

面向文献提供服务的版权管理解决方案*

□ 王莉 / 中国科学技术信息研究所 北京 100038

摘要: 面向数字图书馆的文献提供服务这一特定应用场景, 文章提出基于Handle架构的D-DOIs解决方案, 通过对文献传递行为的唯一标识, 结合数字水印和二维码技术, 实现对数字对象传播的持续追踪。

关键词: 数字版权管理, D-DOIs, 数字对象唯一标识符, Handle系统, 数字水印, 二维码

DOI: 10.3772/j.issn.1673—2286.2013.08.005

1 引言

文献提供是图书馆的核心业务, 由于图书馆的公益性质, 各国版权法都针对图书馆制定了专门的合理使用条例, 也称为“版权例外(Copyright exception)”, 支持以个人学习、研究为目的的少量复制行为。当图书馆还处于手工借阅和纸本流通时期, 图书馆的复制行为还不足以引起著作权人的重视, 但是, 随着互联网的飞速发展, 传统图书馆进入数字图书馆时代, 这种平衡被逐渐打破, 由于电子文献更容易复制和扩散, 著作权人的合法权益更容易受到损害^[1]。越来越多著作权人和出版商呼吁缩小合理使用的范围, 一些机构(如AAAS)甚至已经禁止图书馆利用电子文献资源进行文献传递^[2]。

在网络环境下, “合理使用(Fair use)”仍然是平衡著作权人的利益和公众获取利用信息权的重要手段。对图书馆而言, 严格遵守“合理使用”制度, 除了文献提供服务声明、文献提供数量限制、缓存时间限制外, 一项非常重要的手段是对传递行为的追踪。本文提出基于Handle系统构建面向文献提供的版权管理系统, 实现对传递中的

电子文献资源的版权保护, 促进图书馆正常业务的健康、良性、持续发展。

2 国内外相关研究

2.1 数字版权管理

数字版权管理(DRM, Digital Right Management)是指数字内容在生产、传播、销售和使用过程中对版权进行定义、描述、保护和监控的整体机制, 其核心是通过安全和加密等技术手段锁定数字内容, 控制其分发与使用。

目前国内外开展了很多对数字版权管理的研究, 各种类型的解决方案与应用也不断丰富^[3-5]。按照数字内容的类型, 结合不同的应用目的, 可以将这些研究与实践大致划分为三大类: 第一类是针对流媒体的数字版权管理。流媒体即音频和视频, 涉及音乐、电影等广阔的消费市场, 是研究投入最多、成果产出最多的领域, 代表性的产品有苹果的FairPlay系统、微软的Windows Media DRM和IBM的EMMS数字版权保护方案。第二类是针对企业重要文档的版权管理。严格地说, 这个发展方向已经远离

了“版权”的含义, 而是基于DRM框架构建文档安全解决方案, 强调保护企业内部文档的机密性和完整性, 例如微软Windows RM(Rights Management)解决方案和为Secospace文档安全管理系统(DSM)。第三类是对电子书、电子期刊等公开数字出版物的版权管理, 强调网络环境下的访问控制和追踪。代表性的产品有Adobe的Content Server、微软的电子图书DRM系统、方正的Apabi Right Server以及书生的SureDRM等。数字图书馆中的版权管理属于第三类研究领域。

目前DRM系统在数字图书馆中典型的应用是与电子阅读系统相融合的借阅服务模式, 在国内比较成功的应用是方正Apabi数字图书馆整体解决方案^[6,7]。简单地说, Apabi采用复本方式发行电子图书, 当图书馆将一本电子图书通过网络“借阅”给一位读者时, 该读者获得阅读许可, 这时, 该电子图书不能再被其他读者使用^[8]; 同时, 系统会设置借阅时限, 当读者“归还”电子图书, 或者有效时间已到, 读者的使用许可将自动失效, 而图书馆重新拥有该复本的许可, 可以提供给下一位读者。这种模式并不适

*本文为中央级公益性科研院所基本科研业务费专项资金“基于数字对象唯一标识符的知识管理建设与研究”的成果之一, 获得国家科技支撑计划“技术创新服务平台关键技术研究与应用示范”项目的资助(编号: 2011BAH30B01)。

