

●沈芸芸

Metadata 与网络信息资源的书目控制

摘要 对网络信息资源进行书目控制是信息时代书目控制的一个新内容。Metadata 就是在将因特网视为一个巨大的图书馆的理念下产生和发展起来的。它的主要功能是描述、组织和发现因特网资源,进而对网络资源进行有效控制。参考文献 8。

关键词 网络信息资源 书目控制 元数据

分类号 G250.73

ABSTRACT Bibliographical control of network information resources is a new content of bibliographical control in the Information Age. Metadata originates from the idea regarding the Internet as a huge library. Its main function is to describe, organize, find and then control Internet resources. 8 refs.

KEY WORDS Network information resources. Bibliographical control. Metadata.

CLASS NUMBER G250.73

1 网络信息资源书目控制的内涵

谈到书目控制,通常可以理解为利用书目进行控制,其目的是为了准确地描述和取出文献,以满足特定的检索需求。它是基于用户需求与文献资源激增的矛盾以及任何图书馆都不能收全所有出版物这一理论而产生和发展起来的。这里,书目控制的对象指的是各类型的文献资源。随着时代的发展,文献资源的类型也在不断扩充,包括印刷型文献、音像(视听)资料、缩微资料、机读文献等等。

随着信息化社会的到来,现代人如果不懂网络就会被称为文盲。事实上,吸引人们的不仅是网络本身,更大程度上是由于它所蕴含的极其丰富的信息资源。“网络信息资源”作为一个新的专业术语,也可以说是一种新的信息资源类型,一般可以理解为通过国际互联网可以利用的各种信息资源。网上信息资源主要可以分为两大类:一类是相对流动的、即时性的和变化的信息,一般不具有控制的必要;一类是“稳定性的信息资源”,是由一定的机构运营、并经过一定加工和处理的比较稳定的信息,如网络电子期刊、高等院校和研究机构的网页、知名公司的网页等等,这部分信息资源具有较大的实际价值,我们应该控制的也恰恰就是这部分信息。

在因特网上存在着大量可获得的有用信息。为了检索的便易,这些资源需要加以组织。利用一定的数据格式将这些资源组织起来是获得它们的最有效

的方式。

但是,由于网络资源具有数量大、来源和传播方式多样等特点,人们利用网上资源信息的困难日渐显露。其结果必然会导致人们对书目工作的研究。对网络信息资源进行书目控制,成为网络信息时代书目控制的一个新内容。

同图书馆对印刷型文献进行书目控制相比,网络信息资源的书目控制具有新的内涵。首先,它所控制的是网络信息资源,要对网络信息资源加以描述;其次,网络信息资源比印刷型文献资源更加复杂多样,不但具有广泛性、无序性、分散性,而且具有多媒体特征,即集图文、声像、视听等为一体,所以对它进行书目控制更加困难;第三,网络信息的发展以现代信息技术的发展为依托,具有更新快、传播广泛、不易控制的特点,它与用户利用之间的矛盾更加尖锐;第四,网络资源书目控制对于现代信息技术具有较大的依赖性,不仅在控制对象上体现出网络信息技术的特征,而且在控制方法上更需要信息技术的支持。

由此可见,网络信息资源书目控制就是要解决海量的网上信息与用户利用之间的矛盾,具体实践中,就是对网上资源加以描述分类,使大量随机无序的网上信息转变为有序的、可以高速传递的信息,从而在用户与网上信息之间架起一座桥梁。由于人们对网络信息资源进行控制以方便检索的要求越来越迫切,世界各国从事信息工作的专家对如何进行网络信息资源的书目控制做了许多有益的探讨,Metadata 就是

在这样的背景下产生和发展的。

2 Metadata 及其功能

2.1 Metadata 诠释

可以这样比喻,整个因特网就像一个巨大的图书馆,各个网站就如种类不同的图书,而一个网页就可比作图书中的一页,这样,就可以利用图书馆的组织理念来思考网络资源的组织与控制。Metadata 就是在这样的理念下发展起来的,它是关于数据的数据(data about data),一般将它译为元数据。它作为一个术语,是指用于帮助识别、描述和定位网络化的电子资源的任何数据。

