方式和应用领域，为网络时代《汉语主题词表》的修订寻求新的编制方法，同时还提出将《汉语主题词表》转化为本体的设想，目标是通过本体促进知识组织系统的整体进步[6]。

2.2 知识组织系统互操作的功能

知识组织系统互操作的目的主要是为了解决分布式异构系统（如数据库、数字图书馆）的一站式检索，其主要功能如下：一是实现对检索系统的一站式检索和交叉浏览功能；二是通过信息知识集成创建新的知识组织系统；三是提供术语服务。其中，通过Web服务（Web Service，一个基于HTTP协议和XML语言的协议标准）技术在网络上提供分布式的词汇服务是网络知识组织系统（NKOS）服务的一种主要形式。目前已经提供这类服务的词表有《农业多语种叙词表》（AGROVOC）、AAT、CSA/NBII生物复杂性词表（Biocomplexity Thesaurus）、美国国家农业词表（NAL），以及亚历山大数字图书馆项目中的ADL地名词表等。

3 知识组织系统的语义互操作方法体系

秉承第一部分对知识组织系统的划分，本部分重点对传统知识组织系统间和现代知识组织系统中的本体间语义互操作的方法进行论述。

3.1 传统知识组织系统间的语义互操作方法

传统知识组织系统间的互操作既包含同构系统之间也包括异构系统之间的互操作，可以通过多种方式实现，综合国内外KOS互操作研究与实践，常见的互操作方法可以归纳为KOS的演化方法、KOS的映射转换方法、协议连接方法和临时列表方法四类[7,8]。

3.1.1 KOS的演化方法

KOS的演化是为了满足特定需求而对原有KOS进行改造，使得新建KOS与原有KOS形成对应关系，具体又包括下列方法：

（1）继承/仿建法

继承是以现有的复杂的词表为原型来创建专业的或简单的词表，可分为扩展的继承和限制的继承两种。其中，扩展的继承是以原知识组织系统整体为基础的扩展或是以原知识组织系统一部分为基础的扩展。例如，LCSH是迄今为止国际上使用最为广泛的标题表，其通用的受控语言已成为许多国家创建词表的模式。OCLC领导开发的FAST（主题术语的分面应用）研究计划，是一个枚举的、基于LCSH派生出的分面主题标目系统，旨在使LCSH的句法简单化，在保留LCSH丰富词汇的前提下，使词表更易理解、控制和使用。它采用与LCSH向上兼容的方式，每一条正式的LCSH标题都能转换为FAST标题词。目前，FAST已被作为关联数据发布，并根据开放数据共享署名许可来开放使用，可供人类和机器更加便捷地获取FAST主题。此外，OCLC还发布了一个从LCSH到FAST格式的示范转换工具，并更新了网络搜索界面以加强对FAST词表的支持[9]。

（2）翻译改编法

翻译改编法是从其他语言的词表翻译、改编形成自己的词表，许多非英语主题标目都是从LCSH翻译发展而来的。例如，DDC作为世界范围内使用最广泛的图书馆分类法，已被译成30余种文字（我国也曾翻译过其21版）。此外，许多国家也已将MeSH译为本国的语言，目前的译本有法、德、意、葡、俄、西、中、日、韩等语言，可以建立跨语言的医学检索系统。2002年，我国在《农业科学叙词表》的基础上完成了AGROVOC的翻译，基本实现了《农业科学叙词表》与AGROVOC词表之间的互操作。

（3）系列化分类表或集成词表法

系列化分类表或集成词表法是在一个系统内通过有效地组织实现系统内部兼容，如《中分表》作为我国用户最多、影响广泛的分类主题一体化词表，起到了不同程度上兼容各种专业分类表和叙词表的作用[10]。再如，《军用主题词表》与20部专业词表实现了有效兼容，《国防科学技术叙词表》在编制机读版时将11部词表集成起来实现兼容，《中图法》编制的系列版本实现兼容等。

（4）微词表法

微词表法也称卫星词表法，即从较大的词表或分类表中抽取一部分作为新编专业词表的主体或构成部分。微词表的主导思想是将各专门化的词表作为一个上层结构的卫星表，新编词表与较大词表犹如子表和母表，兼容性较好，专业词表在大型词表结构内有机联系在一起。从最初的应用角度看，一个微词表是一组专门化的词汇，是从一个更大的词表中抽取出来的，所以能够与大词表兼容，而且可以完全被容纳在该大词表的等级结构中。例如，《立法词汇索引》就是由LCSH中立法相关的部分进行扩展而成的。

