

科研机构对机构知识仓储认知和服务需求调查分析^{*}

韩珂 祝忠明

(中国科学院国家科学图书馆兰州分馆 兰州 730000)

【摘要】采用调查问卷方式,对中国科学院科研院所内科研人员对机构知识仓储的“认知情况和对 IR 系统所提供的需求”进行调查,对回收的数据作统计分析和归纳总结。指出此次调查存在的不足,部署下一步工作。

【关键词】机构知识仓储 问卷调查 认知 服务需求

【分类号】G250

Survey of Cognition and Requirement of Service on Institutional Repository in Academy

Han Ke Zhu Zhongming

(The Lanzhou Branch of the National Science Library, Chinese Academy of Sciences, Lanzhou 730000, China)

【Abstract】Based on the questionnaire of “cognition and requirement of IR” done by the researchers working in the Chinese Academy of Sciences, the authors make an analysis of actuality of the institutional repository in academy. Then the authors point out the deficiency and plan in the next step of work.

【Keywords】Institutional repository Questionnaire survey Cognition Requirement of service

1 前言

机构知识仓储 (Institutional Repository, IR) 是数字时代学术研究的基础设施,一个基于大学、科研机构的 IR 是指为管理和传播该机构及其成员所创造的数字资源而提供的一系列服务,这些服务包括数字资源的长期保存、组织、检索或发布等^[1]。通过构建和使用 IR 系统,机构可以实现数字资产的管理,重用已有科研成果,并实现对已有成果的长期保存,从而进一步提升机构的社会形象,扩大学术影响力。

目前,国外研究人员已经在 IR 的相关领域开展了大量而系统的研究工作,其中包括对实施 IR 的可行性以及 IR 的构建和运行情况的调研。2002年, Ibitonke Lawal 在美国和加拿大的 9 种学科中随机抽取 24 万名科研人员,调查其使用电子预印本系统的情况^[2]。2005年,加利福尼亚大学的 John L. Ober 等人对本校的后印本服务环境和可行性进行了调研^[3]。同年 5 月, CN I JISC 和 SURF 等组织联合对澳大利亚、加拿大、美国、英国、荷兰等 13 个国家机构知识仓储的发展情况作了详细的统计分析^[4]。2006年,澳大利亚的 APSR 项目组开始对国内负责和实施数字化科研管理系统的管理者进行采访,归纳总结出在今后的 5 - 10 年中,仓储服务存在的关键问题^[5]。2007年 2 月,美国密歇根州大学资助的 MIRACLE 项目发表了关于美国全国范围内高等教育 IR 的构建和实施情况的报告^[6]。

而我国对 IR 的研究仅限于对 IR 基本理论知识、实践的引进和介绍,鲜有针对本土实际情况开展的实证研

收稿日期: 2007 - 11 - 02

收修改稿日期: 2007 - 11 - 08

* 本文系国家自然科学基金项目“中国西部环境与生态科学数据中心”(项目编号: 90502010)的研究成果之一。

究。本研究以中国科学院部分科研院所的科研人员(包括科研工作者、科研管理人员、学生、图书和情报工作人员)为研究样本,调查科研人员对 IR 认知程度和对 IR 所提供的服务需求情况。希望通过本次调查研究,获得有关 IR 的认知和需求的第一手资料,为寻找适合国内科研环境的 IR 建设模式研究提供支撑,对 IR 服务模式的规划和构建工作有所借鉴。

2 调查问卷说明

本研究主要采用了调查问卷的研究方法。

2.1 选择样本

在样本选择上,采用分层抽样调查方法。选取中国科学院部分科研院所的科研工作者、科研管理人员、图书情报工作人员和学生(博士/硕士研究生)。然后根据用户群体的人数比例确定被调查人数,最后以电子邮件的方式发放调查问卷。考虑到地域文化之间的差异,笔者分别在东部、中部、西北、西南、华南地区挑选若干科研院所,然后依照不同的专业领域的划分,最终选定 18 个院所作为调查对象。