尽管 Metadata 是网络环境下产生的新名词,但是它的理念却不是新的。图书情报界于 70 年代提出的 AACR2(英美编目条例 2)就是体系最庞大、定义最详细的元数据,只是它并不叫元数据罢了。我们可以用图书馆的书目(卡片或机读格式)对 Metadata 进行说明。书目记录了某一种图书的题名、著者、出版事项、页码、ISBN 号、分类、主题以及该书在图书馆书架中的位置等信息,它不仅描述了图书的外部特征,而且揭示了图书的内容特征,使人们能够在看到该书之前就有一个概要的了解。Metadata 的功能与书目卡片的功能大体相似,通过 Metadata,我们能够了解到某个网站的资源类型、Web 页的标题、责任者、主题或关键词以及内容摘要等信息。但是 Metadata 的元素集合远比 AACR2 要简单明了的多,只将一些最重要的信息点揭示出来,而且其规则目前来说并不繁琐,一般用户也可以理解和掌握。

2.2 Metadata 的功能

Metadata 的基本功能就是描述、组织和发现因特网资源,进而对网络资源进行有效的控制。它的更具体的用途如下:

- (1) 描述网络数据的内容。
- (2) 使网络中的数据便于搜索。
- (3) 帮助用户决定某些数据是否为其所需。
- (4) 防止一些用户(如孩子)存取某些数据。
- (5) 让用户可以重新得到或使用数据的另一拷贝,例如,指出某篇 Web 页在的 URL,使用户知道可以到哪里获得这一数据。

(6) 指导怎样“读懂”数据,如说明数据的格式是什么,采用什么样的编码方式,怎样进行解密。

(7) 帮助决定利用什么样的数据实例(如果提供多种格式的话)。

(8) 给出影响数据使用的一些信息(如数据使用的法律条文,数据使用的范围,数据使用的时限等)。

(9) 给出数据的历史过程,如数据的最初来源,以后的演变历程。

(10) 给出数据与其他资源的关系,如与前后版本数据的链接,使用本数据所需要的其他数据和程序。

(11) 对数据管理的控制。

(12) 对某些缺少文本的数据(如图像、声音等)进行文字说明。

2.3 Dublin Core Metadata 规范

Metadata 规范的产生是根植于因特网环境中的,目前已有许多种不同的 Metadata 规范,如 Dublin Core, PICS, Web Collections, CDF, MCF 和 RDF。除此之外,还有许多应用于各学科领域的 Metadata 规范。目前最受关注且应用比较广泛的要数 Dublin Core。

“Dublin Core Metadata 元素集合:参考性描述”于 1996 年 12 月最终定型后至今,它的应用只是实验性的,可能会根据实践而改变对它的解释。需要注意的是,元素具有描述性的名称,它们是要传达共同的元素语义理解。为了促进全球互操作性,对于各个元素的值,大量元素的描述可能与受控词表相关。假定在某一局域内其他受控词表会得到开发以达到互操作性。在下面的元素描述中,规定一个正式的单字词标识(用大写字母强调)来设计元素简写码的语法规则,以便将此计划代码化。每一元素都是可任选和可重复的。以下就是其元素描述:

(1) 题名(标识:TITLE)——由创建者或出版者赋予资源的名称。

(2) 著者或创建者(标识:CREATE)——对创建资源的智力内容负主要责任的个人或团体。例如,对于书面文件它是著者,而对于视觉资源则是艺术创作者、摄影者或插图者。

(3) 主题词和关键词(标识:SUBJECT)——即资源的主题。通常,主题词会被描述为关键词或词组,它们描述资源的主题或内容。这里鼓励使用受控词表和正式分类法,如 DDC, MeSH, LCSH 等,以增加描述的精确性。

(4) 资源描述(标识:DESCRIPTION)——即资源内容的文本描述,包括文件型对象的文摘或视觉资源的内容描述。

(5) 出版者(标识:PUBLISHER)——是指以目前形式制作的资源负责的实体,如出版社、大学的系所或法人实体。

(6) 其他对该资源作出贡献者 (标识: CONTRIBUTOR) ——指未在“创建者”元素中规定的个人或团体,他们对资源作出重要的智力贡献,但对“创建者”元素中规定的个人或团体来说其贡献是次要的(如编者、转录者或插图者)。

(7) 日期 (标识: DATE) ——以目前形式制作资源的日期。所建议的最好做法是“年-月-日”(××××-××-××)形式,即 8 位数字,4 位年,2 位月,2 位日,正如 <http://www.w3.org/R/NOTE-datetime> (ISO 8601 资料文档)中定义的那样。在该方案中,日期元素 1994-11-05 对应于 1994 年 11 月 5 日。也可以使用许多其他的方式,但如果使用了其他方式,就应该用正确的方法识别它们。

