

国内机构知识库发展现状与对策研究

赵迎光^{1,3} 曾 苏² 马建霞³

(1. 中国科学院研究生院 北京 100049; 2. 国核工程有限公司 上海 200233;

3. 中国科学院国家科学图书馆兰州分馆 兰州 730000)

摘 要 通过对国内高校和科研机构的知识库调查分析,发现尽管我国机构知识库建设已经起步,但仍存在数量少、资源不足、更新缓慢、开放获取和共享程度低等问题,针对这些问题提出了 IR 需要转变的三个方面,即 IR 发展应在资源征缴政策上寻找资源数量和质量的平衡点,在建设类型上从单一型向综合型转变,以及促进开放获取,提高资源共享程度。

关键词 机构知识库 调查分析 发展对策

中图分类号 G 250.7

文献标识码 A

文章编号 1002-1965(2011)01-0042-05

Research on Actualities and Countermeasures of Domestic IR

ZHAO Yingguang^{1,3} ZENG Su² MA Jianxia³

(1. Graduate University of Chinese Academy of Sciences, Beijing 100049;

2. State Nuclear Power Engineering Company, Shanghai 200233;

3. The Lanzhou Branch of the National Science Library, Chinese Academy of Sciences, Lanzhou 730000)

Abstract Analysis on IR in the domestic higher education and research institutes shows its development has started in China, meanwhile, problems of limited quantity of IR, limited content of resources and low level of Open Access need to be solved. The article put forward three proposals: the change of resources collection policy with more emphasis on balancing the quantity and quality of the available resources; adopting a comprehensive construction type for IR and improving the level of Open Access.

Key words Institutional Repositories; investigation and analysis; development countermeasure

0 引言

机构知识库 (IR, Institutional Repositories), 又名机构仓储、机构库、机构存储、机构典藏, 2002 年 SPARC (Scholarly Publishing for Academic Resources Coalition) 的一个研究报告^[1]中将 IR 定义为“收集和保存一个或多个大学知识产出的数字资源集合”。报告认为 IR 包含四层涵义: 机构性—不同于学科或专题知识库; 学术性—主要收藏学术资源; 累积性和长期保存—长久保留本机构的智力资产; 开放性和互操作性—保证机构内外的用户能方便存取资源, 支持不同 IR 的互操作。IR 中收录的资源类型可以包括期刊论文、学位论文、会议论文、演示文档、图书和章节、音频、视频、科学数据等多种知识产出。由于机构知识资产的类型多样, 知识产权复杂, 并不是所有的 IR 都具有

开放获取的特点。但 IR 建设的愿景不仅仅包括对机构知识资产的管理和保存, 也包括促进机构知识资产的传播和利用, 从而提高机构知识产出的影响力。

1 国内 IR 发展现状

据 OpenDOAR^[2]统计, 截止到 2010 年 3 月, OpenDOAR 中共收录了 1618 个 IR, 我国 27 个机构拥有自己的知识库, 而内地只有 3 个 (中国科学院国家科学图书馆, 中国西部环境与生态科学数据中心, 厦门大学)。从 OpenDOAR 的数据看, 内地开放获取的 IR 所占的比例还非常小, 但事实上, 在国内已有很多高校和科研单位建设了本校的学位论文库和教师文库, 它们也是以收集保存本机构知识产出为目的的, 但是由于这些数据库的元数据标准和数据开放性等问题, 未能在 OpenDOAR、OAIster、ROAR 等开放仓储中注册。

收稿日期: 2010-09-10 修回日期: 2010-10-11

作者简介: 赵迎光 (1986-), 男, 硕士研究生, 研究方向为数字图书馆; 曾 苏 (1986-), 男, 硕士, 研究方向为数字图书馆; 马建霞 (1972-), 女, 研究馆员, 硕士生导师, 研究方向为数字图书馆。

本文通过网站访问和文献调研方式, 在 2010 年 4 月 1 日至 2010 年 4 月 20 日期间, 对国内 112 所“211”工程院校、93 所中国科学院研究所的 R 建设情况进行调研, 重点关注文本形式知识产出的 R, 包括学位论文库和师生文库, 以及直接以“机构知识库”、“学术典藏库”命名的收录多种类型机构知识产出的 R (为便于表述, 本文中直接称其为“综合性 R”)。对于每个 R, 都从其发展概况, 内容征缴政策、软件平台和基本功能、开放获取和访问控制, 内容管理等方面进行了详细的调查统计。

