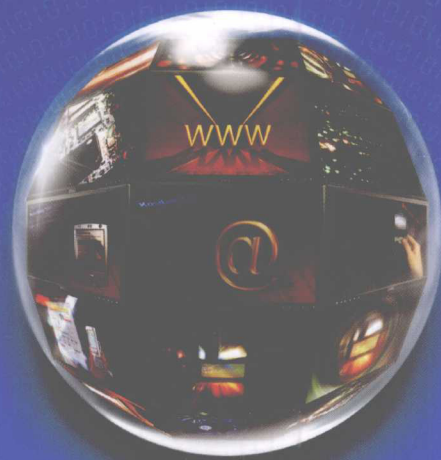




现代图书馆

XIANDAI
TUSHUGUAN

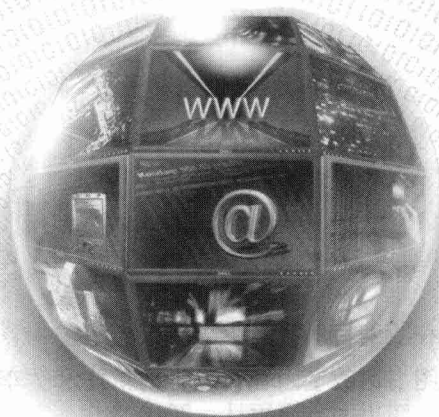
数字内容管理与 创新技术

李欣荣 著

SHUZI NEIRONG GUANLI YU
CHUANGXIN JISHU



四川大学出版社



现代图书馆

XIANDAI
TUSHUGUAN

数字内容管理与 创新技术

李欣荣

著

SHUZHIN
CHU



四川大学出版社

责任编辑:廖庆扬 曾 鑫

责任校对:傅 奕

封面设计:墨创文化

责任印制:李 平

图书在版编目(CIP)数据

现代图书馆数字内容管理与创新技术 / 李欣荣著.

—成都:四川大学出版社, 2010. 3

ISBN 978-7-5614-4772-7

I. ①现… II. ①李… III. ①数字技术—应用—图书馆工作—研究 IV. ①G250.7

中国版本图书馆 CIP 数据核字 (2010) 第 045581 号

书名 现代图书馆数字内容管理与创新技术

著 者 李欣荣
出 版 四川大学出版社
地 址 成都市一环路南一段 24 号 (610065)
发 行 四川大学出版社
书 号 ISBN 978-7-5614-4772-7
印 刷 四川锦祝印务有限公司
成品尺寸 185 mm×260 mm
印 张 9.25
字 数 206 千字
版 次 2010 年 4 月第 1 版
印 次 2010 年 4 月第 1 次印刷
定 价 16.00 元

版权所有◆侵权必究

◆读者邮购本书,请与本社发行科
联系。电 话:85408408/85401670/
85408023 邮 政 编 码:610065

◆本社图书如有印装质量问题,请
寄回出版社调换。

◆网址:www.scupress.com.cn

序

——天道酬勤

我无法告诉你们，我是如何写下自己的故事。在人生的路上，每个人都有着其独特的背景与包袱，等着你去勇敢地扛起或放下。对于那些迷失自己的道路后又重新发现生命的召唤的人，历经艰辛得到成果后的欢愉才更经得起细品与久藏。我可以向你保证，自我探索与努力的旅程，是世界上最值得走的一条路。而且我也可以向你们保证，只要还能够呼吸，就应该去满足你灵魂的渴望，并且与他人分享你的快乐，这些事情都永不嫌迟。

在写作的过程中“蜕掉”了那一层披挂了数十年的，像紧箍咒般紧紧缠住我的旧皮，真正脱胎换骨了。我常常害怕失败，太在乎别人的看法，而且过于羞怯。现在，我甩掉了这些枷锁（至少很多时候如此）。几年前，我从未想过会写下这本书，甚至动笔之前，都怀疑自己的能力。而今，所有的一切都成真了，那是梦想的力量又一个证明。

一路行来，我理解不管自己写或不写这本书，命运已经如此，因此拿起笔，整理多年来的成果。我让自己生命的故事，成为我想要的形式，也为我的女儿作一个榜样。

尽管书中的一些观点仍有待商酌，但我仍渴望能超越自我。但是大部分的时候，我以自己为傲，以我的书为傲，以我生存在这个地球上为傲。因为我相信人类最高的自我是我们存在的光芒与现实，那才是真正的自我。而当我们能够真诚地以心传心，将可以彼此提升，以奇迹净身。

