

多种知识组织方法比较*

常 娥, 夏 婧

摘 要 文章在阐述知识组织环境与目标发生转变的基础上, 比较各种知识组织方法。主要从概念表达方式、是否关联资源、解决语义异构的能力以及开放性等方面论述各种知识组织方法的特点, 重点探讨关联数据与元数据、知识本体整合的必要性与可行性。

关键词 知识组织 元数据 知识本体 关联数据 知识地图

引用本文格式 常娥, 夏婧. 多种知识组织方法比较[J]. 图书馆论坛, 2016 (8) : 1-6.

Comparative Study of Knowledge Organization

CHANG E, XIA Jing

Abstract On the basis of explaining the change of knowledge organization environment and target, the article compares different kinds of knowledge organization methods. The article mainly discusses the characteristics and relations of various knowledge organization methods from the aspects of the concept expression, relevance to resources, the ability of solving semantic heterogeneity and openness. Finally, the article discusses the necessity and feasibility of the integration of linked data, metadata and ontology.

Keywords knowledge organization; metadata; ontology; linked data; knowledge map

0 引言

知识组织的概念最初由美国图书馆学家布利斯(H.E.Bliss)于1929年提出^[1], 他最早阐述了以图书分类为基础的知识组织思想。随后, 美国图书馆学家谢拉(J.H.Shera)全面论述了知识组织在图书馆工作中的重要作用和方法^[2-3]。以此为基础, 国内外学者就知识组织的概念、原理、目标、研究范围、方法和技术等展开研究, 取得了丰硕的成果。早期研究普遍认为, 理想的知识组织是将文献中的知识单元抽取出来并进行分析, 找到人们创造与思考的相互影响与联系的节点^[4]。但由于未找到知识的有效表达方式, 不得不退却

到文献层次, 用文献间接地表示知识^[5], 并产生了一系列以文献分类为代表的知识组织方法。文献组织的实质是知识组织, 但它不完全等同于知识组织。

随着计算机和网络技术的发展, 人类社会已全面步入数字化时代, 并不断迈向数据化时代。如今数字资源成为文献资源的主要形式, 对于文献知识内容的分解与组合变得极为便利, 知识的控制单位由文献资源深化到文献中的数据、公式、事实、结论等最小的独立单位, 即“知识单元”(或称“数据单元”)。知识组织环境的变化使得知识组织不再局限于文献层次, 知识组织的目标亦由文献揭示与序化转变为知识单元的揭示

* 本文系国家自然科学基金项目“图书馆资源组织中的数据关联机制研究”(项目编号: 14CTQ005)研究成果之一

与关联。例如，电子书的内容定位不再依赖任何形式的顺序编码，而是通过给每个内容块加上标识这种完全结构化的方式进行组织^[6]。

有鉴于此，数字时代图书馆必须突破纸质文献组织理念与框架的樊篱，积极探寻新的知识组织方法，以满足数字资源细粒度知识组织的需求。关联数据(Linked Data)是近年来图书馆学领域最为推崇的知识组织方法，学者们从关联数据的概念、发布与消费技术、应用协议、典型数据集等不同层面进行了分析，然而鲜有文献综合各种知识组织方法，从早期的分类索引到 MARC 机读目录、DC 元数据、本体模型，再到关联数据这一系列知识组织方法的继承与发展角度，论述关联数据的时代适应性。为了深入讨论这一问题，本文在介绍各种知识组织方法的基础上，主要从概念表达方式、是否关联资源、解决语义异构的能力以及开放性等方面对比分析不同知识组织方法的特点，最后重点探讨关联数据与元数据、知识本体整合的必要性和可行性。

1 常用知识组织方法

在图书情报领域中，知识组织方法多达数十种，知识组织工具和系统不胜枚举。下文将知识组织方法划分为元数据、知识本体、知识地图和语义网络 4 大类型予以介绍。由于分类法和主题法的知识组织思想已深刻内化到其他知识方法之中，不再单独阐述。