在各类人员的选择上,首先,考虑到一线科研工作者人数在全院总人数中所占比例较大,由他们所创造的科研成果在全院总的科研产出中所占比重也最大,科研工作者也将是 IR 最主要的目标用户群体,所以将调查人数初步设定为总人数的 1/2;其次,是科研管理人员等决策人员如果能够支持 IR 的建设,那么在 IR 系统构建过程中所需要的各种政策环境、设备条件、人员等都可以得到满足,从某种程度上说,管理层是决定是否能够启动 IR 系统计划的先决条件,所以将调查人数比例设定为 20%;对于学生(包括博士/硕士),自中国科学院实施研究生教育改革以来,学生数量在逐年增加,而且许多科研工作都有学生参与,他们所创造的成果在中国科学院科研产出中占据相当的比例,故选定人数比例是 15%;最后一部分是图书馆/情报中心人员,虽然这些工作人员占科学院总人数的比例较小,就目前的 IR 建设模式来看,IR 的构建、维护和管理的大部分工作是依托于各个图书馆/情报中心,图书馆/情报中心人员是 IR 的主要实施和推动者,故本次调查研究设定比例为总人数的 15%。本次选取的调查对象具有相当的代表性,他们对 IR 的认知程度和服务需求一定程度上代表了中国科学院科研人员的观点。

2.2 设计调查问卷

在内容上,此次调查问卷主要从两个方面着手设置了 16 个问题:

(1)对 IR 认知度和意愿的调研:了解被调查对象对 IR 的认识和了解程度;科研机构对机构内部科研产出的管理方式;对学术产出、科学数据的习惯性发表(布)渠道;对 IR 的需求程度。

(2)对 IR 服务需求的调研:了解用户对 IR 具体实施过程中的内容提交模式、提交类型、使用权限、质量控制和知识产权等问题上的意见和要求。

在形式上,本次调查问卷共设 16 个封闭式问题,单项选择和多项选择同时使用。在问卷设计过程中,征求有关专家的意见,并作了多次的调整和修改。对某些问题笔者又进行了二次回访和深度访谈。

2.3 发放与回收调查问卷

问卷确定之后,笔者向选定的对象发放电子邮件。问卷从发放到回收过程历时 1 个月(从 2007 年 9 月 2 日到 2007 年 10 月 1 日),共发放问卷 200 份,回收且有效问卷 135 份,有效回收率为 67.5%。对部分被调查者实施回访时间约为半个月(从 2007 年 10 月 3 日到 2007 年 10 月 16 日)。

3 调查问卷数据分析

3.1 对 IR 的认知和意愿的分析

(1)对机构知识库的认识和了解程度

被调查对象中有 53.3% 的用户选择了“从未听说”;33.3% 的用户选择了“不常使用”;10.6% 的用户选择了“从未使用”;而选择“经常使用”的用户仅占 2.6%,如图 1 所示。

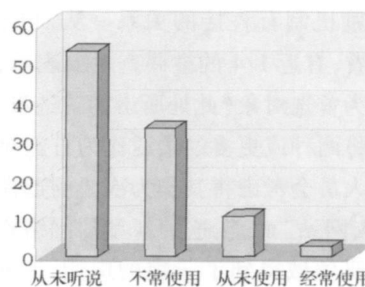


图 1 机构知识库的认识和使用情况

数据显示:将近一半以上的被调查对象“从未听说过机构知识库”,说明这个名词对于国内大多数的机

构来说还是很陌生的;听说但是“不经常使用”的被调查对象也占了相当的比重,说明机构知识库并不是他们经常选用的信息存储和检索、查询的途径;对于“从未使用”的被调查对象来说,一种可能是机构内没有提供类似的服务,另一种可能是不愿意使用机构知识库;而极少数“经常使用”的用户,机构知识库已经得到了一少部分科研人员的认同,并且已经使用。

(2)类似服务系统

从各个科研院所所拥有的数字保存系统和论文提交系统来看,有类似系统的机构占调查总数的61.1%,没有类似系统的占总数的38.9%。调查数据显示,科学院中有相当一部分科研院所已经对机构内部人员的科研产出、科研数据等研究成果的保存和使用予以重视,并且有意识地构建了多种形式的保存系统和论文提交系统,有力地推动了新的科研环境中知识共享和信息交流。

(3)学术产出发布渠道

本次调查问卷,为被调查对象提供了传统信息发布渠道和新兴的信息发布渠道,然后依照使用频次分为4个等级。调查结果如表1所示:

表1 学术产出发布渠道

发布渠道	首选	次要	其他方式的补充	从未使用
印刷型期刊	99.5%	0.5%	0	0
电子期刊	26.3%	42.1%	0	31.6%
个人网站(含博客)	0	6%	18.7%	75.3%
机构知识库(IR)	0.6%	0.5%	1.6	97.3%
学科或专业领域仓储系统	0.8%	0.4%	2.1%	96.7%

从表1显示的数据不难看出,传统的“印刷型期刊”占据重要地位,绝大多数科研人员会将学术成果投到相关领域的学术期刊上,这与现行的学术交流体制以及学术激励机制有直接的关系。从“电子期刊”的发布情况来看,有近1/4的被调查对象表示,可以优先考虑将其作为首选对象(此处所讲的首选没有排他性,即可以同时将两种或更多的渠道作为首选对象),将近一半的科研人员会考虑将其作为次要的选择。

就“个人网站”而言,这种新型的网络资源共享方式已经在科研领域得到了一定的应用,有6%的被调查对象表示可以将个人网站作为次要选择的对象,占总数18.7%的调查对象将其作为补充方式,但是更多的研究人员没有使用过个人网站,或者并没有考虑将个人学术成果放在个人网站上。

而对IR和学科或专业领域仓储系统的调查数据近乎相同,绝大多数科研人员没用使用过IR,没有使用的原因可归为几类,近半数的科研人员从来没有听说过,还有一部分是因为机构没有提供相关提交系统。

(4)调查/实验数据的存放

考虑到存放调查和实验所获得的原始数据有多种不同方式,有时可以选择多个存放地点,所以本题设计的是多项选择。调查数据显示,所有用户都会选择存放在“个人电脑”中。在全部调查对象中,有75.6%的对象将“个人电脑”作为唯一的选择,而有18.1%的人员选择“数据中心”,另有6.3%的人员选择“个人网站”作为数据存放的辅助方式。

统计数据说明,所有的被调查对象将会在科研过程中把调查和实验得到的科学数据放在“个人电脑”上,大量的数据没有得到长期有效的保存和保护。而值得注意的是,已经有相当一部分被调查对象开始使用“数据中心”,这种由机构提供的统一存放方式,有助于长期保存科研数据。另有一少部分被调查对象选择“个人网站”,使得科研数据不再是“隐藏资产”,而是以公开的方式展示给其他相关人员,用以进行学术交流和共享。

(5) IR的使用意愿

被调查的用户中有87.7%的对象表示愿意向IR提交数字资源,其余的12.3%表示不愿意提交。此项数据表明,一方面,IR为科研机构提供收集、组织、存储和管理机构内部数字形式的科研产出的这种服务模式,能够被大多数科研人员、学生以及管理者接受;另一方面,如果IR管理者采用合适的提交政策和保存机制,那么IR在使用过程中不会受到太大的阻力。

3.2 服务需求

(1)提交模式

IR的内容提交需要进行必要的元数据描述,被调查的用户中,有93.7%表示可以接受这一要求,另有6.3%的调查对象不愿意接受这一要求。数据显示,在向IR提交数据资源时,进行简单的元数据描述工作不会成为大多数提交者的负担。

就具体的提交模式而言,51.2%的被调查对象愿意使用“简单快速提交”方式,46.8%被调查对象希望由“IR管理员提交”,仅有2%的被调查对象能够接受“自助提交”,如图2所示。

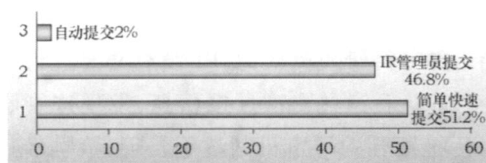


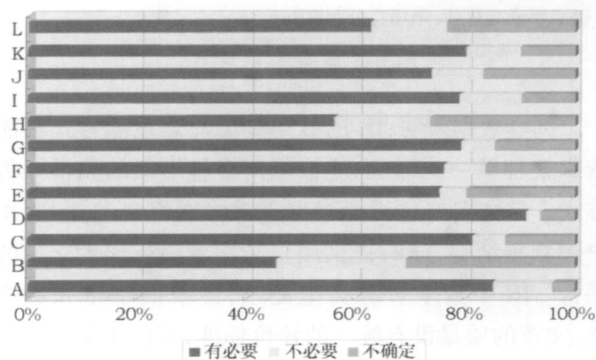
图 2 内容提交模式

从统计结果来看,“简单快速提交”是全部被调查对象最容易接受的提交方式,在提交电子资源的同时,提交者同时提供资源的简单元数据,又一次证明了 IR 采用自助提交的方式,并没有增加科研人员的工作量,科研人员不会因此而拒绝使用或少量使用 IR。