希望通过对多年来成果的总结，不只让我精神一振，也能够让你们相信，在生命中，真的有更多的可能。

李欣荣

2009年6月于西华大学图书馆

序

——天道酬勤

前 言

进入 21 世纪以来,数字内容的管理与创新研究作为现代数字图书馆的一个新兴领域得到广泛的重视和迅速的发展,主要体现在:

- 解决数字内容的盗版贩卖和非法使用的问题。
- 解决非法及有害内容的破坏和污染社会环境问题。
- 解决数字内容使用者的安全合理支付问题。

数字内容的管理与创新研究与实践在这样短的时间内取得如此迅速的发展,无疑是一件令人振奋的事。但从整体上看,也存在一些亟待解决的问题,集中表现为:理论层面上,虽然成果丰硕,但多数研究缺乏深度和广度,尚未形成公认的核心概念和完整的理论体系,理论研究的滞后制约了数字内容的管理与创新研究的可持续发展;实践层面上,由于缺少理论的指导,数字内容的管理与创新研究发展迅速的同时,也带有不同程度的盲目性与随意性,在实际操作过程中出现了诸多问题,需要从理论与实践的结合上给出解释和回答。

所以,在国内外学者研究的基础上,结合数字图书馆的具体情况,通过必要的整合、凝练和新的研究以达到更高的发展水平,是数字图书馆内容服务研究领域的发展途径。

写作本书即是对这一发展方向的努力。概括地说,本书立足于数字内容管理与创新的技术实践,在现有研究成果的基础上,对现代图书馆数字内容的管理与创新研究进行了全面的梳理,提炼出既具普适性又有针对性的原理与方法,从整体上构建起数字图书馆内容管理与创新研究的框架。

需要说明的是,现代图书馆数字内容的管理与创新研究内容非常丰富,在有限的空间中不可能面面俱到,为兼顾全书的体系结构,突出重点,避免过于臃肿,有些内容只能从略。

现代图书馆数字内容的管理与创新研究是一个正在迅速发展的学科领域,虽然在本书中我们力图反映最新的研究动向,但还是难以跟上现代图书馆数字内容的管理与创新研究快速发展的步伐。一些新的内容没有能够在书中得到充分反映,有些重要的理论、方法与技术问题的探讨还要继续下去;特别是 DOI、DOM 以及新技术方面的突破性进展,揭示了数字图书馆内容管理与创新研究的新篇章,热切的期待同行们能在后续的著作中予以全面充分地反映。

数字内容的管理与创新研究是一个全新的研究领域,许多方面的研究都才刚刚开

始。本书的宗旨是以著者对现代图书馆数字内容的管理与创新研究的理解为基础，着重论述基于数字内容管理与创新的数字图书馆特色技术，揭示数字内容管理与创新研究的本质及主流发展趋势。因此，书中包含了许多著者个人的看法，目的在于抛砖引玉，引起思考。真诚地希望得到学术界同仁的赐教和指正，共同推进现代图书馆数字内容的管理与创新研究更加全面、系统并走向深入。

目 录

第 1 章 绪 论	(1)
1.1 数字内容的管理与创新概述	(1)
1.2 数字内容管理与创新研究的意义	(7)
第 2 章 数字版权的保护技术	(8)
2.1 DRM 技术	(8)
2.2 数字水印技术	(20)
第 3 章 IPSec	(29)
3.1 IPSec 体系结构	(30)
3.2 安全策略系统概述	(33)
3.3 移动 IP	(34)
3.4 移动 IP 中的 IPSec	(36)
3.5 IPSec 在移动 IP 中的应用	(38)
3.6 IPSec 在 IPv6 上的应用	(39)
3.7 基于 IPSec 技术的访问控制研究	(40)
3.8 基于 IPSec 的移动 IP 用户访问控制解决方案	(42)
3.9 FreeS/WANIPSec 安全技术数字图书馆虚拟专用网控制信息访问中的实现	(44)
第 4 章 数字对象标识符 DOI	(48)
4.1 DOI 系统的构成	(48)
4.2 DOI-EB 工程	(51)
4.3 DOI 在国内外的发展现状	(52)
4.4 DOI 在学术出版中的应用	(54)
4.5 DOI 的注册	(57)
4.6 DOI 技术在数字内容的商业化中的应用	(59)

