

【数字图书馆】

图书馆外文网络资源元数据描述

王轶琚(南开大学泰达学院,天津 300457)

[摘要]本文首先论述了图书馆外文网络数据库的元数据描述对图书馆网络资源利用的重要性,进而结合语义网的技术,说明选择DC和RDF(S)两种技术来描述网络数据库资源元数据的必要性和意义,叙述了描述的过程和技术应用并对元数据之间的关系进行了简单的分析和说明。

[关键词]图书馆网络资源;元数据描述;DC;RDF(S)

[中图分类号]G250.73 [文献标识码]B [文章编号]1005-6610(2007)06-0025-03

由于网络技术的飞速发展,各高校图书馆的网络电子资源在文献总量中的比重增长得很快,较大的高校馆一般拥有几十到上百个国内外网络数据库,因此如何利用新技术的优势,来解决目前对于高校科研和教学来说都越来越重要的电子资源,尤其是外文网络数据库资源的综合利用问题,显得尤为重要和有意义。

目前,从图书馆网络电子资源描述和组织方面看,多数图书馆对于外文网络数据库资源的描述一般只在OPAC(Online Public Access Catalogs,联机公共访问目录)的MARC(Machine-Readable Catalogue,机器可读目录)数据的个别字段里著录,或是用简单的数据库或期刊导航来体现,缺乏统一完善的电子资源著录、组织体系,不利于网络资源综合利用的实现。

元数据是“关于数据的数据”,它是面向某种特定应用的用于描述资源属性的计算机可理解的信息。通过规范语法结构和语义结构,使得计算机能够无二义性地表现和获取信息。元数据可以用于网络资源的标识、集成、交换、检索等方面。Locan Dempsey认为元数据的功能主要在于几个方面:描述、定位、发现、评估、选择。^{[2][5]}因此元数据对图书馆网络资源描述来说必不可少,它是网络资源组织、检索和利用的前提和基础,具有相当重要的地位。

目前,语义网、网格等新技术的兴起,为数字图书馆提供了更好的理论和技术支持,也为新技术在数字图书馆方面的应用带来了广阔的发展空间。语义技术在数字资源、网络电子资源的组织、检索与利用方面的应用也必将受到大家的关注。

因此,本文提出用DC都柏林元数据和RDF(S)资源描述框架及模式相结合来描述图书馆外文网络数据库资源的元数据信息。一方面,可以更准确、更全面、更灵活的描述图书馆网络电子资源,形成标准、完善的资源组织体系;另一方面也是为元数据增加语义信息,是语义网的技术在图书馆网络资源的知识组织等方面的另一种应用。它便于电子资源综合利用中的索引、导航、检索的实现,同时,标准的元数据信息也便于交换和共享,符合网络发展的需要。

1 DC与RDF(S)^{[1][2][3][5][6]}

Dublin Core都柏林元数据(简称DC)是为描述网络资源、支持网络检索而建立的元数据模式。它起源于1995年在美国俄亥俄州都柏林市召开的元数据讲习班。通过国际性的合作逐步完善,DC现已在网络中得到了

广泛承认和应用。DC 的特点包括:简易性,通俗易懂;可重复性;通用性好,DC 支持对任何内容的资源进行描述,能较好的满足对网络信息描述的需要;扩展性好,引入了 DC 修饰词,便于更详细的描述资源。

当前的 DC 元数据集包括 15 个核心元素:Title、Creator、Subject and Keywords、Description、Publisher、Other Contributors、Date、Resource Type、Format、Resource Identifier、Source、Language、Relation、Coverage、Rights Management。

根据 DC1.1 版本,DC 的每个元素都根据 ISO/IEC11179 定义十个属性:Name、Identifier、Version、Registration Authority、Language、Definition、Obligation、Datatype、Maximum Occurrence、Comment。其它更多更详细的 DC 标准和规范,可参考 <http://dublincore.org/>。

RDF(S)资源描述框架及模式是 RDF 和 RDFS 的合称,位于语义网层次模型中的第三层。RDF 资源描述框架(Resource Description Framework),是一种通用的元数据结构,是描述和交换元数据的框架,是处理元数据的基础。RDF Schema(简称 RDFS)是 RDF 的补充,为 RDF 模型提供了描述领域相关的语义的机制,是 RDF 的类型系统,目的是定义被描述的资源类以及定义领域相关的属性等。RDF 通过基于 XML 语法的明确的结构化约定来帮助建立语义协定和语法编码之间的桥梁,从而促进元数据的互操作能力。更多更详尽的 RDF(S)规范,请参见 <http://www.w3.org/RDF/>。

