网络引文不可追溯性及其解决方案研究*

陆 伟 韩曙光 沈祥兴

摘 要 互联网的发展使得信息资源的存取更加便捷,但网络引文的不可追溯现象也随之凸显。国内外关于网络引文可追溯性的研究大多集中在可追溯性现象及规律方面,对不可追溯问题解决方案的研究偏少,与此相关的实际应用系统更少。网络引文追溯平台的构建应该着重解决网络信息资源位置和内容的"变动性",该平台可由网络引文库构建模块和网络引文集成检索模块构成,以实现最大限度地追溯呈现网络引文。图3。表1。参考文献31。

关键词 网络引文 不可追溯性 网络信息保存 网络引文库 集成检索系统
分类号 G252.34

ABSTRACT The developments of Internet make information retrieval more convenient, but the inaccessibility of Web citation also follows. Research papers mostly focus on the phenomena and rules of accessibility, but pay fewer attentions on the solutions and applications of inaccessibility. The construction of web citations accessibility platform should focus on the locations of Web information resources and the content mobility. This platform should include two models, such as Web citation database and Web citation retrieval system, which can access Web citation as much as possible. 3 figs. 1 tab. 31 refs.

KEY WORDS Web citation. Inaccessibility. Web archive. Web citations base. Integrated retrieval system.
CLASS NUMBER G252.34

1 引言

互联网的日益普及和在线出版的进一步发展,使得网络学术资源的获取变得更加方便快捷,国内外的期刊和图书的引文中,已经有不少以互联网网址出现的网络引文,并且比重呈现逐年增长的态势。Robert P. Dellavalle 等[1]通过对 *Science* 等影响因子较高的三种杂志的文献加以分析,发现30%的文献中至少含有1篇网络引文,所有引文中有将近2.6%的网络文献。也许我们曾经遇到过这样的困惑:几天前访问的网页,再次打开的时候竟然提示"找不到网页"、"禁止访问"等,本来十分重要的信息一夜之间就可能不复存在。据 Alexa 的调查数据,网页平均75天后就会消失[2]。此外,网页内容的频繁变动也导致了读者无法回溯到文献作者引用时的网页状态。由于网络引文缺乏持久性和稳定性,正式的学术交流中对它的引用仍然存在争议[3]。基于此,网络信息资源的可追溯性问题被提上日程,对网络引文的认可程度及其对传统的引文分析方法产生的影响[4-6]等已成为国内外研究的热点。如何解决期刊网络引文不可追溯问题,也显得尤为重要。

2 国内外研究综述

2.1 不可追溯现象的研究

国外关于"网络引文不可追溯"现象的研

* 本文为国家自然科学基金重点项目"基于生命周期理论的数字信息资源深度开发与管理机制研究"(项目编号:70833005)成果之一。

究,用 Google Scholar[7] 检索到被引次数最多的 3 篇文章是 *JASIST* 上的 An Analysis of Web Page and Web Site Constancy and Permanence 和 Web Page Change and Persistence—A Four-Year Longitudinal Study,以及在 *Computing Practice* 上的 Persistence of Web References in Scientific Research。这 3 篇文章基本上概括了国外目前研究的现状。前两篇文章对随机选取的 361 条 URL 进行了跟踪,对网页和网站的持久性和恒定性加以分析,通过前后 3 次(以 6 个月为间隔)的数据分析发现:6 个月之后,12.2% 的网站及 20.5% 的网页不稳定,97% 的网站和 98.3% 的网页内容会加以变化;1 年之后,数值分别增加到 17.7%、31.8% 以及 99%、99.1%[8-9]。第 3 篇文献从 CiteSeer 数据库抽取了 67577 条 URL,分析了网络引文增长规律和不可追溯性的基本现状,进一步从链接和内容的可靠性两个角度,提出了网络引文链接长期保存的解决方案[10]。此外,文献[11-15]也分别研究了各自相关领域期刊的不可追溯现象。