（5）宏词汇法

宏词汇法的思想与微词表相似，但实现方法与其相反。宏词汇的想法仅是创造一种词的“属”的上层结构，它可以包括一组不同领域的叙词表或其他类型的词汇，各专业词表中的词在这一上层结构下互相联系起来。例如，美国UMLS采用的方法与宏词汇法相似。

3.1.2 KOS的映射转换方法

KOS的映射转换方法是直接以互操作为目的的方法，也是目前KOS互操作最常见、最有效的方法，具体又包括如下方法：

（1）直接映射

直接映射是指直接在不同知识组织系统间的词语之间或者词语与分类号之间建立等价关系。例如，OCLC的LCSH-ERIC计划就是采用MARC21规范数据格式，将ERIC叙词表转换为MARC21格式，与LCSH主题词表进行匹配，建立起两者词汇间的链接关系。DDC与LCC、DDC与MeSH、DDC与NLMC（《美国国家医学图书馆分类法》）均采用了直接映射方法。

（2）共现映射

共现映射是指为不同系统中的共现词汇而建立的映射。它通过知识组织系统词语在元数据记录中的共现关系建立术语间的映射关系，相当于整合词表模式，将多种或一个网络的所有检索语言或数据库的词汇混合按字顺排列，并注明某词出现在哪些词表或数据库中及其累计标引的频率。这种共用词典对选择数据库进行跨文档检索特别有用。例如，OCLC从LCSH到LCC的映射研究计划采用的就是这种方式，而且已被美国一些主要的联机书目服务中心所采用。国内学者开展的基于文献语料的术语映射也是一种共现映射[11]。与直接映射相比，

共现映射是一种更为松散的映射方法。OCLC在DDC与LCSH的映射中，便采取了直接映射与共现映射相结合的方法。

（3）中心转换

中心转换是将参与互操作的多个知识组织系统映射到一个共同选定的中心知识组织系统上。这样，两个知识组织系统之间的互操作就可通过中心知识组织系统实现转换。例如，欧盟Renardus项目就是利用DDC作为转换中介词典，实现了跨库检索和浏览。

3.1.3 协议连接方法

各种协议是系统间信息交流的基础，不同系统间的互操作也需要由各种协议来实现信息交换和通信。该方法指通过建立知识组织系统服务协议供其他应用程序访问，创建不同知识组织系统的连接环境，从而实现互操作。例如，欧洲图书馆员会议（CENL）主办及赞助的MACS计划，就旨在通过创建连接管理系统与查询系统实现图书馆目录的多语言主题检索。该计划的领导者是瑞士国家图书馆，成员有法国、英国和德国三家国家图书馆，通过分析SWD、RAMEAU、LCSH三种标题表之间的匹配关系，建立三者间的等同连接，允许用户以法语、英语、德语三种语言连接主题标目，以解决数据库的多语种主题访问，实现图书馆资源的共享。

3.1.4 临时列表方法

临时列表是一种散点式的映射方案，即根据查询词临时从不同的知识组织系统中提取相匹配的对象，组建临时对应列表。

在以上四类互操作方法中，KOS的演化虽然其初衷并不是以互操作为目的，但客观上却支持了KOS互操作；对于独立创建的知识组织系统，映射和服务协议是实现知识组织系统互操作的主要方式，即当参与互操作的知识组织系统比较明确，如在几个特定的机构间进行资源共享时，映射方式比较适用；而当参与互操作的知识组织系统并不明确，如知识组织系统的拥有者只是希望提供一种知识组织服务而并不明确自身的知识组织系统要与哪些知识组织系统进行互操作时，协议方式较为合适；临时列表基于对查询提问的字面匹配、互操作的效率和准确性虽然不高，但实现起来比较简单。以上四类互操作方式的比较如表1所示[12]。

表1 知识组织系统互操作实现方式比较

实现方式	KOS演化	KOS映射	协议连接	临时列表
以实现互操作为目的	否	是	是	是
参与的KOS是否独立生成	否	是	是	是
是否保存映射或连接	是	是	是	否
参与的KOS个数是否固定	是	是	否	否
实现的自动化程度	手工	人机	协议	机器
适用范围	KOS的改造和利用	KOS的互操作	知识组织、知识服务	简单联合检索