第 5 章 DOM (文档对象模型)	(62)
5.1 DOM 简述	(62)
5.2 DOM 的发展	(68)
5.3 DOM 解析 XML 文档	(70)
5.4 基于 DOM 的数字图书馆 Web 资源信息提取服务	(71)
5.5 基于 DOM 中间件技术在数字图书馆资源发现中的应用	(72)
第 6 章 SFX 技术在资源整合利用中的应用	(77)
6.1 概 述	(77)
6.2 SFX 在信息资源整合中的优势	(78)
6.3 资源整合利用的基本思路	(80)
6.4 sfx 技术在资源整合利用中的应用	(81)
6.5 资源整合利用检索的“度”	(84)
第 7 章 数据访问控制技术	(86)
7.1 数据访问控制技术综述	(86)
7.2 数据访问安全保护的 3 个环节	(90)
7.3 XACML 在控制信息访问中的应用	(92)
7.4 SAML 安全技术数字图书馆控制信息访问中的应用	(96)
7.5 数据访问技术的发展与展望	(99)
第 8 章 数据存储技术	(104)
8.1 数据存储技术综述	(104)
8.2 图书馆数据储存的选择: NAS-SAN 存储技术	(119)
8.3 网格存储	(123)
8.4 数字图书馆对存储方案的要求	(125)
8.5 基于网格的数据存储技术在数字图书馆中的应用	(126)
8.6 基于 NDMP 协议的数据管理	(130)
8.7 基于 NDMP 技术的现代数字图书馆数据管理模型设计	(132)
参考文献	(135)

第1章 绪 论

1.1 数字内容的管理与创新概述

数字内容的管理主要涉及数字内容的加工、共享、创新3个方面。现在分别概述其研究情况：

(1) 数字内容的加工主要是对数字内容进行整序和清洗，提供查询、统计和分析，建立链接，以便找到原始地址。还可以进行二次、三次等的数字内容加工。这里主要是对数字内容进行主体标识、分类、关联、浓缩等，以供检索查询用，这在图书馆中较为常用。

(2) 数字内容共享是实施数字内容管理的关键所在。能否在平衡各方利益的基础上实现数字内容长期的、有效的资源共享是抑制数字内容管理与创新发展的瓶颈所在，所以数字内容的安全保护技术就显得非常的重要。

无论是学术界还是产业界，关于数字内容安全保护的内涵研究尚未形成比较统一的认识。目前的研究从一般的信息安全的概念出发，主要涉及数字内容的隐私性、完整性和真实性。从理论上讲，这种概念是没有问题的，但面对实际应用却显得抽象和空泛。针对目前数字内容在开发制作、传递配送和消费使用中的主要问题，保障数字内容安全的关键是：

- 解决数字内容的盗版贩卖和非法使用的问题。
- 解决非法及有害内容破坏和污染社会环境问题。
- 解决数字内容使用者的安全合理支付问题。

针对第一个问题，提出了数字版权管理技术，采用加密手段对数字内容进行保护，使其只能在授权的情况下被使用；针对第二个问题，提出了基于内容的过滤（CBF）技术，采用文字识别、语音识别、图像识别、文本分类等模式识别的方法，将非法或有害的内容进行过滤和封堵；针对第三个问题，目前正在大力研究微支付技术，基于公钥基础设施（PKI）和第三方代理等平台来保证使用者资金的安全。

目前有关数字内容的版权保护标准方面的研究已经展开：开放移动联盟（Open Mobile Alliance, OMA）标准、互联网流媒体联盟（Internet Streaming Media Alliance, ISMA）标准等已经逐步完善。数字水印理论方面的研究也比较成熟，具体的应用尚待开展。

(3) 数字内容的创新已成为数字内容发展的灵魂,是提升数字内容产业竞争力,以及在竞争中立于不败之地的重要手段。目前的研究侧重于单纯的内容创新,而对于生产活动中所涉及的管理上的创新、制度上的创新、组织文化上的创新、经营活动的创新、生产技术上的创新以及知识发现与创造研究不足。

1.1.1 数字内容及其安全的基本概念

随着数字化技术的发展,数字内容的内涵日益丰富,主要包括数字音像、科学出版、远程教育、动漫游戏、金融信息、政府公告、网络博客、网络论坛、短信彩信、彩铃音乐等,涉及教育、科学、金融、文化、娱乐、商业、通信等多个领域。围绕着这些数字内容的开发制作、传递配送和消费使用,一个影响全社会的大规模的产业链正在形成。人们认识到这个产业链无论在市场规模方面,还是在对现有产业结构的影响方面都潜力巨大。

从技术方面来讲,数字内容开发、数字内容传递和数字内容安全是数字内容产业的三大支撑。数字内容开发一方面与文化创意和艺术创造紧密结合,同时也与图像、音频、视频、Web2.0 等技术不可分割;随着宽带技术的发展,数字内容传递正在由传统的离线配送向互联网在线传递和移动传递的方向急剧转变,网络门户、搜索引擎、无线宽带、移动交互等技术成为数字内容传递的核心技术;数字内容安全则包括数字版权管理(DRM)、非法及有害内容过滤、网络支付安全等重要内容。