在统计过程中,笔者发现,不同人群对“IR 管理员提交”这种方式的接受程度有明显不同。对科研人员、科研管理者和学生而言,有 40% 以上的被调查对象选择该项,这符合“省力法则”;而 60% 以上的图书馆情报中心人员更倾向于此类提交方式,原因是由专业人员对信息进行组织加工,更便于管理和提高信息检索的精确度。

(2) 提交内容

科研机构的数字学术资源类型多样,这里笔者归纳总结了 12 种常见的资源类型,对于各种类型存放的“必要性”,被调查对象可以选择“有必要”、“不必要”和“不确定”。调查统计数据如图 3 所示:



(A: 经过同行评议的已经出版的文章; B: 预印本; C: 学术会议论文; D: 博士/硕士学位论文; E: 技术报告; F: 研究报告; G: 专著; H: 多媒体资料; I: 教学课件; J: 专利; K: 标准; L: 原始科研数据)

图 3 IR 提交内容类型

从统计数据总体来看,科研人员对使用 IR 来统一收集、组织、管理和使用数字学术产出持肯定的态度。除预印本之外的所有类型的学术产出都有半数的被访者认为有必要存放进去,尤其是对于已发表文章、博/

硕士论文、学术会议论文、技术报告、研究报告、专著、教学课件、专利和标准等。

被调查对象的存放态度的差异主要集中在“预印本”、“多媒体资料”和“原始科研数据”这三种类型的科研产出上。就“预印本”而言,被访者持否定态度的人数最多,经过回访,笔者发现,这些被调查对象不同意或者不确定的原因主要来自两个方面:

易引起版权纠纷,多数出版商不同意将未发表的文章以网络形式公开发布;

未公开发表的学术成果因为没有受到相应的法律保护,而被人滥用。

对于“多媒体资料”而言,科研人员认为 IR 主要是存放有价值的数字学术资源,而对于音频、视频等资料并非主要存放对象。对于“原始科研数据”,一直以来就是传统学术交流体制的盲区,缺乏适当的激励机制,而大多数重要科研数据都掌握在科研人员手中,对这类资源的保存和保护基本上处于无保障的状态,所以,IR 提供对原始科研数据的保护和管理,得到了相当一部分被调查者的赞许。但有些科研人员对这些科研数据的真实性、完整性、可信度以及获取价值表示怀疑,另外一部分则认为,科研数据的产权不明确,有被盗用的危险,所以对原始科研数据是否存放存在很大争议。

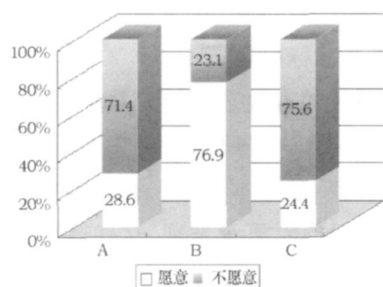
(3) 提交政策

IR 通常会制定内容提交政策来保证有效的数字资源采集工作。一般来说,会采用以下 3 种方式:“制定相关的政策强制存放”、“作者自愿存放”、“开始阶段强制存放”。第 1 种方式有 28.6% 的被调查对象表示愿意接受,其余的 71.4% 表示不接受;而对于作者自愿存放,则有 76.9% 的对象接受,其余 23.1% 表示不接受;第 3 种方式,有 75.6% 的不接受,而有 24.4% 的对象表示愿意接受,如图 4 所示。

统计数据显示:被调查对象对 IR 所提供的 3 种不同的提交模式的接受程度有明显差别。多数受访者认为,作者应该自主决定是否将其学术成果提交到 IR,倾向于学术自由的传播和交流;而对于强制存放措施,显然接受程度没有前者高,但有相当一部分科研人员认为可以接受这种方式,因为就目前的学术评价体制和奖励机制而言,就是采用了一定的强制政策。