需要强调的是：①在具体的资源共享活动中需要根据互操作方式的特点和适用范围从自身实际出发来选择合适的方式；②在知识组织系统互操作实践中，采用的方法并不局限于一种模式，有时甚至可以采用两种或多种方法组合使用。例如，美国加州大学CERES研究计划和美国国家生物信息基础（NBII）就分别采用了继承法与微词表法来创建集成环境的叙词表。

3.2 本体间语义互操作方法

为有效消除众多本体之间的异构性，需要解决本体之间的语义互操作问题。本体间的语义互操作是通过语义整合实现的，即确认和实现两个本体系统之间的共同部分。本体之间的语义互操作不仅是本体研究中的重要课题，同时也是网络知识组织系统研究中的热点问题。综合国内外的相关研究，本体的语义互操作目前一般可以采用本体映射、本体调整和本体合并三种方法[13]。其中，本体映射形式比较灵活，更适合分布式动态的网络信息环境，是当前解决本体语义互操作问题的最有效的方法。

3.2.1 本体映射

本体映射（Ontology Mapping）是本体集成、整合的一种方法，是指两个本体存在语义的概念关联，通过语义关联，实现将源本体的实例映射到目标本体的过程。本体映射通过建立本体间的映射规则达到本体互操作，可以解决不同本体间的知识共享和重用。其方法是找出不同本体中实体之间的语义关系，并采用形式

化方法将其表达出来。目前，国内外本体映射系统已经有了一些较为成熟的应用，国外比较著名的本体映射系统有GLUE、MAFRA、S-Match、COMA等，国内的有RiMOM、Falcon等。

本体间的映射关系也包括一对一、一对多、多对一以及多对多的映射等，其中最简单也是最常用的映射关系是一对一的映射，这是一种等价映射，异构本体的等价成分在互操作过程中可以直接互相替代[14]。

影响本体映射效果的因素主要有两个：一是映射本体之间的异构程度（这与语义互操作负相关，即异构程度越高，映射的难度就越大，语义互操作程度就越难）；二是本体映射方法的成熟程度（这与语义互操作正相关，即成熟度越高，映射的效率就越高，语义互操作程度就越好）[15]。而要实现本体映射过程的自动化，系统可以通过机器学习或其他技术实现对结果的自动修正和完善，当两个本体使用半自动化或自动化的方法获得映射后，就可利用本体映射来实现本体对齐甚至本体合并。

3.2.2 本体对齐

本体对齐（Ontology Alignment），又称为本体匹配（Ontology Matching）。两个不同本体的对齐是指两个系统的概念、属性或关系之间的映射的集合。此处对齐的概念、属性或关系可以认为是等值的。值得注意的是，本体对齐并不是关系实体间关系的集合，而是映射的集合。有时为了便于对齐两个系统，必须引入新的子类或父类，但在对齐的过程中，不需改变任何一个本体系统的公理、定义、证明或计算[16]。从本体对齐的概念可以看出，本体对齐方法实际上就是找到所有的映射的方法，一般多为半自动化的方法。本体对齐将两个本体正确地连接起来，可以实现多个本体的联合查询等。

3.2.3 本体合并

本体合并（Ontology Merging）是指在两个不同的本体A和B中发现共同之处并产生一个新的本体C，即一体化的本体，以有助于实现分别基于本体A和本体B的系统之间的互操作。本体C可以是两个本体系统的中间本体，也可以取代本体A或本体B[17]。本体合并一般只用于本体构建过程中将多个本体合并为大本体，或是本体维护阶段将一个迷你本体（mini-ontology，小部分概念和关系）合并到原本体中以对本体进行更新。除了这两种情况，学界提出的本体集成多数是指本体对齐。本体合并与本体对齐最大的区别在于，本体对齐并不生成新的本体，只是在原有的需要对齐的本体之间建立一个映射集合来达到本体互操作，而本体合并则在原有本体的基础上根据实际应用的需要生成一个新的本体。

本体合并的方法是得到映射后根据应用需要执行映射合并本体，并在执行之后进行后处理。目前，本体合并的方法主要有基于范畴论的方法和基于形式化概念分析（Formal Concept Analysis，FCA）的方法。范畴论和FCA都是抽象处理结构和结构之间关系的理论。其中，基于范畴论的本体描述方法可在某种程度上实现自动的本体集成，本体及本体之间的映射构成了范畴，利用范畴论的“态射”方法实现本体映射，“外推”方法实现本体合并，因而利用本体“外推”也是本体合并的一种方法[18]。而FCA是应用数学的一个分支，它源自哲学领域对概念的理解。概念格作为FCA方法中核心的数据结构，从外延和内涵两方面对概念进行符号形式化描述，实现语义信息的计算机可理解。采用FCA技术进行本体的构建、合并、三维可视化展示，可提高用户在合并本体中的查询效率[19]。但是，FCA对同义词（近义词）关系分析不够，生成的本体语义信息也不够丰富，这些都对以后利用本体进行推理带来很大不便，因而FCA方法仅适用于处理轻量级本体（即不具备逻辑推理功能的本体，如叙词表和WordNet等）。