1. 数字内容保护技术

在数字出版方面,我国相关管理部门的政策大多集中在对网络出版和版权问题的管理方面。2002 年,国家新闻出版总署与信息产业部联合颁布了《互联网出版管理暂行规定》,对互联网出版活动的审批、出版内容的审查做出了比较严格的规定。同时,鉴于网络环境下版权对数字内容产业发展的重要性,国家版权局正在计划构建网络反盗版平台,网络版权保护的法律体系也在不断完善之中。另外,国家版权局颁布的《互联网著作权行政保护办法》、2006 年 7 月生效的《信息网络传播权保护条例》,以及有关法律规定和规范性文件对数字内容产业知识产权的保护和打击网络侵权盗版行为也起到了一定的监管作用。

在数字版权方面,数字内容产业系统中的交易规则还未建立,特别是版权交易管理和规范非常复杂,基本上还处于一个“打官司”的阶段。侵害版权事件屡屡发生,以因特网为例,在因特网上大量没有合法版权的数字电影、数字音乐正在非法传播。国际知识产权协会(HPA)估计,美国电影业每年由于盗版而导致的损失达到 13 亿美元,在唱片和音乐行业这个数字是 17 亿美元。就移动下载而言,在移动运营商获得收益的同时,服务提供商也获得了相关的商业利益,这更要求对网络内容进行保护,以避免由于版权侵害使得内容提供商不愿意提供高质量的内容,影响整个数字内容产业链的良性发展。

法国 2006 年 6 月通过了《信息社会版权法案》,为其他国家解决信息产品版权纠纷提供了一些宝贵参考。可以说,在数字内容产业交易规则体系发育之初,版权的法律规范和界定尚不明确,极大地影响着整个产业的各个环节,包括从“原创”生产系统到集

成环节、传播环节和交易环节等。只有对版权保护进行规范,在注重保护版权、打击盗版的同时,对数字内容产品的“通用兼容性”也加以明确保护,才能真正地“从头保护”这个产业。

(1) 基于网络的数字内容的资源发现。在变化多端的电子出版世界中,由于资料版权所有者的更迭,电子文件的存放位置变化频繁,尤其那些在各主机之间来回转换的文件、新版的软件、不同文档格式的数字化音乐作品,或经过更新内容的科学文章等目标位置改变时,给合法使用者的使用带来了极大的麻烦。目前迫切需要一种技术,以保证即使被链接的版权所有者、出版时间和地点有了变化,仍能正确无误地链接。

应该有一种类似图书 ISBN、期刊 ISSN、录音录像资料 ISRC 等让使用者很容易检索的标识码(ID code),以满足电子环境中各种识别与贸易的需求,同时使开发商在保障提供来源稳定的资料与保护版权两者间获取最大利益。

(2) 基于网络的数字内容的加密管理。通过加密数字内容并在其中附加判断用户是否具有合法使用权的方法,来达到保护数字内容的目的,从而在技术上防止数字内容的非法复制,使最终用户必须得到授权后才能使用数字内容,让作者和出版社的电子版权得到保护。通过数字内容的销售,作者和出版社得到相应的收益。

数字内容销售数量是可计数的,数字内容销售网站能从出版社得到数字内容的销售许可,读者通过网上支付购买数字内容。出版社也可以把数字内容通过销售渠道卖给图书馆,图书馆购买数字内容就像购买纸书一样,按复本数购买,对读者提供借阅的服务。

2. 数字学习

教育部所提出的教育信息化概念中包含了数字学习的部分内容,2006年颁发的《全国教育科学“十一五”规划纲要》中规定的教育科学的3个重点领域也包括教育信息化。该纲要提出,要“深化数字化学习环境建设、数字化教育资源建设和共享机制研究,推动网络课程开发研究”“加强信息技术在教育过程中的应用研究,促进信息技术环境下学习型组织建设与培训模式研究、远程教育发展与质量保证研究、农村远程教育可持续发展研究、学习者能力发展研究”。

由此可见,政府的管理政策中存在对数字化学习的一定关注,但相关管理部门的针对角度是针对教育信息化,具有一定的公益性质。

随着以计算机网络为代表的信息科学技术的蓬勃发展,带动了数字化学习的浪潮,使学习不再局限于固定的时间、地点及教材,而发展成为可在任何时间、任何地点、任何环境下的“数字学习”。目前在“数字学习”方面的研究,主要针对各种不同使用目的的数字学习平台的开发,学习者可以通过各种方式存取平台上的数字化内容。

