



DOI及DOI检索

An Elementary Introduction to DOI and DOI Search

曹红兵 (广西大学图书馆 广西 南宁 530004)

[摘要] 本文介绍了一种新的网络资源标识系统DOI及其在网络信息检索中的作用,并介绍了网上数据库检索系统SpringerLink中的DOI检索。

[关键词] DOI 数字对象标识符 电子出版物 检索

[中图分类号] G252.7 **[文献标识码]** B

[Abstract] DOI, Digital Object Identifier, which is a new kind of identification system in the environment of network, and DOI search in the Springer Link database online has been introduced in this paper.

[Key words] DOI; Digital object identifier; Electronic publications; Search

因特网的盛行推动了电子出版物的蓬勃发展,这在给电子出版商及信息用户带来益处的同时,也带来了问题。对出版商而言,一方面,电子出版物的方便快捷、易于使用为他们赢得了大量的用户和市场;另一方面,电子出版物的版权难以保护又使他们蒙受了重大的经济损失。对信息用户而言,他们一方面为网络资源的无比丰富而欣喜;另一方面又为网络资源的变幻莫测而烦恼。据报道,仅仅半年时间内,因特网上因URL资源地址变动而造成的“断链”、“死链”、“坏链”数就占到了总链接数的1/6。由此而形成的网上信息经常“NOT FOUND”的现象令网络检索用户们抱怨不已。而这些情况很大一部分都是由于现行的网络资源标识系统URL的不稳定和易变造成的。因此,推出并实行一种稳定可靠的、既能保护知识产权又能实现版权所有者商业利益的数字对象标识系统的想法应运而生,DOI由此而来。

1 DOI的定义及特征

DOI,即Digital Object Identifier(数字对象标识符)的缩写。广义地说,DOI是指在数字环境(即网络环境)下标识和交换知识产权对象的一种开放性系统,它遵从URI(统一资源标识符)规范,并为基于数字对象结构公认标准的知识内容管理和知识产权管理提供了一个可扩展的框架。它包括4个组成要素:标识符、解析系统、元数据和规则(Policies)。狭义地说,DOI是指标识任何数字化对象的一种标识符号。

DOI具有以下显著特征:

(1) 唯一性(Unique)。DOI标识符能唯一地标识网络环境下的任何数字化对象,如文本、图片、声音、影像、软件等。

(2) 持久性(Persistent)。DOI标识符一经产生和分配就永久不变,不随其所标识的数字化对象属性(如版权所有者、存储地址等)的改变而改变。

(3) 可驱动性(Actionable)。DOI标识符类似于面向对象模型中的“事件”(Event),能用来驱动某个动作。基于DOI驱动的最简单的动作就是为它所标识的实体(即数字化对象)定位(Locating)。

(4) 兼容性(Interoperable)。DOI系统具有与过去、现在及未来系统的兼容性,具体包括:① DOI标识符的兼容性。DOI标识符中可包含任何已有的标识符,如ISBN、ISTC(国际标准文本代码)、PII(出版者款目标识符)等;② DOI处理系统的兼容性。DOI处理系统具有高效性和可扩展性,能与任何现有或将有的因特网系统进行互操作;③ DOI元数据的兼容性。DOI元数据能够最大限度地实现不同系统间的数据互用性。

2 DOI标识符的构成

DOI标识符是一组由字母、数字或其他符号组成的字符串,包括前缀、后缀两个部分,中间用一个“/”分隔。前缀由DOI管理机构(目前只有一个,即CrossRef)分配给各出版商;后缀由各出版商自行指定并分配给其数字化出版物。前缀是一个由“10”开头的6位数的代码(必要的话,前缀也可以加长),如“10.1006”;后缀则是一组由出版商指定的任意形式的代码,如“jmbi.1995.0238”。一个完整的DOI标识符的形式是“10.1006/jmbi.1995.0238”。

DOI标识符不区分大小写,因此,“10.1006/abc”和“10.1006/ABC”没有分别。

原则上,每个出版商只分配有一个前缀,如分配给德国

Springer-Verlag 科技出版集团的 DOI 前缀是“10.1007”。但必要时也可为一个出版商分配多个前缀,如可为它的每种数字化出版物各分配一个前缀。由于 DOI 前缀一经分配就永久不变,而出版物的版权所有者却有可能变更,因此,单凭 DOI 前缀来识别出版商并不总是可靠的。

与前缀相比,DOI 后缀的语法十分灵活,CrossRef 对此不做具体规定,而只给出几条总的指导原则,如简洁性原则、惟一性原则(在同一个前缀下的后缀必须惟一)、一贯性原则、逻辑性原则、兼容性原则、层次性原则、扩展性原则等。在此原则指导下,各出版商可自行指定其 DOI 后缀的语法规则,如:它可以是任何字母数字字符串;该字符串可以分节也可以不分节;若分节则必须要用分节符分隔;分节符可以是圆点、冒号、管道符或其他符号等等。