合网络环境下的文献提供服务,其核心原因在于:文献提供服务是由传统复制行为衍生而来的,版权管理关注的焦点是少量提供与以个人学习、研究为目的的利用,需要解决的关键问题是版权的描述与跟踪;而借阅模式下并不关注读者的使用目的,强调的是对使用权限的严格控制。

2.2 数字对象唯一标识

唯一标识符是针对某个对象的持久名称,国外的数字文献生产商较早采用它来标识其出版的电子文献,并形成了很多标识符方案。但大多数标识符方案仅仅定义了标识符的名称空间及构成机制,尚未构成一个包含解析的完整的标识符系统,如连续出版物及其单篇文献的标识SICI (Serial Item and Contribution Identifier)、图书及其内容片断(章节、前言、索引、段落)的唯一标识符BICI (Book Item and Contribution Identifier)以及出版物件标识符PII (Publisher Item Identifier)等。相对于这些唯一标识符方案,美国出版协会建立的DOI系统则比较成熟,它并不是简单提出的一种命名机制,而是包括名称空间、唯一标识符、命名机构、命名登记系统和解析系统5个部分的一整套体系,在数字出版领域得到广泛应用,国外已经有Elsevier、Springer、Blackwell、John Wiley等上千家出版商使用DOI作为数字资源的唯一标识符,注册的DOI数量超过千万条。随后一些相关增值服务也相继出现,例如“Simple Text Query”的DOI反查服务,CrossRef采用DOI实现的参考文献链接服务

等。除美国外,日本cIDf和欧洲的MEDRA也在各自地区内开展了类似的研究和服务^[9]。

我国对唯一标识符进行系统研究开始于2002年的科技部项目“我国数字图书馆标准与规范建设”子项目“数字资源唯一标识符应用规范”。该项目提出了我国数字对象唯一标识符的应用体系建设和发展方案,认为我国应遵循开放的Handle System协议和规范,建立中国本土的唯一标识符系统。中国新闻出版总署也从国内出版界的角度对正式出版物的唯一标识符进行了研讨。中国版权中心从版权保护的角度出发,探讨基于DOI的正式出版数字作品权益归属的明确描述问题。虽然相关各界对建立中国本土唯一标识符系统的呼声很高,但是目前,属于我国自己的中文数字对象唯一标识符的标准规范体系并没有建立起来^[10,11]。

2007年3月,中国科学技术信息研究所和万方数据股份有限公司成为IDF的DOI注册机构,负责中文DOI的运营,旨在通过与国内外相关机构的合作,在中文数字资源领域推广DOI标准,从而使读者及其他用户能够可靠地访问数字资源。然而,从使用角度看,中文DOI的应用并没有真正解决全文链接问题,往往是从一个题录页面跳转到另一个题录页面,无法真正满足用户的信息需求。

在数字图书馆领域,DOI标识符的应用主要集中在参考文献链接和开放链接服务两个方面。前者是通过加入CrossRef联盟获得的增值服务;后者则是通过DOI标识与OpenURL技术的结合,构建跨资源类型、跨资源平台的开放式链接服

务体系。笔者认为,图书馆并不是DOI标识符的注册者,而是使用者;除了合理利用DOI标识符本身外,应该积极探索利用DOI系统成熟的命名-解析框架管理数字图书馆范围内的各种复杂数字对象。

3 面向文献提供的版权管理解决方案

图书馆的文献提供服务是由其传统的复制行为衍生而来的,版权管理关注的焦点是少量提供与以个人学习、研究为目的的利用,需要解决的核心问题是版权的描述与跟踪。本文以Handle系统为底层解析框架,将DOI技术引入DRM,实现对传递中的电子文献资源的版权管理(以下简称为D-DOIs^①)。其基本思路是:遵循DOI编码标准定义D-DOI名称空间;在文献提供服务时,为当前传递行为分配一个唯一标识符D-DOI (Delivery DOI),相关元数据不仅要描述文献本身的信息,还包括读者和服务提供者的描述信息,以及发生传递行为的时间;采用数字水印和二维码技术将D-DOI等信息嵌套进电子文献中,这些信息将跟随该电子文献的整个生命周期。D-DOIs提供唯一标识符的解析功能,当未来发生版权纠纷时,能够从原始文档中提取相关信息,追查传播源头,判断是否存在超越合理条款非法复制和传播行为。