国内关于网络引文不可追溯现象的研究集中在学术期刊上的网络引文方面。文献[16]通过对《软件学报》和《中国图书馆学报》1999-2003 年间所刊载的 1589 篇含网络引文的文章进行分析,研究了可追溯率和网站类型、时间等因素的关系,探讨了不可追溯性的各种情况,建立了可追溯性的回归模型。文献[17]以中国学术期刊网上的情报学、情报工作专题中各个期刊的网络引文为研究对象,检测了 1997-2003 年的网络引文,说明了其可获取的百分比随时间变化的情况。研究表明,引用时间越远,网络引文可获取的百分比越低。

2.2 网络引文不可追溯原因分析

网络引文不可追溯的原因很多,主要分为如下两类[10,18]:

(1)链接的变动。链接的变动往往会导致读者根据提供的链接,无法获取作者引用的相关资源。这种情况产生的主要原因有:硬件的问题,如服务器关闭、网络故障等;原有链接的更新,如链接被删除、修改等;访问网络环境的限制,不同 IP 地址的计算机访问某个网络信息资源时,有可能得到不同的结果;此外,作者的网址链接拼写错误也会导致网页的无法访问。

(2)内容的变动。网页内容的变动主要表现在网络信息资源内容的动态更新,以及网页排版结构的变化等。此外,网络引文的不规范使用,如引而不用、用而不引等也往往会导致网页内容的不可追溯。

2.3 网络引文不可追溯的解决方案

为了提高网络信息资源的可追溯性,已有一些国外学者和机构做了相关研究。网络引文通常利用 URL 进行标识,而 URL 所代表的仅仅是资源和位置的映射关联,当位置发生移动时,这样的映射也自然消失了。因此,IETF 于 1993 年提出了一项统一资源命名计划——统一资源名称法[18](Uniform Resource Names,简称 URNs),对每一个数字化信息赋予一个永久的、唯一的且独立于信息资源的存储位置名称,通过这个名称就可以映射到该电子资源上。典型的 URNs 项目有 IDF(The International DOI Foundation)开展的数字式对象标识符 DOI[19]、OCLC 的 PURL[20](Persistent URL)项目等。此外,M. L. Creech[21] 提出了一种面向作者的链接管理(CLT/WW)技术,通过记录某站点的操作日志,自动识别和修复该站点下的死链接。

上述方案对于解决网络资源位置的改变有一定的帮助,但对于网络资源内容的变动(修改和删除)却力不从心。Joachim Feise[22] 提出了利用配置管理系统(Configuration Management System),收集、存储和组织历史网页资源,用户可以通过代理服务器向系统发送获取特定时间下的网页资源的请求,系统将返回请求的内容并呈现给用户。美国的 Internet 档案馆[23],通过自动抓取或接受 Alexa 公司和其他机构捐赠的数据,存储了将近 850 亿个网页存档,当遇到"找不到网页"的错误时,还可以检索到历史网页。类似的还有北京大学天网实验室开发的关于中文历史网页信息的存储与展示系统 Web Infomall[24]。此外,国外的 WebCite 系统[25] 专门就解决网络引文的保存问题提出了解决方案。

在该系统中,每一个 WebCite 引文都是一个存储在 WebCite 上存档的网络引文,它不是直接链接到作者访问的网页,而是在作者访问的时候将文档存储起来(相当于备份),当读者需要查看作者的引文时,其实际访问的是 WebCite 上的备份内容。实际上,该网站在一定程度上充当了可靠的第三方机构的角色。

从国内外的研究论文中可以发现,原生数字信息资源的不可追溯现象已经十分普遍,即使有的网址能重现网络引文,但其版面布局和内容的频繁变化亦会导致原有网址下获取的网络引文并非作者真正想要的引文信息,这势必就失去了引证文献标著的意义,严重影响着人们使用和引证网络信息资源的积极性。此外,网络引文可追溯性研究也大部分集中在可追溯性现象及规律的研究上,对不可追溯问题解决方案的研究偏少,与此相关的实际应用系统更少。Internet Archive、Web Infomall 等项目在对良莠不齐的网络信息资源进行采集和存储时未加以过滤,这种全采集模式缺乏针对性,导致采集过程的大部分时间花费在重复或无价值的信息上。WebCite 通过和各期刊杂志社的合作,存储了期刊文献中的所有网络引文,但是其对于未在系统中存储的网络引文的追溯性支持不够,同时限于其规模及易用性,WebCite 实际上并未得到广泛的使用。