我国教育部教育信息化技术标准委员会制定的平台与媒体标准引用实践指南 CELTS-17。该标准是对教学系统定制和资源建设过程所参照引用的已有标准和规范进行描述,提供几类规范化的标准引用谱,即标记语言、音频格式、视频及图形格式、页面描述语言、Java 技术、JavaScript 技术、文字处理格式、演示文稿格式、电子表格格式、文档服务。本标准也将针对教学系统所处的操作环境(如浏览器、工作站等),提供常见的标准引用组谱。标准引用谱和标准引用组谱可直接用于现代远程教育系统运

行环境定制和教育资源的描述。这样,无论是远程教育系统运行环境,还是教育资源,都符合标准化,有利于资源重用,确保系统间的互操作性和兼容性。

不过,在针对数字学习平台的多样化特性衍生出的数字内容无法兼容及再利用的问题,数字学习平台在支持带有相同角色或不同角色的多人合作学习,能根据个人特点和具体情况,调整学习内容和学习活动,从而达到支持个性化学习目的等方面的研究还不足。

数字技术给数字内容提供了可能性,对内容的创作收集、存储和传输都带来了革命性的影响,怎样针对数字内容商业化过程中存在的数字内容进行加密保护,以及如何开展数字学习是目前数字产业发展中需要解决的问题。

1.1.2 数字版权管理技术

数字版权管理并不仅仅是个技术问题,它要通过技术、法律、商业等各种有效手段保证数字内容在制作、传递和消费各个环节中不受到盗版、侵权和滥用,以保护所有者的知识产权,但本文只讨论通过技术手段实现的 DRM 系统。

目前,常见的 DRM 系统由 3 部分组成:数字内容供应者(CP)、许可证发放器(LD)和用户播放器(UP)。

CP 利用打包程序将数字文件进行加密。目前常用 128 位或 156 位的对称加密算法。密钥利用与 LD 共享的密钥种子和一个全局唯一的密钥标识生成。内容加密后,再添加作者、版本号、发行日期、密钥标识等头信息。打包后的数字文件可以存放在 CP 的网站服务器上,也可以制成光盘发行。

UP 在访问 CP 网站服务器或通过光盘播放打包的数字文件时,首先在自己的许可证库中查找所需要的许可证(解密密钥)。如果存在,便可播放;如果不存在,则必须向 CP 指定的 LD 申请播放该数字文件的许可证。

LD 接到 UP 的许可证申请后,对用户的身份进行验证,如果是合法用户或通过付费等手续成为了合法用户,则向 UP 发放播放该数字文件的许可证。许可证可根据需要设置有效期和不同的收费标准等。

在 DRM 系统中,加密的作用也可以用数字水印或数字签名技术来替代。学术界和产业界在这些方面开展了大量的研究和开发。DRM 系统用不同的信息安全技术来实现,系统的性能和实现成本是不同的。

近十年来,图像、视频、音频等多媒体的加密算法得到了深入研究。人们越来越多地将加密过程与压缩编码过程相结合,以同时获得较高的安全性和较高的压缩率。同时还进一步考虑多媒体网络、无线网络、移动网络的带宽和可靠性的特点,研究开发满足异构网络环境下可伸缩性和实时性要求的加密算法。

相对于互联网,DRM 在移动网络上发展得更加迅速。主要原因如下:移动网络相对封闭,DRM 系统易于建立,且不易受到攻击。移动网络用户数量巨大,受 DRM 保护的数字内容在这一平台上大量发布会降低数字内容的成本,有利于正版数字内容的推广和知识产权保护。

2002 年 11 月,开放移动联盟(OMA)发布了移动 DRM 国际规范——OMA DRM

V1.0 Enabler Release, 为如何建立移动网络上的 DRM 系统提供了指南。OMA DRM V1.0 标准推出后, Nokia、Motorola 等著名国际厂商纷纷进行了相应开发, 对其存在的问题进行了公开讨论。

2005 年 6 月, OMA 公布了 OMA DRM V2.0, 制订了基于 PKI 的安全信任模型, 给出了移动 DRM 的功能体系结构、权利描述语言标准、DRM 数字内容格式 (DCF) 和权利获取协议 (ROAP)。

1.1.3 基于内容的过滤技术

基于内容的过滤是数字内容安全的重要内容。CBF 的主要对象包括非法内容和有害内容, 如非法广告、黄色信息、谣言、网络病毒、黑客攻击等。早期的 CBF 技术主要采用串匹配的方法对文本文件和可执行文件进行过滤, 防范的对象是有害文本信息和病毒。随着多媒体技术的发展, 非法和有害的信息开始大量地利用图像、视频、音频等形式传播, 使得简单的串匹配技术无法对内容进行有效识别。在这种情况下, 人们开始将模式识别、自然语言处理、机器学习等智能技术引入 CBF。此外, 基于上述智能技术的文本分类和挖掘也取得了长足的进展, 从而推动 CBF 全面进入了以智能技术为依托的阶段。