1.1 国内 IR 发展概况

1.1.1 R 及其收录资源的类型。从 R 的类型来看, 学位论文库在高校中占有很大的比重 (84%), 而教师文库和综合性 IR 所占比例较小。从 R 所收录的资源来看, 学位论文库主要收录学位论文, 形式和内容都比较单一。而综合性 IR 所收录的资源种类较多, 图 1 以厦门大学和中科院力学研究所的综合性 IR 为例分析了其各种资源类型所占的比例。

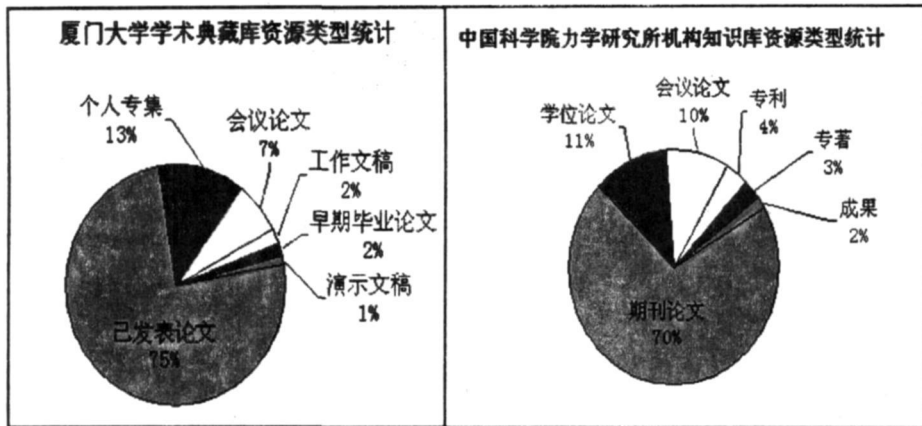


图 1 综合性 R 资源类型统计

从图 1 来看, 对于高校来说, 综合性 R 解决了高校中除学位论文外其他知识产出的保存和使用的问题。对于科研院所来说, 保存了本机构几乎所有的知识产出。同时也可看出, 综合性 IR 可以覆盖多种知识产出类型, 和内容单一的学位论文库相比有很多优势, 对于发布者来说, 可以集中组织管理和展示各种知识产出, 并可增加其发布资源的引用几率和关注程度; 对于使用者而言, 可以扩展其获取资源的途径, 加快知识的传播和利用。

1.1.2 资源数量。根据 R 类型, 将 R 分为学位论文库、科研院所综合性 IR 和高校综合性 IR 三类, 然后在每个类中按照各个 IR 的资源数量降序排列, 找出每个类中的四分位点所在的 IR 的资源数量, 如图 2 所示。

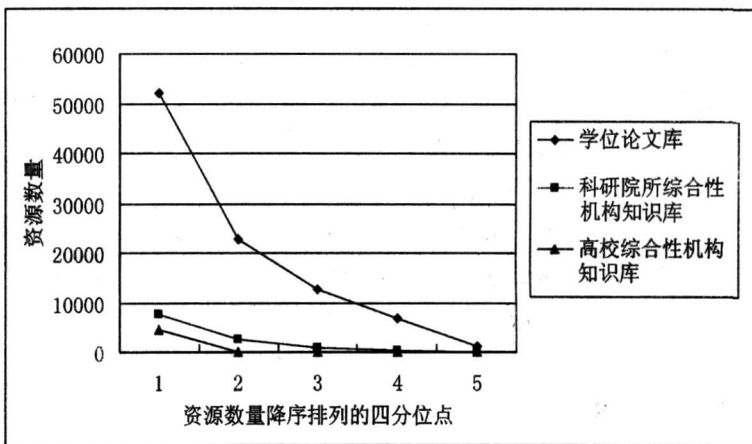


图 2 三类 IR 资源数量比较

从图 2 可以得出以下几个结论: a 学位论文库的资源数量多于综合性 R, 尽管学位论文库内容单一, 但是由于建库时间早, 资源收集制度成熟规范, 因此资源量明显多于综合性 R。b 综合性 IR 建设缓慢。在仅有的几个高校的综合性 R 中, 半数以上的 R 资源数量接近于零, 部分 IR 有库而无内容。这一方面是由于综合性 IR 的建设刚刚起步, 积累不够, 但同时也说明了 IR 的建设在规划和政策方面严重不足, 缺少有力的

资源征缴的政策机制和长期持续发展保障机制。c 高校和科研院所相比, 前者综合性 R 建设较为缓慢, 需要借鉴后者发展模式。不论是在综合性 R 的数量还是在资源数量上, 科研院所明显要多于高校。尽管科研院所刚刚起步, 但发展较快。因此高校在建设知识库过程中可以借鉴科研院所的发展模式