分节表示的 DOI 后缀常可以反映出信息的层次性,具有结构清晰、明白易懂的优点。因此,许多出版商都采用了这种 DOI 后缀表示方式。现举两例如下:

出版商	DOI 后缀语法	DOI 标识符示例
美国学术出版社	/ 4 位期刊代码.文章录用年份.文章序列号	10.1006 / jmbi.1998.2354
美国物理学会	/ 刊名缩写.卷数.文章代码	10.1103 / PhysRevLett.88.088302

许多出版商还采用已有的标识符来作后缀,如美国数学学会采用了 PII (出版者款目标识符)作为其 DOI 后缀,依此生成的 DOI 标识符形如“10.1090 / S0002-9939-00-05422-8”。

3 DOI 的注册、使用及在网络信息检索中的作用

DOI 必须注册后才能使用。DOI 注册机构受 DOI 国际基金会 (IDF) 的委托,受理 DOI 的注册事务,并提供已注册 DOI 信息的存储、解析和检索服务。出版商负责 DOI 的注册,并负责更新和维护该注册信息的可靠性。DOI 连同它所标识的数字化对象的 URL 和其他书目信息 (DOI 元数据),以及与该数字化对象有关的所有 URL 链接信息被出版商送往 DOI 注册机构进行注册,并存储在那里的 DOI 中心服务器。这些集中存储的数据最终形成一个 DOI 解析数据库,利用特殊的软件,这一数据库可以链接或解析某一个 DOI,使其与同它有关的所有数字化对象的 URL 联系起来。当用户欲检索某一个数字化对象,或有关这一对象的信息时,就会发出如下形式的 DOI 请求:“http://dx.doi.org/10.1007/s00412005078”,该 DOI 请求会被送到 DOI 服务器,由该服务器作出响应,并将所找到的 DOI 记录连同有关的对象的地址 (大部分是 URL) 返回给用户的浏览器,由浏览器从中进行检索,再将检索结果显示给用户。

传统的网络检索是以 URL 作为检出资源的标识符,而 URL 变化频繁,常令所检出的资源因资源地址变动而出现

“NOT FOUND”的局面。这种以 URL 为标识的检索结果的不可靠性使得传统的网络信息检索带有很大的缺陷;加之随着网上资源的增多,用户已不再满足于仅仅检出原始文献正文,而是希望通过原始文献正文能够再比较方便地查找到其引用的文献,以探寻知识的起源和学术的关联。于是,DOI 检索应运而生。由于 DOI 的惟一性和持久性,使得 DOI 具有追踪文献当前地址和参考文献链接的能力。因此,利用 DOI 进行检索,可以保证即使所链接的资源存储地址或版权所有者发生变化,检索者仍能找到所需的信息;而且,利用 DOI 检索,检索者不但可得到该 DOI 所标识的对象的完整的书目信息和文献正文,还能得到该对象所引用文献的书目及正文信息,且不论这引用文献与被引文献是否在同一个数据库中或为同一个出版商所有。因此,DOI 检索在网络信息检索、尤其是在参考文献链接检索方面具有很大的优势。

CrossRef 作为经 IDF 授权的,针对学术期刊、会议录和图书方面目前惟一的 DOI 注册机构,它所提供的 DOI 检索服务使其成为当前网上最大的开放式参考文献链接服务系统。目前,CrossRef 的 DOI 中心目录中存储有来自 6 000 种期刊、超过 450 万篇文献的 DOI 记录,并预计以每年 100 万篇左右文献的速度增长。

4 SpringerLink 中的 DOI 检索

作为最早参加 CrossRef 的出版商之一,Springer-Verlag 在其网上数据库检索系统 SpringerLink 中为其所有以“Online First”方式出版的网上出版物提供了 DOI 方式的检索。在 SpringerLink DOI 检索中,用户只要以 DOI 作为检索入口词,如在检索词输入框中输入“10.1007 / s00214990 m180”,就可以得到该 DOI 所标识资源的完整的书目信息 (文摘、题录) 和文献全文,以及所有与该资源有关的资源的 URL 链接。这些 URL 链接所指向的资源既可能位于 SpringerLink 系统内,也可能位于 SpringerLink 系统外 (Off-Site)、属于 CrossRef 其他成员数据库中的记录。SpringerLink DOI 检索得出的书目信息 (即题录、摘要) 对所有用户是免费提供的,而全文信息则只有授权用户或通过单独付费方式才能得到。

参考文献:

- 1 Dr.Norman Paskin.The DOI Handbook.http://www.doi.org/handbook_2000/index.html, 2003.6
- 2 DOI Info:&Guidelines http://www.crossref.org/01company/15doi_info.html, 2003.6
- 3 http://link.springer.de/search.htm, 2003.6

[作者简介]

曹红兵 女,1968 年生,华东师范大学图书情报专业本科,副研究馆员,已发表论文 8 篇,现在广西大学图书馆工作。

[收稿日期:2003-07-15]