(8) 资源类型 (标识: TYPE) ——指资源的类别,如主页、小说、诗歌、工作文件、技术报告、散文及字典。资源类型应在 Dublin Core 专家组控制之下的列举表中作选择。当前有关这一元素应用的思考,可参阅 <http://sunsite.berkeley.edu/Metadata/types.html>。

(9) 资源格式 (标识: TYPE) ——即资源的数据格式,用于识别软件;可能的话,还可识别用以显示或管理资源的硬件。

(10) 资源标识 (标识: IDENTIFIER) ——用于唯一标识资源的字符串或数字。对网络化资源来说,例子包括 URLs 和 URNs (在实现时)。其他全球唯一标识,如 ISBN 或其他正式各称,也可作为脱机资源下的候选元素。

(11) 来源 (标识: SOURCE) ——如果可用的话,该元素是一个字符串或数字,用来唯一标识该资源是从哪一作品中得来的。例如,一个 PDF 版本的小说可能具有“来源”元素,包含了得到 PDF 版本的图书的 ISBN 号。

(12) 语言 (标识: LANGUAGE) ——资源知识内容所使用的语言。在实践方面,这个字段的内容应与 RFC1766 (即 Request for Comments 1766) 保持一致。见 <http://ds.internic.net/rfc/rfc1766.txt>。

(13) 关系 (标识: RELATION) ——指该资源与其他资源的关系。该元素的目的是提供表述资源之间关系的一种手段,这些资源与其他资源有正式关系但都作为各自独立的资源而存在。例如,文档中的图像、图书中的章节或馆藏中的文献条目。关于“关系”的正式规定当前正在发展中,该元素的使用目前尚属实验性的。

(14) 覆盖范围 (标识: COVERAGE) ——资源的

空间和时间特征。关于“覆盖范围”的正式规定当前正在发展中,该元素的使用目前尚属实验性的。

(15) 版权管理 (标识: RIGHTS) ——它是对版权通告、版权管理声明的一个链接,或是对一项服务的链接。该服务提供能够检索到某资源的检索词方面的信息。关于“版权管理”的正式规定当前正在发展中,该元素的使用目前尚属实验性的。

这个元素集是可扩展的,其中的每一个元素都是可重复使用或有选择地使用。在使用时,主要利用 HTML 语言格式进行描述。目前因特网上的网页,大都带有符合 Dublin Core 规范的 Metadata。与其他 Metadata 相比,Dublin Core 简单易懂,可灵活运用于各个领域,并基本取得了国际一致,得到包括美、英、德、日、加等国家的积极参与和支持。

2.4 其他 Metadata 规范

PICS 即因特网内容挑选平台 (The Platform for Internet Content Selection),是由 W3C (WWW Consortium) 组织开发的标准。它的显著特征是引入了内容分类定级机制,以此对网上资源进行分类定级,过滤不符合它们规范的内容。

Web Collections 是伴随着 XML (eXtensible Markup Language) 应用环境的日渐成熟而发展起来的,采用与 HTML 语言相似的样式风格,引入了一种表示 Metadata 的层次结构的方法,对 Web 页进行描述,因此它的数据可以紧紧嵌入到 HTML 内容当中,从而将许多文件组合在一起进行深层次的结构化内容描述。

CDF 即频道定义格式 (Channel Definition Format),它是微软提出的一项 Metadata 规范。它借助于 HTML 语法来对其频道内容进行描述,当浏览器接入这个频道后,这些 Metadata 数据就会展现出来。

MCF 即 meta 内容框架 (Meta Content Framework),它是网景提出的一个利用 XML 的 Metadata 方案。在 MCF 中,对 data 和 Metadata 并不做明显地区分,在 MCF 组织看来,人们以同样的方式利用数据 (包括 data 和 Metadata)。它采用对象、属性类型、属性值的形式描述 Metadata 数据,并允许它们的重复使用,从而使它能描述复杂的文件关系。

RDF 即资源描述框架 (The Resource Description Framework),它是一批专家试图解决许多 Metadata 数据的互操作性——也就是它们的兼容性而制定出来的。它符合多种需要,健壮而且灵活。它提供的框架,包括范围之广,足以使它能容纳得下所有领域的