1.1 元数据

元数据最简明抽象的定义是“关于数据的数据”(data about data)^[7]，其本质含义是指按照相应的句法结构与语义结构组织元素集，使其成为所描述对象的“缩影”或“替代物”，然后利用这个“缩影”或“替代物”来对相应对象进行识别、组织、检索、保存与显示等管理操作，这是元数据的本质职能^[8]。元数据方法擅长揭示文献资源的整体内容，主要用于文献资源的描述、定位、选择与检索。

元数据的元素项主要为隶属关系和相关关系，然而这种语义关系简单、粗糙，且隐含在元

数据标准的语法结构中，灵活性较差。元素与子元素之间的隶属关系在定义具体元数据标准时得以确定，而这些元素所描述对象之间的复杂关系则需要借助外部程序实现。例如，DC 的“主题”元素、CNMARC 的“主题附注”元素都是为了描述文献资源的主题概念，然而所描述主题概念间的相互关系则需要通过其他途径进行显示。元素之间的相关关系以及通过相关关系所描述的对象之间的复杂关系，同样存在着无法进一步识别的问题。例如，DC 通过“关系”元素连接了在内容、版本、形态等方面相关的文献资源，然而这些文献资源之间的复杂关系无法进一步识别。这些是元数据方法的局限所在。

此外，元数据在描述文献资源时，还存在着特殊性与一般性的矛盾，这亦是无法克服的困难。虽然 DC、MARC 等少数元数据格式占据主导地位，然而永远无法统一到仅有的少数几种格式^[9]。由于只有专业的元数据对于特定的领域应用最合适，因此专业领域存在着大量的不同元数据方案，由此引发了不同元数据的映射与互操作问题，这是学术界和实践机构都致力于解决的难题。近年来，国外学者提出了建立“元数据的元数据”机制，以灵活地实现不同信息系统之间的互操作，如美国著名知识仓储 HathiTrust 采用 METS 元数据来编码和传输数据库中的不同元数据标准。然而这只是一个将不同元数据标准整合在一起的上层元数据方案，仍无法完成不同元数据标准之间元素项的语义映射和转换。

1.2 知识本体

20 世纪 90 年代初，起源于哲学范畴的知识本体(Ontology)逐渐被引入人工智能、知识工程等领域，而后进入图书情报领域，用于文献资源的组织与检索。知识本体定义众多，但本质区别不大，目前公认的定义是 Studer 于 1998 年提出的：“知识本体是指概念体系的明确的、形式化、可共享的规范。”^[10]知识本体善于描述信息资源内在的知识内容，从功能上看，它与主题法、分类法相通，但比主题法、分类法更加灵活和复杂^[11]。

知识本体主要采用面向对象的分析方法，将

事物(领域知识)作为整体对象,对其属性和过程进行处理,而分类法和主题法则是将概念抽取出来进行处理。例如,在描述一台电脑时,知识本体就会生成一台电脑的对象,它包含电脑的品牌、CPU、硬盘、显示器等属性信息,还包含电脑的修理、采购等操作函数;而主题词表在描述电脑概念时,则使用电脑的品牌、CPU型号,以及用电脑加物品修理等多个概念进行组配完成;分类法在描述电脑概念时,则会按照学科将其进行归类处理。

显然,知识本体在描述领域知识时,将领域知识的一些重要属性和操作与知识本身紧密联系,因此其表示方法更加形象、准确与合理。总体来说,知识本体充分吸收了面向对象编程语言的经验,具有抽象、继承和封装的3大特征,是一种面向对象的知识表示方法,某种程度上接近于分面分类法的思想,但明显有别于主题法和分类法,因为分类法和主题法本身虽包含了概念关系,但它们的关系描述简单而松散。