3.1 名称空间

D-DOI名称空间由前缀和后缀两部分组成,两者间用半角“/”分隔。前缀由若干级命名授权段构成,为注册图书馆代码。各级采用6

^①文中用D-DOIs表示整个版权管理解决方案,D-DOI则是解决方案中使用的唯一标识符。

位数字编码表示,用半角“.”分隔。后缀是在注册图书馆本地域下的唯一标识码,没有固定的格式,由图书馆内部自定义,只要保证内部唯一即可。这里建议直接嵌套本地业务系统中的系统唯一号,方便D-DOIs与本地系统之间的互联互通。整个唯一标识符采用ISO/IEC 10646字符集,字符串使用UTF-8规则编码。

以国家科技图书文献中心联合服务场景为例加以说明。首先为NSTL共建共享体系分配一个前缀,作为预留名称权威“0.NA”之下的Handle来注册,例如“0.NA/110000”,表示注册了前缀“110000”。同理,包括我所在在内的9家文献信息服务机构加入了NSTL共建共享体系,各个机构经过注册得到各自的授权前缀,比如我所的前缀为“110000.110001”。当某个用户提交一条文献请求时,NSTL网络服务系统自动生成文献订购号“D1002250045BZ00”,该编号在系统内部是唯一标识。最终,我所承担了该篇文献的全文提供服务,则这次文献提供行为可以标记为“110000.110001/D1002250045BZ00”。

3.2 数据模型

(1) 元数据

在D-DOIs架构中,唯一标识符的注册与解析,以及相关的权限管理功能与Handle系统没有太多变化,因此唯一标识符的核心数据模型基本遵从了原来Handle系统中的定义,这里不再赘述,重点讨论的是针对具体应用的扩展部分,文中称之为D-DOI对象元数据。

D-DOI对象元数据的定义借鉴

了DOI应用实践。但是,与DOI唯一标识文献资源不同,D-DOI标识的是一次文献传递行为,不仅与传递的文献相关,也与请求服务的用户以及完成服务的机构实体相关。因此,D-DOI对象元数据应该完整地描述一次具体的文献传递行为,如图3所示。

图1给出了D-DOI对象元数据定义的树形结构。一个完整的D-DOI对象由document、user、institution和timestamp四部分组成,其中,document是被传递文献的描述元数据,可以是一篇期刊论文、会议论文、图书的某个章节、标准文献或其他文献类型;user是用户的描述信息,包括原系统标识、姓名、单位、电话、E-mail等联系方式;institution是完成文献提供服务的机构,根据需求可以精确到具体的人,用institution-department描述;timestamp则是发生本次传递行为的时间戳。由于用

户信息涉及个人隐私,因此D-DOIs规定对象元数据信息不对公众开放,即使是请求-解析的访问也需要身份验证。

(2) 数据存储

Handle系统缺省采用的是独有的内置数据库,没有图形化界面,直接访问很不方便。为了实现应用扩展,必须要有相应的数据支持,这时最好的实现方法就是:抛弃Handle内置存储模式,将其迁移到关系数据库上,对于构建在Handle核心之上的注册/解析等功能和自定义的应用扩展功能采用统一的数据库存储。

Handle系统本身是支持多种数据库的,因此实现起来并不困难。以MySQL为例,基本迁移步骤如下^[12]:

步骤1 安装MySQL的JDBC驱动包Connector/J。

步骤2 修改Handle系统的配置文件config.dct。

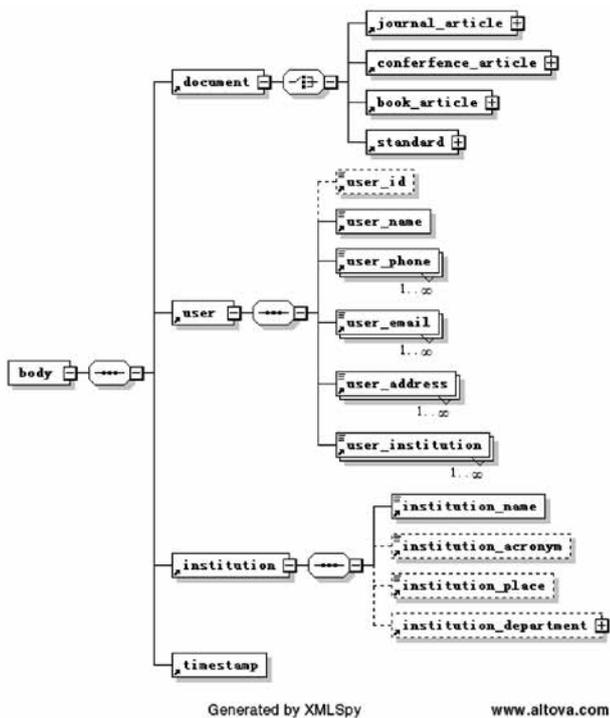


图1 D-DOI对象元数据结构

◆ 通知服务器唯一标识符的存储与检索由SQL支持

```
storage_type = "sql"
```

◆ 设置连接数据库的相关参数

```
sql_settings = {
    sql_type = "MySQL"
    sql_url = "jdbc:mysql://localhost:3307/DDOIHandle"
    sql_driver = "com.mysql.jdbc.Driver"
```

```
sql_login = "DDOIAdmin"
sql_passwd = "123456"
sql_read_only = "no"
```

步骤3 手工创建核心表na和handles。

```
create table nas (
    na varchar(255) not null,
    PRIMARY KEY(na));
Create table handles (
    handle varchar(255) not null,
    idx int4 not null,
    type blob,
    data blob,
    ttl int 4,
    timestamp int4,
    refs blob,
    admin_read bool,
    admin_write bool,
    pub_read bool,
    pub_write bool,
    PRIMARY KEY(handle,idx));
```

需要注意的是, Handle系统是大小写敏感的, “handle” 字段中要求使用大写字母。

3.3 主要功能分析

首先必须明确: Handle系统是建立在互联网架构之上的分布式信息系统, 提供有效的、可扩展的、可靠的全球名字服务。DOI系统是在Handle技术体系上建立的商业应

用, 其功能已经远远超出Handle系统本身。D-DOIs解决方案针对图书馆文献提供业务这一特定应用场景, 以Handle系统为底层解析框架, 借鉴DOI实现中的相关技术, 力图实现对传递中的电子文献资源的版权管理。其主要功能包括D-DOI生成、D-DOI解析、D-DOI嵌入、D-DOI提取和身份认证等, 如图2所示。

(1) D-DOI的生成和解析

D-DOI的生成和解析交由Handle核心程序完成, 如图3所示。

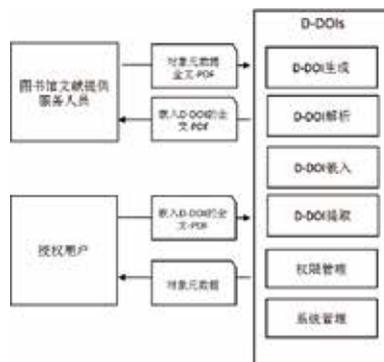


图2 D-DOIs功能框架

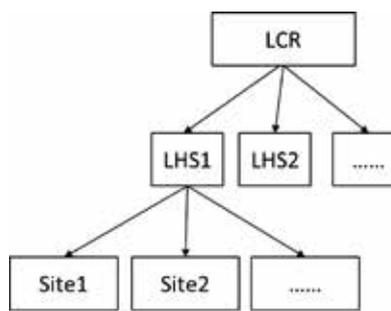


图3 D-DOIs注册-解析体系结构^[13]

LCR (Library Consortia Registration) 表示图书馆联盟注册中心, 其功能是管理命名授权; LHS (Library Handle System) 是图书馆本地服务, 实现本地域内的D-DOI对象命名授权和解析。LHS和其管理的所有D-DOI标识都要求在LCR中注册, 因此解析服务可