(4) 权限设置

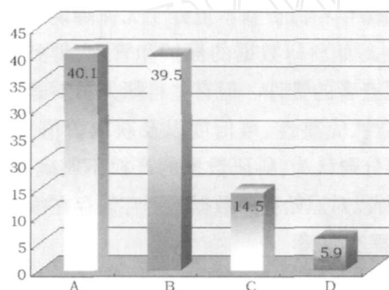
就 IR 的使用权限设置而言,40.1% 的被访对象选择对“部分机构开放获取”;39.5% 的被调查对象选择



(A:制定相关的政策强制存放;B:作者自愿存放;C:开始阶段强制存放)

图4 R的提交政策

“部分资源开放获取”;而选择“开放获取”和“仅限内部使用”的用户分别占14.5%和5.9%。也有相当一部分用户同时选择了几种权限设置方式,通过再次反馈意见,这些用户认为,不同类型资源的开放程度应该不同,如图5所示:



(A:部分机构开放获取;B:部分资源开放获取;C:开放获取;D:仅限内部使用)

图5 R的权限设置

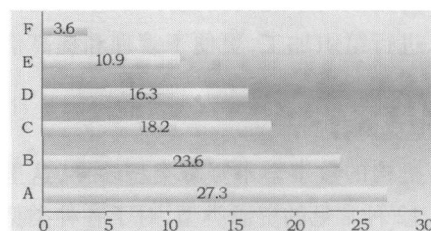
调查数据表明:R的数字化资源全部的开放获取或者限制范围在机构内部这两种模式是被调查者不愿接受的,如果资源仅限制在内部使用就大大降低了资源利用率,很难提高机构的学术地位和声誉;如果全面向互联网开放获取,则会出现类似资源质量参差不齐、知识产权等问题。绝大多数被调查者倾向于对部分机构开放获取,不同类型的资源采取不同的开放政策。

(5) 质量控制

对于R是否进行质量控制的问题,有64.3%的被调查者表示有必要对提交内容进行质量评审,而其他35.7%的被调查者认为没有必要进行质量控制。调查数据说明,多数被访者认为,数字资源的质量控制需要

通过其管理部门制定必要的质量控制政策来实现;也有相当一部分科研人员认为,R没有必要进行繁琐的质量控制。笔者通过再次反馈信息,发现这些研究人员认为传统期刊的评审机制过于繁琐,给学术信息的交流和传播带来延时,所以,R如果不采用方便、快捷的传播学术信息运行模式,那么其优势将无法体现。

而对于R的质量控制方法,此次问卷设计了6种常用的网站评选和学术评审办法,采用多项选择的形式。统计结果显示,“引用分析”和“发送E-mail与作者讨论”两种方式受欢迎程度最高,分别有27.3%和23.6%的被调查用户选择这两项;其次是“通过点击次数、下载次数进行数据统计”、“在文章后面在线发表评论”,选择对象所占比例分别是18.2%和16.3%;其余分别有10.9%和3.6%的被访者选择“向R提交评论文章”和“采用打分制度”,如图6所示:



(A:引用分析;B:发送E-mail与作者讨论;C:通过点击次数、下载次数进行数据统计;D:在文章后面在线发表评论;E:向R提交评论文章;F:采用打分制度)

图6 质量控制方式

调查数据表明:R的数字资源质量控制方式不能依靠某种单一的形式,应该采用多种形式的质量评价方式。具体实施过程要根据需求将网络自动评判、用户评判相结合,同时提供用户与作者进行学术交流的平台。R中数字资源类型复杂,涉及的学术领域广泛,文章的质量没有统一的评价标准,所以质量评价不宜于使用硬性的评判制度,如打分制度。

(6) 知识产权

对于R中数字资源的知识产权归属权问题,有82.3%的被调查者认为应该归“作者”本人所有,有16.4%的被调查者认为应归属于“机构”,而认为知识产权归属于“出版商或第三方机构”的仅占1.3%。调查数据显示,从用户的角度来看,大多数数字资源的提交者都希望拥有更多的对数字资源的支配权。

知识共享许可证 (Creative Common, CC) 是一种网络自由的许可授权机制。IR 的管理者会为用户起草两个许可协议: 一是机构与资源提供者之间的存储许可证; 另一个是资源提供者与资源利用者之间的知识共享协议。采用 CC 让用户获取信息更为方便, 减少了版权纠纷, 所以有 94.5% 的被调查者表示愿意接受这种内容授权协议。