知识本体虽可表达知识概念之间的各种复杂关系,解决语义异构的问题,但它无法建立知识单元与资源实体之间的直接关联,这是知识本体与元数据方法的重要区别。原因在于,元数据侧重信息资源的描述与定位,而知识本体擅长于知识内容的组织与管理,在元数据方法中设有“来源”项,可以将一条元数据描述与资源实体进行关联,而知识本体则较少考虑与资源实体的关联与揭示。例如,对于《信息检索与利用》这本书而言,元数据方法直接对其进行描述,并使用“来源”项标识这本书的空间位置,从而建立了与这本书的直接关联;知识本体则是将书中的特定知识内容抽取出来,结合信息检索这一特定领域,进行统一的形式化定义,并对它们之间的关系进行规范化描述,而对于这本书的实体特征与存在形式往往不做揭示。此外,和元数据方法一样,知识本体亦无法解决特殊性和一般性的矛盾,目前以领域本体的构建与应用为主。

1.3 知识地图

在众多知识组织方法中,除元数据和知识本

体外,还包括知识地图、主题地图、概念地图等方法,简称知识地图类组织方法。知识地图概念最早由布鲁克斯于1988年在其经典著作《情报学的基础》中提出,他认为人类的客观知识结构可绘制成以各个单元概念为节点的学科认知地图^[12]。由于概念地图和主题地图都以知识单元为节点,并以图解的方式提供知识结构的可视化表达,因此都可视作知识地图的一种,然而这三者之间存在明显的不同。

知识地图表达的知识结构范围广于概念地图和主题地图。主题地图和概念地图主要针对一定学科领域或者主题范围类的知识概念及其关系进行揭示,不能实现所有类型的知识概念描述,如揭示整个组织知识资产分布状况和结构等,而知识地图则可以。Eppler按功能将知识地图分为知识资源图、知识资产图、知识结构图、知识应用图和知识开发图5种类型^[13]。

主题地图起源于计算机科学,作为网络资源知识结构描述工具被提出,具有定位和导航某一知识概念所在资源位置的功能,ISO13250S是主题地图的国际标准^[14]。概念地图起源于教育学领域,作为教与学的工具被提出,不具备知识资源导航和定位的功能。在主题地图中,可通过“事件”来描述与主题相关的知识资源,而概念地图并不存在类似的结构来指明知识资源。有学者指出,以超文本方式存在的概念地图可作为资源导航工具^[15],但至今未见类似的概念地图应用成果报道。此外,在概念地图中,知识概念及其关系类型的设计比知识本体要灵活,没有严格定义概念之间的关系类型和约束函数,正因为如此,概念地图不具备知识推理功能。

1.4 语义网络

语义网络是万维网创始人蒂姆·伯纳斯·李(Tim Berners-Lee)于1998年提出的概念,完整的语义网络模型从下至上包括Unicode与URI层、XML Schema层、RDF Schema层、Ontology Vocabulary层、Logic层、Proof层、Trust层这7层结构^[16]。由于语义网络模型过于复杂,其应用基本局限于学术研究领域的试验性

开发^[17]，至今仍未在真实的计算机网络世界中得以普及，但语义网架构中的核心关键技术 URI 和 RDF 如同星星之火得以保存。

2006 年，蒂姆·伯纳斯·李抛开语义网络模型中的复杂成分，在 URI 和 RDF 技术基础上，再次提出“关联数据”的概念。关联数据实质上是一套应用规范，它规定在网络上发布的数据必须满足 4 个原则：(1)使用 URI 作为任何事物的标识名称，不仅是标识文档；(2)使用 Http URI，使任何人都可以参引这一全局唯一的名称；(3)当有人访问名称时，以 RDF 形式提供有用的信息；(4)尽可能提供链接，指向其他的 URI，以使人们发现更多的相关信息。