3 解决方案及系统构建

通过上文的分析,笔者认为网络引文追溯平台的构建应该着重解决网络信息资源位置和内容的"变动性",最有效的解决方案是采用类似于 WebCite 的功能,在作者引用行为产生时及时将网络引文的内容存储下来,构建网络引文库和网络引文检索平台,使得读者能够准确追溯并原貌呈现出网络引文的"引用版本"。需要注意的是,该平台存储的网络资源通常是即将审核和出版的论文文献中包含的网络引文("现刊"引文),而对于另外的未能及时存储的网络引文(一般为"过刊"引文)的追溯问题,网络引文追溯平台也应该加以解决。基于此,笔者提出了新的网络引文追溯平台解决方案,下文将就该平台的功能需求、模块划分和实现做详细介绍。

3.1 解决方案与思路

为解决现有系统的问题,本文构建的网络引文追溯平台将网络引文划分为"现刊"引文和"过刊"引文两部分。对于"现刊"引文,系统将为在线期刊编辑出版系统提供接口,对网络引文数据进行获取、识别和存储,彻底解决网络引文的不可追溯性问题。对于"过刊"引文,通常的研究认为网址无法访问是不可追溯的标志,但国内外开展的一系列项目(Internet Archive、Google Cache 等)采集和保存了大量的历史网页,在一定程度上能够重新定位无法访问的网络信息资源。据相关研究[1],在 60 个不可直接访问的网络引文中有 31 个能在 Internet Archive 中找到,另有 2 个可在 Google Cache 中找到。因此,本系统将提供 Internet Archive、Web Infomall 和 Google 等的接口,并通过整合得到的数据,实现最大程度的网络引文追溯。

值得注意的是,网络资源内容的变动往往会导致实际获取的网络信息资源并非同作者引用时的内容完全相同,即使追溯获取网络资源以前的版本,也无法确认是否为引用行为产生时的版本,因此对于某网络引文的追溯应考虑到 URL 和访问时间的结合。但另一方面,尽管网络引文在不同时间的内容可能不同,为了避免冗余存储同一个网址下不同时刻的网络资源,系统还需要自动检测冗余的内容,及时剔除和规整相关引文信息。

考虑到浏览器的设置不同,IP 地址的限定、访问权限的差异等,不同用户在访问同一个页面时也可能产生不同的结果,系统将以服务器所在的网络环境为准存储网络引文。网络引文主要有两类,即普通的网页文件(包括. html、. htm等为后缀名的静态页面和. asp、. jsp 等为后缀名的动态页面)和以 HTTP 协议传输的文档文件(包括以. pdf、. doc 等后缀名结尾的文件)。鉴于 PDF 文档在数字资源长期保存以及可信度确认等方面的强大支持力度[26],系统自动将以

上两类网络引文转换生成 PDF 文档,并加以存储和标引。

3.2 系统的模块划分

基于上述分析,笔者认为,网络引文追溯平台主要可划分为两大模块,即网络引文库构建模块(图 1)和网络引文集成检索模块(图 2)。

3.2.1 网络引文库构建模块

网络引文库构建模块将结合在线编辑出版系统(如各期刊杂志社的在线投稿平台),构建“网络引文引用”功能的插件,规范作者和审稿编辑的引文引用行为,在论文作者引用行为产生时,通过调用该插件,就可以及时将网络引文保存下来,并加以序化规整,形成网络引文库,期刊编辑和出版商在论文审核的过程中,通过调用该插件亦可发现网络引文库中未及时存储的引文数据,并及时反馈给用户,存储和更新网络引文的获取链接。网络引文库构建模块将分为两部分,即作者参与下的网络引文存储模块以及编辑、出版商参与下的网络引文检测模块(见图 1)。