以通过本地服务器完成, 也可以提交给顶级服务器。图3显示, 每个本地服务可由一个或多个服务节点组成, 解析数据可以在不同节点间复制; 同时, 每个节点的工作也可以由多台服务器来完成, 能够有效地提高解析效率^[14]。

(2) D-DOI的嵌入和提取

D-DOI等相关信息采用数字水印和二维码两种方式嵌入待传递电子文献, 随着服务人员的传递行为发布出去, 并跟随着该电子文献传播。当出现版权纠纷时, 即可通过水印提取和二维码识别技术获得原始D-DOI编码, 进而通过系统解析查看原始版权信息, 从而保护图书馆和用户的权益。

◆ 数字水印

数字水印是DRM系统中最常采用的技术手段之一, 其基本思想是将水印信息隐藏在数字内容的原始数据中, 然后发布该水印版本作品。水印信息带有隐蔽性, 不会影响被保护数据的正常使用; 同时, 水印信息是足够安全的, 只有在数字内容被严重破坏的情况下才可能被抹去。当出现版权纠纷时, 所有者即可从水印版作品中提取水印信号作为依据, 从而保护所有者的权益^[15]。

图4描述了D-DOIs中采用的水印技术模型。其中, L为加入D-DOI体系的图书馆, L执行一次文献传递服务时, 向D-DOI中心注册唯一标识符, 然后选择一个L个人用的私钥 K^{Emb1} 向待传递的电子全文F嵌入含有本次传递标志(D-DOI)的数字水印watermarking₁, 得到F1, K^{Emb1} 由L的口令产生, 具有唯一性。再次将L标志L-DOI结合中心私钥 K^{Emb2} 对F1嵌入第二个数字水印watermarking₂, 得到F2。至此完成水印嵌入操作, F2随即进入流通领域。

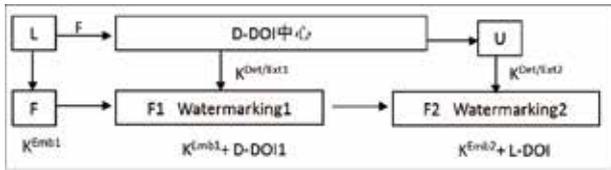


图4 D-DOI的嵌入和提取模型：采用水印技术实现^[16]

当L将嵌入水印的电子全文F2提供给用户U时，U可以用中心的公钥 $K^{Det/Ext2}$ 提取watermarking2水印，得到L信息；进而采用L的公钥 $K^{Det/Ext1}$ 提取D-DOI唯一标识符进行验证。

关于水印算法问题，虽然图书馆传递的电子文献一般都是PDF文件，但是在真实的应用场景中，这些PDF文件可能是可解码的，也可能是扫描图片转存的，因此需要考虑集成不同的水印算法，这涉及更深的专业内容，已经有很多可借鉴的研究成果，本文不再赘述。

◆ 二维码

随着移动网络环境下智能手机和平板电脑的普及，二维码应用迅速渗透到人们的日常生活中，图片、声音、文字等可以数字化的信息均能够进行编码，适用于产品追溯、物流仓储、身份识别等。将D-DOI标识符制作成二维码形式，显式嵌入待传递的电子全文中，可以方便地通过移动网络，实现电子全文的适时跟踪和追溯，防止文献的滥用。

只要掌握了二维码的编码原理，编程实现并不困难；同时，也有很多开源包可供使用，实现中需要注意的是二维码的容量问题。二维码的最大容量是由版本、纠错级别和编码模式共同决定的。一般来说，版本越高容量越大；同一个版本下，纠错级别越低容量越大；同版本同纠错级别下，可容纳的数

字、字符和多字节字符的数量又呈递减趋势^②。根据D-DOI唯一标识符的应用规模估算，可以采用版本9、Level M（15%纠错率）的编码设置，这时，一个二维码最多可以存储432个纯数字，或262个字母数字混合串，或180个8bit字节数据。图5是按照该设置生成的二维码样例，表示的是“110000.110001/D1002250045BZ00”。