由于关联数据的框架简单，目前商业、媒体、出版、政府、图书馆等诸多领域的关联数据集发展迅猛，构建了庞大的数据网络^[18]。关联数据技术上实现虽不困难，但围绕数据集的发布、消费与应用而形成的开放应用标准、URIs 复用、RDF 的动态链接维护、RDF 关联发现等一系列重要问题，研究尚待深入^[19]。值得注意的是，关联数据仅仅是一套网络数据发布原则，类似于行动指南，对于不同应用领域，还需设计具体

数据关联模型，以支持领域关联数据集的发布。

2 知识组织方法的比较

大数据时代，文献资源组织的对象由文献单元转变成知识单元，并将采用细粒度、关联与开放的方式对其进行组织。本文主要从概念表达方式、关联资源、支持开放获取、解决语义异构等方面，详细分析各种知识组织方法的功能特点。其中，概念表达方式指的是知识单元的颗粒度，以及知识单元间关联最终产生的结果。文献本身被认为是粗粒度的知识单元，而文献中的数据、公式、事实、结论等则被认为是细粒度的知识单元。知识单元关联的结果存在线性和网络两种形式。关联资源指的是能否直接在知识单元与资源本身建立链接。无论馆藏知识网络中建立了多么丰富的关联，知识单元始终无法取代文献资源本身，因此知识单元与资源的关联必不可少。开放获取指的是能否支持资源或知识单元的网络访问。一般来说，使用 Http URI 标识，并进行了网络发布的资源，才能进行开放获取。解决语义异构指的是能否提供表达一词多义或一义多词的机制。各种知识组织方法的比较结果如表 1 所示。

表 1 知识组织方法的比较结果

知识组织方法	起源学科	概念表达方式	关联资源	解决语义异构	支持开放获取
分类法与主题法	图书情报学	线性，粗粒度	否	否	否
元数据	图书情报学	线性，粗粒度	能	否	否
知识本体	哲学、计算机科学	网络，细粒度	否	能	否
主题地图	计算机科学	网络，细粒度	能	能	否
概念地图	教育学	网络，细粒度	否	否	否
关联数据	计算机科学	网络，细粒度	能	否	能

由表 1 可知，分类法与主题法起源于图书情报学，它们以文献本身为知识单元进行描述和组织，是最基本的知识组织方法，功能简单，两者既无法直接关联资源，又无法解决语义异构问题，更无法支持资源的开放获取。

元数据亦是一种线性、粗粒度的知识组织方法，虽使资源有了基本的微观结构和语义基础，但无法解决资源描述的异构性和语义性问题。而本体模型虽可实现资源的语义化描述和细粒度组织，但在资源关联的广度和深度上未能提供更多

的帮助。元数据和本体模型，均不支持资源的开放获取。

概念地图实质是一种思维导图，概念定义极其灵活，既无法关联资源，又不能解决语义异构问题，因此不适合馆藏资源组织。

主题地图和关联数据不仅可满足资源概念的细粒度、多维网状表达，并且可关联资源本身，因此两者的组织功能较为接近，但也存在明显区别：(1)主题地图设有公共主题指示符(PSI)，可以有效解决语义异构和互操作问题^[20]，而关联数据

没有这样的机制，无法直接解决语义异构问题。(2)主题地图类是网络资源的主题索引，不支持资源的网络发布和获取，而关联数据则可通过万维网直接存取资源。值得注意的是，虽然关联数据本身无法直接解决语义异构问题，但它可通过 URI 复用或借助 RDF 框架构建领域数据关联模型来实现概念的多元化表达。

综合而言，关联数据最具时代适应性，能满足资源细粒度、关联与开放的知识组织需求，然而由于它无法直接解决语义异构问题，因此需要吸收其他知识组织方法的优点。从分类索引，到 MARC 机读目录、DC 元数据、本体模型，再到关联数据，这一系列知识组织方法都是在特定时代背景下产生的，它们之间本身就存在着一定的继承与发展关系。下文将重点阐述各种知识组织方法的继承与融合问题，尤其是关联数据与元数据、知识本体等方法的融合，从而分析关联数据的集成性和时代适应性。