网络引文存储模块通过构建用户接口,获取论文作者提交的网络引文数据,并使用网页采集器依次遍历采集每条网络链接地址映射的相关资源,存储、规整网络引文数据,并以服务器时间为准,为存储的网络资源加盖时间戳,对无法直接访问的网络资源,将反馈给作者。规整和存储子模块需要检测内容的重复性,为了防止冗余存储,对于已存储的内容相同的网络资源,只需要生成新的映射关联,此模块将返回存储后的永久性网络引文获取路径等相关信息。

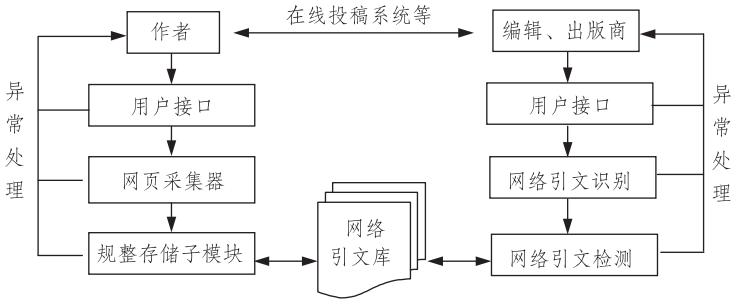

图 1 网络引文库构建模块流程图

网络引文检测模块将通过用户接口获取在线投稿系统中论文的引文信息,使用事先定义的特定引文著录规范,自动识别引文数据中的网络引文信息,发现其中的网络引文获取链接地址,通过调用网页检测器,检测用户提交的网络引文是否可以追溯。不可追溯的网络引文将及时通知论文作者,以确保网络引文的可追溯性。

3.2.2 网络引文集成检索模块

网络引文检索平台是读者追溯网络引文的主要途径,它将构建获取网络引文地址和引文访问时间的用户接口,并顺序实现网络引文库、Internet Archive、Google 等的检索接口,试图返回读者需要的特定时间和特定网址下的网络引文。系统的基本流程如图 2 所示。

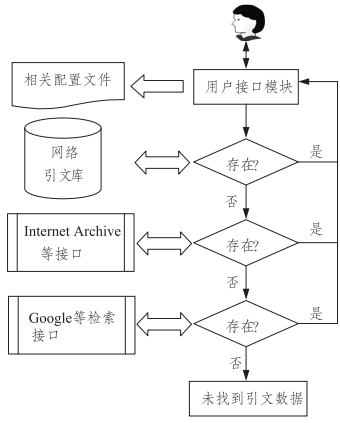

图 2 网络引文集成检索模块流程图

系统通过用户接口模块获取读者请求的网址和时间,在网络引文库中进行检索,若未找到匹配的引文信息,则调用 Internet Archive 等网页保存系统的接口;若还未找到,则通过对网络引文的分析,调用 Google 检索接口,寻找获取该网络引文的线索;若能匹配到网络引文,则将检索结果返回用户接口模块,并对未存储的网络引文加以存储,以实现最大程度的网络引文追溯。

3.3 系统实现和评价

3.3.1 网络引文库构建模块的实现

网络引文存储模块:该模块一方面通过构建用户接口获取网络引文数据,另一方面负责采集、读取和存储网络引文的内容。为了避免在编码之间相互转换的过程中遇到的乱码问题,系统将以 Java 字节流的形式读取文件,并调用开源项目 HttpClient[27]。在文件存储的过程中,系统将调用 HtmlToPdf[28] 和 IECapt[29] 开源软件将文件转换为 PDF 文档和网页图片快照文件(JPEG 格式)。

网络引文在下载存储之后,还需要按照特定序化的格式将引文内容加以标引、规整和内容冗余检测,将网络引文的标题、关键词、存储时间(服务器当前时间)、URL、网页源码等相关信息存入数据库,进而方便读者的追溯行为。本系统使用了 Mysql 数据库。

网络引文检测模块:网络引文库构建模块的实现需要在线编辑出版系统(编辑或出版商)提供引文数据,并根据定义的引文著录标准,对网络引文加以分析和检测,判断作者提交的数据是否已经存储和可获取。由于不同标准的参考文献标引的格式不尽相同,网络引文自动分析方法也有所差异,本系统目前实现的处理网络引文标准是 GB/T 7714 – 2005,该标准的基本形式如下:

主要责任者. 题名:其他题名信息[文献类型标志/文献载体标志]. 出版地:出版者,出版年(更新或修改日期)[引用日期]. 获取和访问路径。

在该标准下,系统需要析出引文获取方式中含有[OL]的引文作为网络引文,调用 HttpClient 等开源软件,自动分析和判别网络引文的可获取性,并将结果返回给作者,杜绝不可获取的网络引文。

3.3.2 网络引文检索平台的实现

网络引文检索平台为读者提供了追溯特定时间和特定网址下网络引文的途径:系统将首先通过动态构建 SQL 语句,实现在网络引文库中的数据库检索,若未匹配到合适的网络引文,则调用 Internet Archive 和 Google 的检索接口,寻找合适的网络引文信息。关于 Google、Internet Archive 和 Web Infomall 等检索接口的实现,可以参见文献[30]。该平台的界面如图 3 所示,用户可以通过提交 URL 和 URL 存储的时间来追溯网络引文,亦可以通过提交网页相关的描述信息,匹配数据库和各个接口的网页信息。通过点击各个选项卡,就可以切换到不同的接口下查看相关信息,用户还可以点击时间下拉菜单查看存储的不同版本的网络引文。

图 3 网络引文检索平台的 Internet Archive 接口实现

3.3.3 网络引文检索平台的评价

由于“现刊”引文存储了所有使用“网络引文引用”插件的期刊杂志的网络引文,对于该类引文将彻底解决不可追溯的问题,因此评价的过程主要针对“过刊”网络引文的追溯性。此外,由于网络引文著录的不规范性,很多网络引文并未给出作者访问时间,因此本次数据的采集和检测也尚未考虑到同样 URL 在不同存储时间上内容的差异。

笔者选择万方数据库作为链接获取的数据源,自动抓取了 2001 – 2006 年发表在《情报学报》、《软件学报》和《系统工程理论与实践》上的 3987 篇文献(剔除万方数据库没有收录的刊期论文)的 4395 条链接,使用 Xenu[31] 链接检测软件进行检测,对各条网络引文进行循环检索,处理结果见表 1。

表 1 三大期刊网络引文可追溯性检测结果(2001 – 2006)

各阶段的结果(单位:条)	软件学报	情报学报	系统工程理论与实践
文章总数	1561	701	1725
网络引文总数(T)	2730	1432	233
可直接访问引文数(S)	1946	892	148
本平台增加可访问引文数(R)	393	218	35
稳定性:(S/T)	71.3%	62.3%	63.5%
弱稳定性:((S + R)/T)	85.7%	77.5%	78.5%
不稳定性:((T – R – S)/T)	14.3%	22.5%	21.5%

从表 1 数据看出,Internet Archive、Web Infomall 等接口在一定程度上提高了网络引文的不可追溯性,平均提高 15%(R/T),但仍有 20% 左右的网页不可追溯。

3.4 系统的不足之处

由于网络环境的不同,许多网络引文信息在不同的 IP 地址下的访问权限也有所不同,尤其是对于有些收费的数据库资源而言。网络信息在保存的过程中,还需要考虑到权限屏蔽和知识产权保护的问题,而在本系统中尚未考虑。

Web2.0 的进一步发展使得用户的参与性愈来愈强,同传统的纸质和电子期刊的发文流程大相径庭,由于 blog 等 Web2.0 网站上的发文情况并没有固定的审核流程,使得互联网上充斥着大量表现“个人观点”的文章,这样的文献能否成为网络引文也是值得思考的问题。笔者 2008 年 9 月 28 日在中国期刊网上进行检索发现,“参考文献”域中含有“blog. sina. com. cn”的文献记录已经有 1217 条,而 2005 年仅有 1 条,2006 年有 82 条,2007 年骤增到 847 条,如何面对新类型的网络引文也是系统需要考虑的问题。