4 研究结论和下一步工作设想

4.1 研究结论

科研人员对 IR 的认知度和需求度呈现出以下特点:

(1) 中国科学院的多数科研人员都从未听说或使用过 IR, 对 IR 的认识还是空白, 只有少量试点单位的研究人员对 IR 有初步的了解和认识。

(2) 中国科学院的科研院所已经开始关注数字资产管理 and 长期保存工作, 但是目前还没有系统而可行的解决方案和统一的管理模式。

(3) 中国科学院绝大多数科研人员对 IR 这种服务模式表现出较高的热情, 表示愿意使用 IR 来实现各种类型数字资源的全面管理和长期保存。

科研人员对 IR 所提供的服务需求主要集中在以下方面:

(1) IR 能够实现对科研机构多种类型的数字资产的长期保存、管理和使用等功能, 对不同类型的资源制定不同的管理政策。

(2) IR 的管理者应该制定较为宽松的内容提交政策, 尤其是对于那些容易引起版权纠纷的预印本文献, 不宜采用强制措施。

(3) 科研机构目前还不能接受完全开放的信息交流模式, IR 的管理者需制定灵活完善的访问控制策略 (权限设置), 根据科研院所的特定需求制定半开放式的开放政策。

(4) 科研人员赞同对提交到 IR 的数字资源进行规范的质量控制。针对 IR 的数字资源类型多样的特点, IR 应针对不同的资源定制灵活多样的质量控制方法。

(5) IR 的管理者应明确规定各种数字资源的版权归属, 与出版商、内容提交者和用户签署相关的协议书, 避免引起不必要的法律纠纷。

4.2 下一步工作设想

本次研究是首次在国内的科研机构对 IR 进行的

实证研究, 并获取了相关的一手资料, 所得到的研究数据和结果对于今后 IR 的规划、构建和管理能够起到一定的指导和借鉴作用。但是本次研究存在一定的局限, 主要体现在以下两个方面: 一方面, 选择的样本相对比较局限, 样本量相对较小, 结论不具有广泛的代表性; 另一方面, 本次调查仅从用户的角度考察了目前 IR 的发展情况。所以本次研究只是一个起点, 今后还要开展相关的后续研究, 主要从以下两点着手:

(1) 在本次调查的基础上, 扩展研究范围到高等院校。高等院校是重要的科研产出源, 每年会涌现出大量的学术论文、报告、教学资料等不同形式的研究成果, 是学术信息生产、交流的重要组成部分, 因此, 在高等院校构建和实施 IR 存在其合理性。研究方法主要采用问卷调查方法。

(2) 选择国内已经运行或正在计划构建 IR 的大学和科研机构进行深度访谈, 具体了解其具体政策制定、实施过程中遇到的实际问题。通过具体个案的分析和比较, 总结其运行的经验和教训, 为计划实施 IR 的机构提供参考。

参考文献:

- [1] Clifford A Lynch Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age [EB/OL]. [2007 - 08 - 29]. <http://dspace.uniroam2.it/dspace/bitstream/2108/261/1/ir.html>
- [2] Ibironke Lawal Scholarly Communication: The Use and Non - Use of E - Print Archives for the Dissemination of Scientific Information [EB/OL]. [2007 - 08 - 29]. <http://www.isdl.org/02-fall/article3.html>
- [3] John L Ober Postprint Repository Services: Context and Feasibility at the University of California [EB/OL]. [2007 - 08 - 29]. http://osc.universityofcalifornia.edu/responses/materials/UC_postprintstudy_final.pdf
- [4] Gerard van Westrienen, Clifford A Lynch Academic Institutional Repositories Deployment Status in 13 Nations as of Mid 2005 [EB/OL]. [2007 - 08 - 29]. <http://www.dlib.org/dlib/sepember05/westrienen/09westrienen.html>
- [5] Margaret Henty. Ten Major Issues in Providing a Repository Service in Australian Universities [EB/OL]. [2007 - 08 - 29]. <http://www.dlib.org/dlib/may07/henty/05henty.html>
- [6] Karen Markey, Soo Young Rieh, Beth St Jean. Census of Institutional Repositories in the United States M RACLE Project Research Findings Jean [EB/OL]. [2007 - 08 - 29]. <http://www.clir.org/pubs/abstract/pub140abst.html>

(作者 E-mail: kakahw@163.com)