2 DC 与 RDF(S)相结合描述图书馆外文网络资源的意义

由于 DC 具有简练、易于理解、可扩展及与其它元数据形式兼容性好等特点,是比较好的网络信息资源描述的元数据集,所以能较好地管理元数据。而 RDF(S),不仅是用来描述资源的框架、一个规范,它还能让机器理解,并且是 DCMI(Dublin Core Metadata Initiative)推荐的描述框架,符合未来语义网的发展趋势。

从图书馆外文网络数据库资源的综合利用角度看,采用 DC 和 RDF(S)结合对外文网络数据库资源进行元数据描述的意义在于:

- (1)便于实现资源的知识组织功能;
- (2)通用、标准的元数据信息可以为高效的资源索引与导航提供支持;
- (3)元数据的语义信息增强,可以为资源的检索提供一定的语义支持。

合理的、标准化的、包含一定语义关系的元数据信息,相当于为外文网络数据库资源作了一定的语义标注,有助于提高检索系统的联想能力和精确性。当然,如果全面考虑网络资源的语义检索问题,不仅需要描述这些资源的内容、版权、形式特征等方面的元数据信息作为基础,同时还需要有对抽象的概念、概念间的关系进行描述的本体信息作为基础,而本文主要就前一部分进行论述,目的是为了对网络数据库资源进行更具准确性和灵活性的、带有一定语义性的元数据描述。

3 元数据描述

由于外文网络数据库实际包含的文献类型很多,比如 Books、Journals、Magazines、Transactions(学报)、Proceedings(会议录)、Conferences、dissertations(论文)等,为了举例具有典型性,将外文网络数据库资源抽象成三级:库—刊—篇,其中,中间一级以数据库中数量最多的电子期刊为例。其它文献类型的描述过程类似,只是选取 DC 元素有所区别。

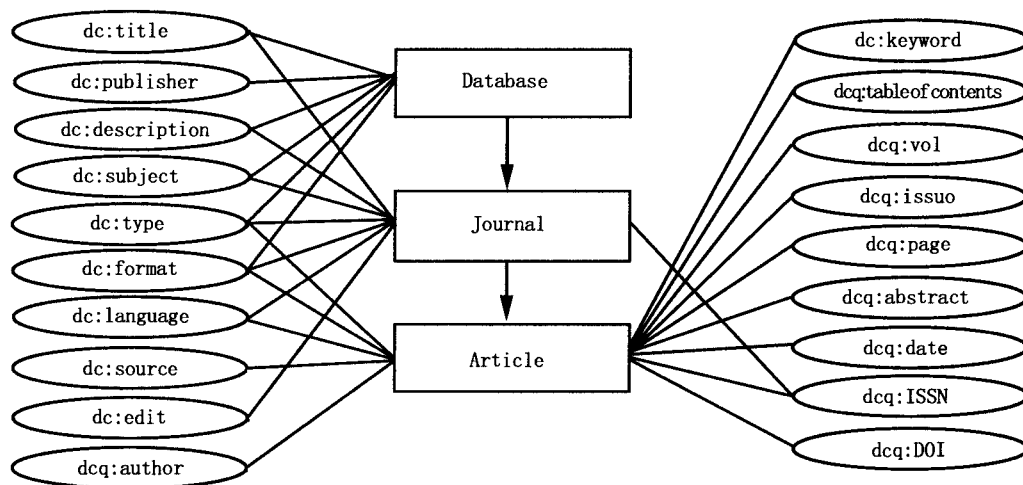
通过对 DC 元素含义的理解和参考,元数据描述之前首先应确定库刊篇三级元数据描述所需的 DC 核心元素,并针对三种对象分别补充适当的自定义元素。本文为数据库描述所选取的元素为:数据库名称、出版社(名称、地址、网址)、简要描述、文献类型、主题等。通过 DC 的 title、publisher、description、subject、format 等元素对于该网络数据库的网络资源地址、出版社信息、数据库简要介绍、数据库涉及的文献主题等内容给予描述。