此外,网络资源的引用行为不仅表现在期刊杂志的论文上,互联网上的“转载”(或者直接的复制、粘贴)行为也十分普遍。目前,blog、网站后台管理系统等大量使用了网络在线编辑器,本系统网络引文库的构建过程还需要逐步深化以实现“网络引文引用”插件的构建,以更好地移植到在线编辑器平台上,从更广泛的范围上解决网络资源的不可追溯现象。

4 结论

互联网的发展使得信息资源的存取更加便捷,引文的不可追溯,尤其是网络引文的不可追溯性便凸显出来并被赋予了新的特征。针对网络引文信息的不可追溯问题,本文设计并实现了网络引文的追溯平台,论文作者能够通过该平台存储网络引文引用行为产生时刻的引文内容,构建网络引文库;论文的编审和出版商可以通过该平台检测网络引文的有效性,并及时反

馈给作者;论文读者亦可以通过该平台100%检索已存储的网络引文。对于未能及时存储的网络引文(如已发表的论文文献中的网络引文),本系统构建了Internet Archive等的接口,实现了最大限度追溯呈现网络引文的需求。当然,网络引文追溯平台尚存在着诸多不足,笔者将正视这些不足,积极寻找补充的解决方案。

参考文献:

[1] Robert P. Dellavalle, Eric J. Hester, Lauren F. Heilig, etc. Going, Going, Gone: Lost Internet References [J]. Science, 2003, 302(5646): 787-788.

[2] Brewster Kahle. Preserving the Internet [J]. Science American, 1997,276(33), 82-83.

[3] Liwen Vaughan, Debora Shaw. Bibliographic and Web Citations: What Is the Difference? [J]. Journal of The American Society For Information Science And Technology, 2003, 54(14):1313-1322.

[4] Liwen Vaughan, Debora Shaw. Can Web Citations Be a Measure of Impact An Investigation of Journals in the Life Sciences[C]//Proceedings of the 67th ASIS&T Annual Meeting, 2004(41): 516-526.

[5] Blaise Cronin. Bibliometrics and beyond: some thoughts on web-based citation analysis[J]. Journal of Information Science, 2001, 27(1):1-7.

[6] Liwen Vaughan, Debora Shaw. Web Citation Data for Impact Assessment: A Comparison of Four Science Disciplines[J]. Journal of the American Society for information science and Technology, 2005, 56(10):1075-1087.

[7] Google. Google Schloar [OL]. [2007-05-21]. http://scholar.google.com/.

[8] Wallace Koehler. An Analysis of Web Page and Web Site Constancy and Permanence [J]. Journal of The American Society For Information Science And Technology, 1999, 50(2):162-180.

[9] Wallace Koehler. Web Page change and persisitence: a four-year longitudinal study [J]. Journal of The American Society For Information Science And Technology, 2002, 53(2): 162-171.

[10] Steve Lawrence. Persistence of Web References in Scientific Research [J]. Computing Practices, 2001,34(2):26-31.

[11] Jonathan D. Wren. 404 Not Found: The Stability and Persistence of URLs Published in MEDLINE [M]. Bioinformatics,2004, 20(5): 668-672.

[12] Emily Olfson, Jeffrey Laurence, etc. Accessibility and longevity of Internet citations in a clinical AIDS journal [J]. AIDS Patient Care STDs, 2005, 19(1):5-8.

[13] Matthew E. Falagasa, Efthymia A. Karveli, Vassiliki I. Tritsaroli. The risk of using the Internet as reference resource: A comparative study [J]. International journal of medical Informatics, 2008, 77(4): 280-286.

[14] Frank McCown, Sheffan Chan, Michael L. Nelson, etc. The Availability and Persistence of Web References in D-Lib Magazine[C]// 5th International Web Archiving Workshop (IWAW05), Vienna, Austria, 2005, eprint arXiv:cs/0511077.

[15] Diomidis Spinellis. The decay and failures of Web References [J]. Communications of the ACM, 2003, 46(1):71-77.

[16] 吴志强. 我国学术期刊上的网络参考文献可追溯性考察[J]. 情报学报,2006, 25(1):80-86.

[17] 胡德华,方平,刘双阳,等. 网络参考文献的可接受性、选择性和可获取性研究[J]. 情报学报,2006, 25(2):179-183.