这里同时采用了两种保护手段，是为了各取所长，互为补充。数字水印以不可感知的方式进行追踪，在网络传播下起主导作用；二维



图5 二维码样例：110000.110001/D1002250045BZ00

码则是为了适应移动环境而设计，方便随时取证。此外，系统并没有将对象元数据直接编码嵌入，这是从信息安全角度考虑的。因为，任何加密技术都不可能做到无懈可击，如果用户个人信息跟随数字内容广为传

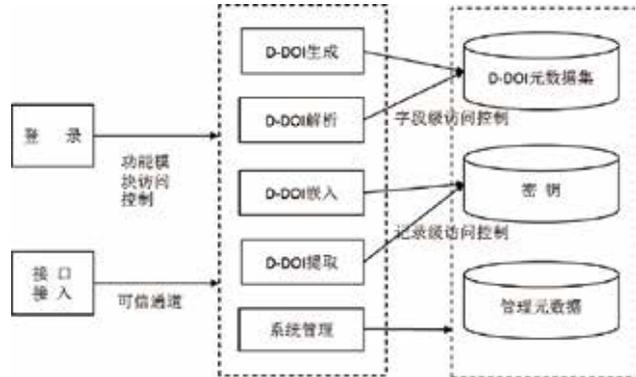


图6 D-DOIs访问控制框架

播，将面临极大的风险；而加密技术和D-DOIs身份认证相结合的设计，将大大降低这种风险。

(3) 权限管理

D-DOIs权限分为功能模块-表记录-表字段三个级别，如图6所示。功能模块的访问控制采用用户-角色-功能三个层次来管理。通过将功能分配给相应角色，然后将角色赋予相应的用户，从而实现对用户操作权限的控制。记录级别的访问控制主要针对密钥信息，也就是说，所有加盟机构的密钥存放在一个库表结构里，每个用户只能读写自己专属的记录。字段级别的访问控制针对的是唯一标识符元数据集，直接采用Handle系统的访问控制模式。

除了直接登录之外，D-DOIs还提供接口接入访问模式，主要支持图书馆文献提供业务与系统之间的无缝连接。通过接口访问也需要验证用户的身份信息，这是通过接口参数传递实现的。

4 结语

唯一标识符与版权管理的结合

^② 查看完整的QR码容量信息，可以访问<http://www.qrcode.com/en/vertable1.html>。

并不是一个创新话题,日本cIDf就是一个成功的范例,中国学者也提出了“基于中文内容唯一标识符的DRM技术方案”^[17,18]。当知识产权问题进入图书馆领域时,越来越多的学者加入研究队伍,然而迄今为止,并没有形成一个真正解决问题的方案,这是因为困在“标识什么?保护什么?”这个局里。DOI的优点和影响无需多说,通过DOI实现的版权保护是针对数字对象的,而大多数情况下,图书馆并不是数字对象的生成者,那么保护这些数字对

象的版权应该不是图书馆来做的事情,其实即使有心为之,也不是图书馆能力范围内能做的事情。从这些数字内容的生命周期来看,图书馆处于流通环节,知识产权侵权风险是在传播过程中产生的,那么监控的应该是从本馆这一流通渠道发布出去的数字内容,显然,对图书馆具体业务进行标识,对发布的特定数字内容进行跟踪保护才是有意义的。

本文锁定在数字图书馆的文献提供服务这一特定应用场景上,提出一个基于Handle架构的解决方

案,通过对文献传递行为的唯一标识,结合数字水印和二维码技术,实现对数字对象传播的持续追踪。这种方式并不能阻止侵权行为的发生,但它可以判别对象是否受到保护,监视被保护数据的传播、真伪鉴别和非法拷贝、解决版权纠纷并为法庭提供证据。D-DOIs解决方案不仅仅可以应用在一个自治的图书馆内部,实现本馆的文献传递服务跟踪;还可以扩展到一个图书馆联盟体系中,比如,在馆际互借机制的支持下可以实现各成员馆的服务跟踪。