为期刊所选元素:刊名、出版社(名称、地址、网址)、全文收录时间范围、文摘书目收录时间范围、出版物类型、ISSN、archive、主题、出版频率、纸本链接等。由于期刊有时会被多个数据库收录,具有两个或更多的网络地址,因此,可以使用 rdf:bag 来描述。另外期刊元数据一般还包括期刊 online archive 的时间范围或卷期范围,可以用 dc:coverage 和 dcq:DCMIPeriod 元素嵌套使用。

为期刊文章所选元素:题名、作者、刊名、table of contents、年代、卷期页码、关键词、类型、书目信息、文摘、全文等。文章元数据一般会包括一个或多个资源链接,一个或多个著者或责任者,一般还会包括来源期刊的卷期、页码、ISSN 及来源期刊同卷期的 contents 列表,此外还有著者关键词、著者机构、DOI 文献号等等,都可以使用 DC 和 RDF(S)来描述,元数据描述的 RDF 文档,由于篇幅所限,不再表述。

4 元数据之间的语义关系

从上面的元数据描述的基本过程也可以看出,由于资源的不同,元数据描述时选取了各自符合需要的元素,其中有些 DC 元素被几种资源类型都用到了,有些则是单独资源描述时用到的。本人将这种简单的关系,用下图表示:



元数据的对象—属性简单关系图

上图只是对元数据间关系的一个最简单的描述,而实际上,由于使用了 RDF(S),引入了类、属性和实例的概念,因此使元数据之间有了更丰富的语义关系。比如:有些电子期刊被不同的数据库收录,存在多个网络资源链接;通过文章的 references 链接可以将很多书、文章等的元数据都联系在一起;通过对著者、出版社、主题、来源等创建索引,可以聚集同一机构、同一著者、同一主题等的各类网络文献等等,这些都图书馆单纯用简单的 MARC 记录或是资源导航进行资源著录和组织所无法做到的。

各种元数据信息汇集在一起,形成图书馆网络资源的元数据库,而元数据库中存在的这些关系,一方面有知识聚类效果:在统一、通用的元数据描述的基础上,同一数据库、期刊、主题、著者、出版社、出版物等很容易集合在一起,有利于图书馆网络数据库资源的组织、索引和导航功能的实现,方便用户浏览。另一方面,元数据之间的各种语义关系、超链接,很容易实现元数据关系网中的多重跳转。比如:著者—作品—参考文献—其他著者,数据库—书—评论—评论人,等等。更多这样的关系,聚集关联在一起,形成了图书馆网络资源元数据库中的语义关系网,对于网络资源的语义分析和检索非常有用,便于在此基础上进一步建立效率更高的图书馆网络检索系统。

[参考文献]

- [1]<http://www.w3.org/>.
- [2]<http://dublincore.org/>.
- [3]宋 伟,张 铭. 语义网简明教程[M]. 北京:高等教育出版社,2004.
- [4]白同强,刘 磊. 语义 Web 的研究与展望[J]. 吉林大学学报(信息科学版),2004,(2):154-159.
- [5]张晓林. 元数据研究与应用[M]. 北京:北京图书馆出版社,2002.
- [6]李远景. 基于网络电子资源和数字图书馆的元数据[J]. 现代情报,2005,(6):66-67.
- [7]庞孝梅. 基于网络信息资源组织管理的和比较研究[J]. 图书馆学研究,2005,(8):43-45.
- [8]<http://dc.library.sh.cn/dcmi-terms.htm>.
- [9]<http://bbs.w3china.org/>.
- [10]柳亚琴等. RDF 在语义 Web 知识表示中的作用[J]. 太原理工大学学报,2006,(2):146-149.
- [11]郭瑞华,张玉莉. 语义 Web 上 DC 元数据的描述及抽取技术[J]. 现代情报技术,2005,(6):212-214.
- [12]陶 兰等. 面向语义 Web 的 RDF 数据处理和应用[J]. 深圳大学学报(理工版),2005,(4):330-333.
- [13]李 玲,唐胜群. 知识网格中基于 RDF 的知识表示技术和应用[J]. 计算机应用研究,2005,(12):223-225,229.
- [14]鲁 舟,朱国进. 基于 RDF 的个性化资源集成模型的研究与应用[J]. 计算机应用研究,2005,(10):69-73.

[作者简介]王轶璐(1975—),女,馆员,本科学历,曾发表论文数篇。