和经验, 结合自身存在的学位论文资源优势等情况, 找出一条适合自己的知识库发展模式。

科研院所和高校作为我国科研知识产出的生力军, 其科研知识产出的管理、保存、发布、利用是非常重要的, 尽管学位论文的组织管理相对容易, 但科研知识产出管理不能仅仅局限于学位论文, 这也说明, 还有其他的盲点有待于信息资源管理部门去开拓, 从而拓展和提升对教学科研人员的服务领域和水平。

1.1.3 资源更新。对于高校论文库而言, 论文库的更新主要是学位论文的提交, 提交时间集中在每年规定的论文提交时间, 更新的资源数量和本校硕博毕业生的数量相关, 更新频率相对稳定。

综合性 R 的资源更新则不同, 经过调查发现, 中科院科研院所更新频率最高, 尤其是力学研究所更新

频率 1274 5 条 / 月,除了与其建库时间较短和研究成果丰富有关外,一些政策的激励也使得科研人员积极存缴自己的科研成果。

对于高校而言,除了厦门大学以外,其他高校的综合性和 IR 使用频率相对较低,资源更新缓慢。厦门大学学术典藏库建库较早,资源增长稳定。在西班牙的赛博计量学实验室 (Cybermetrics Lab) 2010 年 1 月 10 日推出全世界开放获取 IR400 强排名^[3]中,厦门大学是大陆唯一上榜高校。

1.2 内容征缴政策 在所调查的高校和科研院所中,除 8 所高校的学位论文提交信息无法获取外,其余高校的学位论文征缴政策基本一致,即硕士和博士在提交学位论文后方可办理离校手续,这种强制性的规定使得学位论文库的资源得到保证。

在“厦门大学学术典藏库简介”^[4]中,对于内容的征缴政策主要是提倡全校的教学和科研工作者尽可能地把手头的有学术价值的数字资源都提交到学术库,并没有强制性的规定。

科研院所的知识库建设也是刚刚起步,但是部分院所的知识库内容征缴制度已经比较成熟,例如作为中国科学院 IR 示范的力学研究所 IR 规定^[5]: 所内作者的论文统一通过中科院力学所 IR 系统进行提交,作为年终考核绩效和晋职的依据,这一制度提高了机构成员主动提交资源的积极性。

从以上几个类型的 IR 的内容征缴政策中可以看出,当前我国 IR 在内容征缴政策的上存在一些不同,主要表现在以下两个方面。

a 存缴数量与政策强度。根据调查结果,可以按照政策强度将征缴政策分为三种层次: 强制存缴 (mandatory), 鼓励 / 激励存缴 (encouraged), 建议存缴 (may), 这三种层次的政策存缴数量与政策强度成正比。从图 5 中的资源数量也可以看出,学位论文库的激励机制最强,科研院所的综合性和 IR 次之,高校的综合性和 IR 最弱 (如厦门大学只是倡议), 说明知识库资源数量多少与激励机制的强弱存在正相关关系。

b 内容审核制度与资源质量。不同类型的知识库在内容质量要求上也存在差别。学位论文库中存储的只是学位论文,只要通过答辩的学位论文均可以存缴; 综合性 IR 中存储的还是以本机构已发表的论文、会议论文以及会议报告为主,往往经过了同行评议,因此对内容质量的要求较高,而对于演示文档、工作文件等类型,内容质量尚没有严格的审核制度。内容质量的要求不同,也造成了资源质量的不同,因此在进行知识库建设时如何寻找两者的平衡点也是需要考虑的问题。

1.3 软件平台及基本功能 高校的论文库系统主要以购买软件为主,集中于清华同方 TPI TRSD Paper

Apabi 数字资源平台和 CALIS - DL ETD 四种系统。在高校和科研院所的综合性和 IR 中,除了中国农业大学使用的中数创新公司 CDICM 数字图书馆应用系统^[6]以外,其余系统均基于开源的 DSpace 系统开发。

在所提供的接口方面, CALIS 的成员馆都实现了 OA I- PMH 接口,提供元数据收割功能。DSpace 系统也支持 OA I 协议和 OpenURL 协议。通过 OA I 接口,服务提供者可以定期或不定期对 DSpace 系统的元数据进行收割^[7]。厦门大学学术典藏库系统支持 OA I 协议和 OpenURL 协议,而中国科学院 IR 目前尚未正式提供对外开放的 OA I- PMH 接口。