3 知识组织方法的融合

分类法开启了文献资源组织研究的先河，诞生于 1876 年的《杜威十进分类法》(DDC)是世界上第一部正式出版的分类词表。我国的图书分类研究历史更加久远，可追溯到西汉时期刘向父子编制的《七略》。然而主题法研究相对较晚，世界上第一部主题词表《美国国会图书馆标题词表》诞生于 1909 年。20 世纪 60 年代中期开始，国外开始尝试分类主题一体化研究与实践，取得了诸多成果，奠定了早期知识组织方法融合的理论基础。

20 世纪 90 年代中后期，网络信息资源的蓬勃发展促进了元数据方法的发展。通过设置“主题”“分类”等相关元数据项，元数据将分类法与主题法自然地融于一体。然而元数据方法只是罗列了各种主题概念，如作者名、机构名、文献题名，无法解决其语义异构问题。图书馆界的传统作法是建设规范文档数据库，然而规范文档数据库的建设、更新与维护，工作量极其浩大，并且无法从根本上解决元数据方法的局限性。

21 世纪初期，知识本体的出现使得图书情报领域的研究重心曾集中于知识概念的表达以及语义关系的挖掘上，产生了一系列领域本体组织模型。同时，学者们发现知识本体实际上与主题法相通，纷纷致力于研究如何将叙词表转化为知识本体，或者借助各种叙词表构建领域本体，由此可见知识本体可以整合主题法。然而，无论是知识本体，还是主题法，它们都忽视了知识概念与资源本身关联的重要性，未提供资源链接。

数字网络时代，文献资源组织与服务环境发生了深刻的变化，如何融合多种知识组织方法，以构建新型资源组织框架，成为学界关注的焦点。元数据与知识本体的融合，主题地图与 RDF 框架的转换，尤其是关联数据的应用成为学界热烈讨论的课题。在众多知识组织方法中，关联数据最具代表性，可完全融合元数据与知识本体等方法，以克服自身无法解决语义异构性之不足。主要原因是：关联数据的核心技术之一是 RDF 框架，即采用“资源(Resource)- 属性(Property) - 值(Value)”三元组的形式表达资源，该框架与所应用领域无关，是通用的，其中资源被定义为任何有 URI 标识的东西，可以是一篇文档、一个独立的网页，或是一个主题词、一个作者。因此，关联数据将知识概念与资源实体完全置于统一的 RDF 框架下进行处理，属性关系不再是资源实体与知识概念，或者知识概念与知识概念的一重语义关联，而是可实现知识概念与知识概念，资源实体与资源实体，以及知识概念与资源实体的三重语义关联。换言之，若 RDF 框架的“属性”项关联的是资源实体与知识概念，那么各种元数据模型可置于关联数据框架中。将元数据框架与关联数据进行融合，只需将各个元数据项转换为 RDF 的第二组元“属性”项即可。若“属性”项关联的是知识概念与知识概念，那么各种领域本体模型则可置于关联数据框架中。本体模型与关联数据的融合，需要将知识概念与语义关系分别转换，即将知识概念转化为 RDF 的第一组元“资源”和第三组元“属性值”，而语义关系转换为 RDF 的第二组元“属性”项。因

此，关联数据天然具有集成性，可完全融合元数据和本体模型。以 DC 为代表的元数据方案与关联数据融合后，将逐渐从一个面向万维网的元数据集转变为面向语义万维网和关联数据的核心属性词汇表^[21]，而各种分类表、主题表和领域本体将成为关联数据的概念取值和属性关系的重要来源。

此外，关联数据的 URI 具有双重功能：一是资源实体的导航与定位功能；二是知识概念的唯一标识功能，通过 URI 复用可实现概念的多元化表达和规范控制。例如，若词语 A 和 B 都表达了概念 C 的含义，那么则可使用概念 C 的 URI 来标识词语 A 和 B，无需重新定义词语 A 和 B。