在文本文件过滤方面, 通过向量空间模型 (VSM) 或 n-gram 语言模型对文件进行表达, 然后利用正反两方面的样本对需要过滤和不需要过滤的两类文件进行建模, 从而生成可执行特定任务的分类器, 如 Bayes 分类器、SVM 分类器、k-NN 分类器等。将这样的分类器放在网络节点或主机上, 便可实现文本文件的过滤。目前最常见的文本文件过滤器是垃圾邮件过滤器, 国际著名会议 TREC (Text REtrieval Conference) 从 2005 年开始将垃圾邮件过滤器作为测试项目, 有力地推动了该项技术的发展。在中国, 除了垃圾邮件之外, 垃圾短信等短文本中的非法有害信息的过滤也得到了学术界、产业界和政府的高度重视。目前已经有国家自然科学基金、国家信息安全计划、跨国企业资助的项目在加紧研究。

在图像和视频文件过滤方面, 文字识别、人脸识别、人体识别、物体识别等图像识别技术是核心。通过这些技术, 可对文件中包含的字牌、标语、广告等反映不同场景的文字, 以及人脸、人体、物体等反映不同人物和事件的对象进行识别。获得这些关键信息后, 便可以对图像和视频进行分类和过滤。例如, 对黄色图片进行过滤, 对毒品广告进行过滤等。在上述图像识别技术中, 人脸识别和物体识别是当前的研究热点。文字识别是开展较早的研究, 但图像中的文字识别有其特殊性, 如倾斜和光线的影响等。关于人脸识别和物体识别, 近年来人们给予了极大的关注, 并取得了显著的进展。2007 年国家自然科学基金的一个有关物体识别的重点项目吸引了全国 11 个颇有实力的科研单位的申报, 竞争之激烈实属罕见。在国际上, 物体识别的研究正在越来越紧密地与网络图像检索和过滤相结合。

在音频文件过滤方面, 语音识别、语种识别、语音关键词检测技术是核心。对于安静环境下的新闻播报类语音文件, 先通过语音识别技术将其转换为文本文件, 就可以利用文本过滤技术进行过滤了。美国国家标准技术研究所 (NIST) 和国防部的话题检测

与追踪 (TDT) 计划对这项技术进行了长期的研究, 取得了令人瞩目的进展。目前的研究热点是噪声背景下的语音文件或歌曲音乐类文件的过滤。这类文件不易用通常的语音识别方法进行内容识别, 需要研究专用的方法。利用语种识别和语音关键词检测技术进行过滤时, 不需要将整个文件转换成文本, 而只需识别文件中的语音是不是指定的语种或是否包含指定的关键词。语种识别和语音关键词检测常被用于粗过滤, 以提高过滤器的效率。

在网络环境中, 过滤器的效率是一个突出问题。基于智能技术的过滤器通常具有较高的计算复杂度, 时间开销较大。其主要原因是文件表达模型的特征向量维数过高。如在文本分类中, 常常采用几万维的特征向量, 每一维对应一个词。因此, 特征降维已经成为特别重要的环节。简单的特征降维方法是特征选择, 即从现有的特征中优选一部分。还有一种方法是高维空间向低维空间映射变换的方法, 通过去除数据值方差小 (能量小) 的维度, 进行降维。如主成分分析 (PCA)、线性鉴别分析 (LDA)、流形分析、图模型等。这些方法的研究, 具有非常重要的普遍意义, 已经成为本领域的研究热点。

1.1.4 微支付技术

在线数字内容的消费常常金额很小, 例如, 下载一首歌曲、一个彩铃、一篇论文, 甚至书中的一页内容。这样的消费金额难以采用常规的方法进行消费者和商家之间的结算, 因为结算本身的成本相对消费金额太高, 甚至会超过消费金额。例如, 如果下载一首歌是 5 分钱, 那么通过通常的银行手续去交钱, 光手续费可能至少就要 1 元。这样的结算是消费者和商家都不愿接受的。因此, 数字内容的消费离不开微支付技术的支撑。

所谓微支付就是对任意小的消费金额进行电子支付的技术。它要解决的主要问题除了保证消费者在电子银行中的资金和数据的安全、商家不被骗取、交易数据不被篡改之外, 就是以最低的成本实现电子付费, 以保证交易成本不超过消费金额。目前, 常见的微支付方式包括网络在线支付、手机支付、电子支票支付、信用卡支付等。