在软件所提供的功能方面,基本上都实现了常用的检索和浏览功能,但是个性化的功能比较少见。

1.4 开放获取和访问控制 高校论文库为了保护作者的版权等内容,都会采取一定的访问控制方式,比如 IP 限制和账户限制等。在所调查的 102 所高校学位论文库中,所使用的访问控制策略比例如图 3 所示。

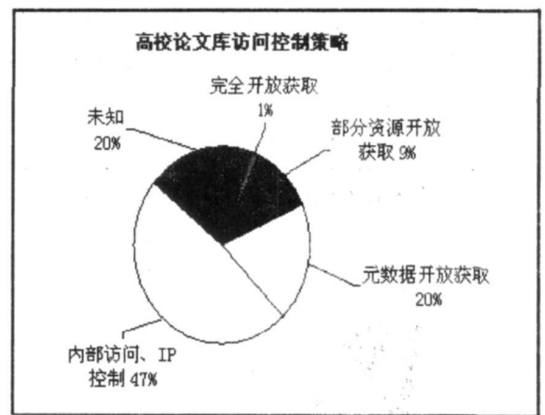


图 3 高校论文库访问控制策略比例分布

(注: 以上数据均通过外部 IP 和非登录用户访问获取)

从图中可以看出,内部访问、IP 控制的论文库在学位论文库中几乎占到了一半,资源开放程度不高。加入 CALIS 系统的高校其学位论文资源可以通过 CALIS 平台实现共享,在 CALIS 平台外的用户只能通过文献传递机制来获取资源。

在高校综合性和 IR 中,开放获取的比例达到了 50%,但除了厦门大学外,其余知识库资源相对较少,而资源较为丰富的清华大学 OAPS 则是按照作者的授权要求提供相应的三种级别的开放方式: 同意在校园网内供校内师生使用; 同意 OAPS 成员单位共享; 同意以电子、网络等数字媒体形式公开出版。中国农业大学 IR 也不提供全文的开放获取。

所调研的 50 个有 IR 的科研院所中,22 个提供外部访问功能,其余 28 个只有本机构内部可以访问,开放获取比例不到 50%。

综合来说,国内 IR 的开放获取程度较低,这不利于机构知识资产的传播、利用和机构知识生产影响力

的提升。所以国内的 IR 在“开放获取”方面还有待改进。

2 对我国 IR 发展的建议

通过对以上各个方面的分析比较,综合国内 IR 各个方面存在的问题,对我国 IR 的发展提出了几点建议,即需在以下三个方面做出转变:

2.1 制定合理的资源征缴政策——寻找资源数量和质量平衡点

在 1.2 节的分析中,已经得出征缴政策的不同会对资源数量和质量产生影响。因此如何根据机构的具体情况制定适当的征缴政策,使资源的数量和质量得到保证,也是 IR 建设的一个重要方面。

2.1.1 内容征缴政策。

上文中所提到的内容征缴激励机制分为三种(强制、鼓励、建议)。个人认为,这三种政策都各有优劣,在我国当前的科研环境下,要根据不同的情况采取不同的策略。

在 IR 建立初期,由于机构成员对于 IR 的认识不足,应该采取强制提交结合激励提交,并辅之以建议征缴的策略。对于核心的内容资源,可以采用强制征缴政策,对于期刊论文等采用激励政策,其他类型的资源采用建议征缴的策略。无论采用怎样的内容征缴政策,都不仅仅需要 IR 的建设承担单位——图书馆的努力,还需要由 IR 建设团队征得学校或者研究机构决策管理方的理解和支持,机制的建设应该自上而下推行。

2.1.2 内容质量评价机制。

内容审核机制决定知识库内资源的质量,应该根据当前 IR 的发展阶段和资源的类型来制定合理的内容审核机制。目前 IR 的内容质量依赖于其他已有的内容质量审核机制,比如学位论文评审、期刊论文和会议论文的同行评审机制,对于 OAPS 规定了“提交—评审—修改—编辑—出版”的流程,而对于其他类型的内容,尚没有合理的内容审核评价机制。对于这些内容,如果采取严格的审核机制,会减少成员提交资源的积极性,为了鼓励成员积极参与到内容的建设,可以对这方面的资源采取类似于中国科技论文在线的资源审核机制,即对内容要求很低,但是资源提交以后根据下载量和反馈来间接评价其资源质量,同时也可采取在线同行评价的机制。