4 结语

网络环境下，知识组织工具间的语义互操作已成为图书馆等机构特别是数字图书馆建设需要解决的重点问题。与传统知识组织工具间的语义互操作相比，本体间的语义互操作研究相对还不够成熟。无论是传统知识组织工具间的语义互操作还是现代知识组织工具间互操作的方法都在不断发展，共同构成一个日益完善的语义互操作方法体系。但是，知识组织工具间互操作特别是语义互操作问题的解决将是一个复杂、长期的、不断完善的过程。相信随着本体技术和语义万维网研究的深入，随着图书馆等机构间的协作以及标准化的进程，知识组织工具间的语义互操作问题将会得到更好地解决。

参考文献

[1] 毕强,牟冬梅,韩毅.下一代数字图书馆知识组织[M].吉林教育出版社,2009:18.
[2] 王松林.资源组织[M].国家图书馆出版社,2011:10.
[3] 司莉.知识组织系统的互操作及其实现[J].现代图书情报技术,2007(3).
[4] 戴维民,包冬梅.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综述[J].图书馆杂志,2009(12).
[5] 司莉.知识组织系统的互操作及其实现[J].现代图书情报技术,2007(3).
[6] 戴维民,包冬梅.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综述[J].图书馆杂志,2009(12).
[7] 司莉.知识组织系统的互操作及其实现[J].现代图书情报技术,2007(3).
[8] 侯汉清,刘华梅,郝嘉树.60年来情报检索语言及其互操作进展(1949-2009)[J].图书馆杂志,2009(12).
[9] OCLC将FAST(主题术语的分面应用)作为关联数据发布[J].现代图书情报技术,2012(1).
[10] 刘华梅,侯汉清.基于受控词表互操作的集成词库构建研究[J].中国图书馆学报,2009(3).
[11] 薛春香,乔晓东,朱礼军.KOS互操作中的术语映射研究综述[J].现代图书情报技术,2010(2).
[12] 王军.数字图书馆的知识组织系统:从理论到实践[M].北京大学出版社,2009:84.
[13] 李冠宇,李琳,郭立群,等.基于映射的实例转换研究[J].计算机工程与应用,2010(6).
[14] 于光杰,胡静娴,朱天.本体映射技术研究[J].福建电脑,2008(10).
[15] 毕强,牟冬梅,韩毅.下一代数字图书馆知识组织[M].吉林教育出版社,2009:188.
[16] 储荷婷,张茵.图书馆信息学[M].人民大学出版社,2007:89.
[17] 于娟,党延忠.本体集成研究综述[J].计算机科学,2008(7).
[18] 杨先娣,何宁,吴黎兵.基于范畴论的本体集成描述[J].计算机工程,2009(6).
[19] 张瑞玲,徐红升,沈夏炯.基于FCA的本体原型系统的设计与实现[J].计算机工程与应用,2008(19).

作者简介

王景侠（1976-），男，南京政治学院上海校区军事信息管理系讲师，对元数据、知识组织系统，尤其是语义万维网环境下的数字图书馆的知识组织有浓厚兴趣。E-mail: wangtuzhongguo@hotmail.com

Knowledge Organization Tools and the Method System of Their Semantic Interoperability

Wang Jingxia / Military Information Management Department of Shanghai School of Nanjing Political Institute, Shanghai, 200433

Abstract: Under the Web environment, the tools of knowledge organization are mainly all kinds of knowledge organization system (KOS), and interoperability of KOS has been the hot problem between the researches and applications. Based on the analysis of knowledge organization tools, the paper mainly discusses the method of Semantic Interoperability among traditional knowledge organization tools and modern knowledge organization tools represented by ontology, in order to provide reference to various types of information agencies digital resource integrating, resource sharing and knowledge services.

Keywords: Knowledge organization, Knowledge organization system, KOS, Interoperability

（收稿日期：2012-12-05）