微支付系统中的核心技术包括 PKI 技术和交易代理技术。通过 PKI 技术对交易所涉及的各方的标识符、交易数据等进行加密, 以防止伪造身份、盗取密钥、破解消息等攻击的得逞。通过交易代理技术, 实现信用担保、身份认证和公平交易, 交易代理通过可转移硬币等技术, 最大限度地降低交易成本。目前微支付研究的重点在于协议和系统模型。微支付协议分为离线方式和在线方式两大类。典型的离线微支付协议包括 MPTP、Payword、Agora 和 MiniPay 等。这些协议以消费者的信用为基础, 消费者在真正付款之前就可以完成交易, 因此对重复消费 (同一凭据反复使用) 和恶意消费 (透支消费) 缺乏有效的控制。典型的在线微支付协议是 Millicent, 它采用交易代理在线实时验证消费者帐户信息的方式, 可以有效防止重复消费和恶意消费, 但也因此降低了协议的运行效率。

微支付协议和模型的优劣, 主要从安全性、公平性、交易成本、运行效率等方面进行评价。安全性主要指交易者的身份不被伪造和不被泄露, 以保证交易者的资金安全和交易的隐私; 公平性主要指在整个交易过程中, 消费者、商家和交易代理受到平等的对待, 消费者的信用得到正确的评估, 商家不受到欺骗, 交易代理得到合理的利益; 交易

成本要尽量地降低,以满足微支付的要求;运行效率要尽量地提高,协议的时间开销和空间开销要尽量小。

微支付协议和模型与系统所基于的网络有密切的关系。例如,基于移动通信网络的微支付系统、基于 WWW 网络的微支付系统、基于 P2P 网络的微支付系统等相互之间有明显的差别。总体上讲,现有的系统还不能满足目前数字内容产业迅猛发展的要求。OECD 的报告认为,微支付系统技术上的滞后,制约了数字内容的消费。

数字内容产业的规模正在迅速膨胀,对 IT 产业结构的冲击正在日益加强。因此,数字内容技术已成为各国奋力抢占的一个战略至高点。数字内容安全是数字内容技术的重要组成部分,没有它的保证,数字内容产业就难以发展。同时,数字内容安全涉及众多具有挑战性的科学和技术难题。例如,高效的多媒体加密和隐藏算法、网络多媒体内容识别和过滤算法、面向异构网络的微支付协议等。攻克这些挑战,需要学术界、产业界和政府的共同努力。同时,也只有这样,才能使中国的数字内容产业迅速发展起来,成为中国 IT 产业强劲的推动力。

从技术方面来讲,数字内容开发、数字内容传递和数字内容安全是数字内容产业的三大技术支撑。数字技术给数字内容提供了可能性,对内容的创作收集、存储和传输都带来了革命性的影响,如何针对数字内容商业化过程中存在的数字内容的加密保护以及如何开展数字学习是目前数字产业发展的趋势。

在各种信息资源都走向数字化、多种媒体不断融合的趋势下,数字内容管理成为一个全新的课题。目前主要研究方向集中在数字内容的内容提供、平台搭建(宽带技术、接口技术等)、监管体系等,而对于数字内容的数字版权保护标准、数字内容的传递、基于数字内容的资源挖掘、异构资源的整合与检索、数字内容的可视化等方面的研究还有待深入。

1.2 数字内容管理与创新研究的意义

1. 保护数字内容供应商的利益

避免数字内容商业化过程中,因使用者滥用数字资料而使数字内容供应商蒙受严重的经济损失,让公益性政策(如公平使用原则和馆际互借等制度)延续下去。

2. 为使用者带来方便

使用者需要一种技术,以保证虽然被链接的版权所有者或出版时间和地点有了变化,其识别符仍能保持一种持久性,需要一种识别途径,将使用者和相关的资料链接在一起。

3. 数字学习

目前的数字学习研究以内容导向为主,学习者仅能以直线式的存取方式来阅读数字教材,缺乏学习者间的互动及情境式学习。制定基于数字学习的国内标准、开发以使用者为主体的数字学习平台是当前急需解决的问题。

第2章 数字版权的保护技术

随着互联网的发展,网上电子书、音乐、电影、图片等数字内容的传播越来越多,我们可以实现随时随地的下载。据有关机构调查,数字内容很容易复制、修改,网络上传播的数字内容存在大量的盗版和侵权现象。这种现象也引起了越来越多作者和发行者的关注,数字版权管理开始成为互联网时代的一个热点话题。