[18] Uniform Resource Name [OL]. [2008-10-04]. http://en.wikipedia.org/wiki/Uniform_Resource_Name.

[19] PURL [OL]. [2008-10-04]. http://purl.org/.

[20] The Digital Object Identifier System [OL]. [2008-10-04]. http://www.doi.org/.

[21] M. L. Creech. Author-oriented link management [J]. Computer Networks and ISDN Systems, 1996, 28(7):1015-25.

[22] Joachim Feise. An Approach to Persistence of Web Resources[C]//ACM HT'01, Aarhus, Denmark, 2001:215-216.

[23] Internet Archive. Waybackmachine homepage [OL]. [2008-10-04]. http://web.archive.org/.

(下转第118页)

力量的功能,在当今贫富差距加剧所引发的社会对立、分化、冲突中,可以发挥促进理解、促进凝聚的作用,成为我国和谐社会发展至关重要的聚合力量。

参考文献:

[1] 曲晓玮. 我国图书馆社会捐助的缺失分析[J]. 图书馆论坛,2005(12):144-146.

[2] 中华人民共和国民政部. 慈善:大有作为的社会公益事业——我国慈善事业30年发展辉煌成就综述[EB/OL]. [2009-01-04]. http://cbzs.mca.gov.cn/article/shxw/yw/200812/20081200023199.shtml.

[3] 刘喻. 美国私人基金会捐赠高等教育的研究[D]. 武汉:华中师范大学,2008.

[4] 于良芝. 公共图书馆服务体系研究[J]. 中国图书馆学报,2008(2):79-80.

[5] 梁幸枝,邢婷. 人情的选择,还是制度的依赖——中外社会信任机制的概况研究[J]. 社会,2003(5):49-54.

[6] 杨岭雪. 美国图书馆基金会的类型与运作[J]. 图书馆杂志,2005(5):69-71,64.

[7] 中华人民共和国民政部. 2007年民政事业发展统计报告[R/OL]. [2009-01-04]. http://cws.mca.gov.cn/accessory/200806/1214811949213.doc.

[8] 江希和. 税收激励慈善行为的政策原则分析[J]. 生产力研究,2007(5):33-34,49.

[9] [美]米切尔·拉伯夫. 世界上最伟大的管理原则[M]. 徐海波,编译. 北京:科学技术文献出版社,1989:10.

[10] 孙慧明. 关于鼓励社会力量参与图书馆建设的两个提案[J]. 图书与情报,2008(2):34-36.

张秀梅 中山大学图书馆馆员。通讯地址:广州市中山二路74号中山大学医学图书馆。邮编510089。

(收稿日期:2009-01-08;
最后修回日期:2009-02-25)

(上接第105页)

[24] 北京大学网络实验室. Web InfoMall [OL]. [2007-08-29]. http://www.infomall.cn/.

[25] WebCite. How do I use WebCite [OL]. [2008-10-04]. http://www.webcitation.org/.

[26] 刘家真. 文件保存格式与PDF文档[J]. 档案学研究,2002(2):46-51.

[27] http client [OL]. [2008-10-04]. http://hc.apache.org/httpclient-3.x/.

[28] Java Html Pdf[OL]. [2008-10-04]. http://java-html-pdf.qarchive.org/.

[29] IECapt-A Internet Explorer Web Page Rendering Capture Utility[OL]. [2008-10-04]. http://iecapt.sourceforge.net/.

[30] 陆伟等. 信息检索实验[M]. 武汉:武汉大学出版社,2008:33-42.

[31] Xenu. Find broken links on your site with Xenu's Link Sleuth (TM) [OL]. [2007-08-29]. http://home.snafu.de/tilman/xenulink.html.

陆 伟 博士,武汉大学信息资源研究中心副教授。通讯地址:武汉珞珈山。邮编430072。

韩曙光 武汉大学信息资源研究中心硕士研究生。通讯地址同上。

沈祥兴 武汉大学信息资源研究中心教授,图书馆研究所所长。通讯地址同上。

(收稿日期:2009-01-05)