4 结语

现有知识组织方法，从文献分类、主题、目录、索引，到元数据、知识本体，以及关联数据等，它们在知识组织与检索功能上既有区别又有联系。本文主要从原理上介绍了各种知识组织方法，并从概念表达方式、是否关联资源、语义异构的能力和开放性等方面分析了它们的功能特点，重点探讨了关联数据与元数据、知识本体整合的必要性与可行性。本文研究结果表明，大数据时代，关联数据最具适应性和集成性，可完全融合元数据和本体模型，以满足资源细粒度、语义化和开放性的知识组织需求。

参考文献

- [1] Bliss H E, Dewey, John. The Organisation of Knowledge and the System of the Sciences [M]. New York : Henry Holt and Company, 1929 : 356-357.
- [2] Shera J H. An epistemological foundation for library science[J]. Frontiers of librarianship -syracuse university, 1968 (8) : 7.
- [3] Shera J H. Mechanization, Librarianship, and Bibliographic Enterprise[J]. Journal of documentation, 1974, 30 (2) : 153-169.
- [4] 文庭孝, 刘晓英, 刘进军. 知识关联的理论基础研究[J]. 图书馆, 2010 (4) : 9-11.
- [5] 马费成. 知识组织系统的演进与评价[J]. 知识工程,

1989 (2) : 39-43.

- [6][21]赵亮, 张春景. DC 元数据年会综述[J]. 数字图书馆论坛, 2012 (11) : 33-42.
- [7] Review of Metadata Formats [EB/OL]. [2016-04-26]. <http://www.ukoln.ac.uk/metadata/review.html>.
- [8][11]王绍平. 组织数字信息资源的元数据方法和Ontology方法[J]. 上海高校图书情报工作研究, 2005 (4) : 10-13.
- [9] 刘炜, 李大玲, 夏翠娟. 元数据与知识本体[J]. 图书馆杂志, 2004 (6) : 50-54.
- [10] Studer R, Benjamins V R, Fensel D. Knowledge engineering: principles and methods [J]. Data and Knowledge Engineering, 1998, 25 (1) : 161-197.
- [12] B C Brookers. 情报学的基础 (四) ——第四篇 情报学: 变化中的范式[J]. 王崇德, 邓亚桥, 刘继刚, 译. 情报科学, 1984, 5 (1) : 66-77.
- [13] Eppler M J. Toward a pragmatic taxonomy of knowledge maps: classification principles, sample typologies, and application examples [C]//Information Visualization, 2006. IV 2006. Tenth International Conference on. IEEE, 2006 : 195-204.
- [14] 吕元智, 王心裁, 谭必勇. 基于主题地图的电子政务信息资源组织研究[J]. 中国图书馆学报, 2007 (4) : 73-76.
- [15] 李秀芬, 路线. 概念图的研究及其应用现状[J]. 吉林工程技术师范学院学报, 2009 (7) : 65-68.
- [16] 李洁, 丁颖. 语义网关键技术概述[J]. 计算机工程与设计, 2007 (4) : 1831-1836.
- [17] 刘炜, 胡小菁, 张春景, 等. RDA 与关联数据[J]. 中国图书馆学报, 2012 (1) : 34-42.
- [18] Linking Open Data [EB/OL]. [2016-03-20]. <http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>.
- [19] Health T. Linked Data: Evolving the Web into a Global Data Space [M]. San Rafael, California: Morgan & Claypool, 2011.
- [20] 刘茜. XML 主题图和资源描述框架的比较[J]. 现代情报, 2006 (4) : 216-220.

作者简介 常娥, 东南大学图书馆副研究馆员; 夏婧, 南京邮电大学图书馆馆员。

收稿日期 2016-04-06

(责任编辑: 付伟棠)