数字版权的保护技术主要有两类,一类是采用数字水印技术,另一类是以数据加密和防拷贝为核心的 DRM (Digital Rights Management) 技术。

2.1 DRM 技术

以数据加密和防拷贝为核心的 DRM 技术,是把数字内容进行加密,只有授权用户才能得到解密的密钥,而且密钥是与用户的硬件信息绑定的。加密技术加上硬件绑定技术,防止了非法拷贝,这种技术能有效地达到版权保护的目的,当前,大部分计算机公司和研究机构的 DRM 技术采用这种方法。

DRM 系统是内容保护的研究方向之一,它的目的就是保护版权和数字多媒体资源的经济价值。目前,多媒体资源的一部分应用中,要求用户能从加密的资源中得到一部分信息,可以进行浏览,但是未解密时得到的图像质量会比较差,比如付费电视节目、远程教育等。另外,由于在传输过程中可能要经过不同的网络,还要求视频在传输中能适应不同的传输协议和带宽。数字版权管理,是流媒体中进行版权控制的关键性技术,它可以在一定程度上避免媒体文件被数字化后的盗版问题。

DRM 技术的工作原理是,首先建立数字文献授权中心,编码压缩后的数字文献内容利用密钥 (Key) 加密保护 (Lock),加密的数字文献头部存放着 KeyID 和文献授权中心的 URL。当用户提出访问需求时,根据文献头部的 KeyID 和 URL 信息通过数字文献授权的准授权后送出解密 (Unlock),文献方可传送。

需要保护的文献被加密,即使被用户下载保存,没有得到数字文献授权中心的准授权,也无法使用,从而有效地保护了数字文献的产权。

密钥一般有两把,一把公钥 (Public Key),一把私钥 (Private Key)。公钥用于加密内容本身,私钥用于解密。当文献头部有被改动或破坏的情况,利用密钥就可以判断出来,从而阻止文献被非法使用。上述这种加密的方法,有一个明显的缺陷,就是当解密的密钥在发送给用户时,一旦被黑客获得密钥,即可方便地解密,不能真正地确保数

字内容商的实际权益。

另一种更加安全的加密方法是使用三把密钥，即把私钥分成两把，一把存放在用户的 PC 机上，另一把放在验证站（Access Ticket）。要解密数字内容，必须同时具备这两把密钥，才能正常使用数字内容。这样，当解密密钥在发送给用户时，即使被窃取，也仍然无法解开加密的内容。

2.1.1 DRM 特点及应用

所有的 DRM 应用都是基于“系统辨别”和“信息记录”的。记录里面包括了合法版权所有者的情况和版权信息。这些情况或信息是通过元数据（Metadata）、版权管理信息（Rights Management Information, RMI）或数字对象标识符（Digital Object Identifiers, DOI）来实现的。

1. DRM 几大要素

（1）元数据

元数据对于查找信息是非常有帮助的，它能描绘出作品的基本信息。使用元数据存在两个问题：一是不同的人在不同的时间，不同的背景下对同一事件采用不同的关键词来描述；二是元数据的结构兼容性，比如，A 标准会用<姓>、<名>来记录一个人，而 B 标准用<姓名>来记录，那么，除非进行转换，否则 A 标准无法和 B 标准兼容。因此，世界上很多组织做了大量的工作，就是为了建立持久通用的元数据结构标准。我国在这方面也做了许多工作，如 1996 年的《中国文献编目规则》，1998 年的《音像资料机读目录》等。

（2）版权管理信息

简单地说，版权管理信息就是一些可以进行版权管理的信息。在美国 1998 数字千年法（DMCA）中，将破解版权保护技术视为违法行为，并定义了版权管理信息。这些版权管理信息，如作者、联系方式、授权条件和权利有效期等，不仅可以标示权利人，方便用户获得作品使用许可，还可监控用户的使用情况，跟踪侵权行为的发生。这在很大程度上，既杜绝了数字作品的不正当使用，又能够保证数字出版、改选和传播等环节能在合理的法律环境下有秩序地进行。而目前我国数字出版行业的法律环境还远未达到这样的程度。

（3）数字对象标识（DOI）

DOI 系统类似于用来识别物品的条形码，注册代理机构为每一个数字对象分配识别号码，并记录描述该对象的元数据。假如能够用相对简单的办法跟踪对象和版权所有者情况，那么元数据就可以被不断更新。

2. 保护版权的手段

在 DRM 系统的实现方案中，采用了许多技术用来避免知识产权被侵犯。每种手段都有各自的优缺点，采用什么方案需要综合考虑其优缺点、维护成本、被保护作品的重要性、作品被侵权的可能性和打算投入的资金量。

没有任何一个技术系统是 100% 安全的，DRM 为数字内容提供了相对高水平的安