参考文献

- [1] 胡锦涛,赵春艳.试论中外版权法中图书馆的“合理使用”[J].图书馆学刊,2004,26(6):18-19. DOI:10.3969/j.issn.1002-1884.2004.06.011.
- [2] 梅海燕.网络电子期刊的许可协议研究[J].情报资料工作,2003(3):34-37.
- [3] 冯尉.数字版权保护技术的发展探讨[J].数字图书馆论坛,2009(5):54-59. DOI:10.3722/j.issn.1673-2286.2009.05.010.
- [4] 谢俊,陈明,李晓明,等.流媒体版权管理系统的研究与实践[J].计算机应用与软件,2009,26(7):23-25,65. DOI:10.3969/j.issn.1000-386X.2009.07.009.
- [5] 向宏,周贤林,徐光侠,等.DRM技术在电子文档安全保护中的研究与应用[J].计算机时代,2006(11):24-26. DOI:10.3969/j.issn.1006-8228.2006.11.010.
- [6] 赵继海.基于DRM技术的电子书服务模式的构建[J].情报学报,2002,21(3):323-327. DOI:10.3969/j.issn.1000-0135.2002.03.013.
- [7] 刘芳,赵玥琪,卢国强,等.数字版权管理在数字图书馆的应用进展[J].医学信息学杂志,2011,32(7):62-66. DOI:10.3969/j.issn.1673-6036.2011.07.019.
- [8] 向林芳.论DRM在数字图书馆中的应用:以方正Apabi为例[J].高校图书馆工作,2011,31(6):85-87. DOI:10.3969/j.issn.1003-7845.2011.06.027.
- [9] 乔晓东.主要数字文献生产商唯一标识符应用现状研究[EB/OL]. [2013-03-15]. cdl2.nstl.gov.cn/mt/blogs/2nd/archives/docs/CDLS-S14-04.pdf.
- [10] 任瑞娟,刘丽斌,濮德敏,等.中文DOI路在何方:从参考文献著录与DOI的关系探讨中文数字对象唯一标识符的发展方向[J].中国图书馆学报,2010,36(2):115-121.
- [11] 任瑞娟,米佳,濮德敏,等.建立中文内容唯一标识符规范研究[J].图书情报工作,2011,55(10):15-18,14.
- [12] HANDLE.NET Technical Manual Version 1.1 [EB/OL]. [2013-03-15]. http://www.handle.net/support.html.
- [13] 梁健.Handle System在中文DOI系统中的应用[C]//《图书情报工作》杂志社第20次图书馆学情报学学术研讨会论文集,2008:178-181.
- [14] 毛伟,孙洵,王峰,等.一种互联网资源标志与寻址技术:Handle System[J].计算机应用研究,2004,21(5):252-254. DOI:10.3969/j.issn.1001-3695.2004.05.090.
- [15] 潘晓东.基于数字水印的PDF文档版权保护技术[J].科技资讯,2007(35):158-159. DOI:10.3969/j.issn.1672-3791.2007.35.130.
- [16] 孔祥维,杨德礼,胡祥培,等.数字水印和数字产品的版权保护[C]//管理科学与系统科学研究新进展,2001:221-226.
- [17] 郝军,顾君忠.开放式数字图书馆的Handle System实现机制[J].计算机工程与应用,2003,39(19):142-145.
- [18] 任瑞娟,张欣,张雨涛,等.基于中文内容唯一标识符的DRM技术与应用研究[EB/OL]. [2013-03-16]. http://www.docin.com/p-485746176.html.

作者简介

王莉, 硕士, 中国科学技术信息研究所研究员, 全国信息和文献标准化技术委员会技术协作分技术委员会委员。研究方向: 数字图书馆。E-mail: wangli@istic.ac.cn

Solution of Digital Rights Management for Document Supply Service

Wang Li / Institute of Scientific & Technical Information of China, Beijing, 100038

Abstract: For the scenario of document supply service in digital libraries, this paper proposes a digital rights management solution D-DOIs. D-DOIs solution is based on the Handle system. It identifies a specific document delivery behavior using D-DOI, combining with the digital watermarking technology and the QR code technology, which can enable the continuous tracking of the dissemination of the digital object.

Keywords: Digital rights management, DOI, D-DOIs, Handle system, Digital watermarking, QR code

(收稿日期: 2013-04-03)