应用。它是如今被十分看好的一种 Metadata 规范。

由此可见, Metadata 注重对网络信息资源的版权以及外形特征和学科主题实施有效的控制, 这与书目控制的初衷——即以检索出所需文献为目的而对文献信息实施描述、揭示、分析、选择、评论、综合——不谋而合。Metadata 格式实际上可以看作书目的一种变体, 它与书目的功能又极其相似, 所以说, Metadata 与书目控制之间的关系是密不可分的。

3 Metadata 对网络信息资源书目控制的意义

Metadata 是网络环境下的产物, 尽管对它的研究以及它的应用仍处于实验性的阶段, 但通过分析可以看出, 它离不开书目控制论的指导, 同时又丰富和发展了书目控制理论的内容、手段和方法。所以说, Metadata 对于研究书目控制理论与实践在网络化环境下所发生的变化与发展具有深远的意义。

(1) Metadata 在一定程度上解决了网络上海量可用资源与用户需求之间的矛盾。由于 Metadata 数据本身较为简单, 一般用户无须培训就能掌握它的格式并能正确使用。它所具有的强大的组织和发现网上信息的功能, 使用户能够极其方便地利用有用的网上信息资源, 不致在网上迷路而无谓地浪费宝贵时间, 而且能够防止各种垃圾信息和不良信息的过多传播。

(2) 提高了网络资源的利用率。在网络信息的浩瀚海洋中, 某些有用的特定信息很容易就被淹没其中, 但通过对它们实施控制, 揭示出它们的一些重要特征后, 用户在需要时就能方便迅速地找到这些信息, 从而大大提高网络信息的利用效率。

(3) 通过 Metadata 可以掌握网络信息资源在地域上和学科上的分布状况, 从而做好网络资源的建设工作, 使其分布更合理。

(4) Metadata 还为图书馆的各类型馆藏资源的数字化提供了很好的组织描述工具, 使我们在组织信息上网的同时就对它们进行有效控制。例如将古籍拓片上网, 我们就可以参照其他 Metadata 制定出一系列标准的描述数据。

此外, 由于 Metadata 还未形成国际通用的规范标准, 许多关于数据细节的各种规定都处于发展过程中, 例如著者或创建者的规范控制、主题词的使用、HTML 语言的代码化等。因此, 加快 Metadata 数据的规范化和标准化工作, 以便有效地对网络资源进行书目控制, 这是信息时代最为紧迫的任务之一。

参考文献

- 1 Weibel, Stuart. The State of the Dublin Core Metadata initiative April 1999. D-lib Magazine, April 1999. <http://www.dlib/apri199/04weibel>
- 2 Dublin Core Metadata Element Set: Reference Description. <http://purl.org/dc/July2,1999>
- 3 <http://mirrored.ukoln.ac.uk/about/elementset.htm>
- 4 黄纯元. 图书馆与网络信息资源. 中国图书馆学报, 1997, 23(6): 13 ~ 19
- 5 张智雄. Metadata: 组织和发现 Internet 信息资源的数据. 现代图书情报技术, 1999(1): 21 ~ 25
- 6 索传军. 论网络化图书馆的信息资源建设. 图书馆, 1999(1): 22 ~ 25
- 7 黄俊贵. 书目控制的思辨. 中国图书馆学报, 1995, 21(5): 56 ~ 63
- 8 王岩. 书目控制的含义及实用性研究. 中国图书馆学报, 1992, 18(4): 69 ~ 71

沈芸芸 北京大学图书馆工作。通讯地址: 北京市中关村。邮编 100871。

(来稿时间: 1999-10-09)

“世纪之初的图书馆”学术研讨会召开

由中国图书馆学报编辑部和重庆图书馆联合举办的“世纪之初的图书馆”学术研讨会, 于 2000 年 9 月 21 日至 24 日在重庆市召开。会议得到重庆市文化局的大力支持。会议评出优秀论文 13 篇, 三等奖论文 38 篇, 二等奖论文 26 篇, 一等奖论文 7 篇, 合计 84 篇。论文作者 102 人, 分布在全国 24 个省、自治区、直辖市及科学研究、军队、党校、教育系统和企业。会议向论文作者颁发了获奖证书。与会作者们认为, 通过学术交流, 活跃了学术气氛, 开阔了视野, 提高了研究能力; 通过对论文格式规范的学习, 对今后写作论文大有裨益。