2.2 转变知识库类型——从单一型向综合型转变

从上述对单一 IR 和综合性 IR 的比较也可看出综合性 IR 在各个方面较之单一 IR 都有很大的优势,因此今后国内在建设知识库方面应以综合性 IR 为主。在建设过程中应注重以下方面:

2.2.1 政策支持。

国内综合性 IR 建设方面缺乏有力的政策支持,相关部门没有把建设 IR 提上议事日程,没有把 IR 建设作为机构评价指标,因此建设积极性不高,另一方面高校对于 IR 的认识不到位,没有给

予足够的重视。

结合学位论文库建设成功的经验,自上而下的 IR 建设体制是实现 IR 顺利建设和转变的前提和保障。IR 作为一个国家知识产出存储和利用的重要工具,对于国家的知识创新和长远发展都有着至关重要的作用,因此把 IR 建设纳入国家的长远规划,作为衡量高校和科研机构科研水平的重要指标,是当前亟待解决的问题。

2.2.2 与原有系统的数据交换和数据重用。

国内 IR 发展过程中要考虑到机构原有的知识产出系统和资源,探索适合国内科研信息环境的 IR 发展模式。对于存在学位论文库的机构来说,建设综合性 IR 应该充分考虑并利用已存在的学位论文库资源,避免资源的重复建设和浪费,也可以使学位论文库通过综合性 IR 更好的发挥其潜力和价值。例如中国农业大学 IR 就将本校已有的学位论文库、教室文库、专题文献信息服务平台三大部分集成起来,中国科学院的 IR 建设中也实现了与原有的科研管理系统 ARP 的数据交换,并正在进行与原有的学位论文数据库等其他已有系统数据的交换与互操作。

2.2.3 资源收录范围。

综合性 IR 应尽可能覆盖本机构全部的知识产出,从以上的高校和科研院所来看,综合性 IR 基本上实现了这方面的功能,这些都是显性的知识产出。在此基础上对于机构内部的隐性的知识的保存和利用也是一个值得考虑的问题。

同时,不同的机构其产出的知识类型也不同,因此应该充分结合本机构的知识类型确定资源收录范围。

2.3 促进开放获取,提高资源共享程度

“开放性和互操作”是 IR 的内涵之一,而我国 IR 在这方面存在明显不足。究其原因,一是政策和传统观念的制约,二是对于 IR 收录内容的知识产权状况的迷惑和犹疑,三是知识库建设缺乏统一的标准。下面提出了三点解决建议。

2.3.1 开放获取的宣传。

开放获取的益处显而易见^[8]:对大学而言,开放获取在更大范围宣传了机构所承担的科学研究,研究论文的被引率更高;对科研资助机构而言,开放获取提供对所资助的研究成果的更大的利用,促进进一步的研究;对大学教工而言,开放获取将带来更多的读者量和更高的引用,因为学术内容放在开放的网站上比放在封闭的网站上能得到更广泛的阅读;对大学图书馆的用户而言,开放获取提供了对研究成果的利用,而没有订购或获取限制技术所施加的障碍;对政府而言,开放获取因为研究成果得到更广泛的利用,其科研投资得到了更大的回报,使公众了解科学,鼓励纳税人对研究成果给予更多的投资。但开放获取并不为多数人所了解,因此,大力宣传开放存

取,让更多的机构和科研人员认识到开放获取的重要意义是提高资源共享程度的必要手段。

2.3.2 知识产权问题。在开放获取过程中,首先必须处理好知识产权的问题。因此对 R 中资源进行分析,明确各种资源的产权状况并采取相应的产权管理策略也是值得考虑的问题。

于佳亮在《IR 资源建设中的产权策略研究》^[9]中从权利所有人角度分析,将 R 中存储的资源分为三类:本机构拥有著作权的资源、提交者拥有著作权的资源和其他权利人拥有著作权的资源。然后又提出了三种可供 R 采纳的策略:保守的资源收集策略、规避策略、分级管理策略。

各个机构可以根据不同类型的资源的知识产权状况,采取灵活的访问控制策略,例如为作者提供开放获取、在一定范围内共享、一定层次的开放获取、某段时间后开放获取或者不开放的选择。在最大程度上实现资源的共享。

2.3.3 标准的制定完善和实施。在 R 发展初期制定完善的 IR 建设标准,对于今后的发展有重要的指导意义,可以避免各个机构之间建设过程中各自为政,建设完成后形成信息孤岛,资源共享困难的状况发生。

在这方面,可以借鉴中国科学院的 IR 建设模式,即自上而下的知识库建设模式,首先建立一个完善的灵活的系统原型,各个机构再根据自己本身的情况进行适当的修改。既避免了机构之间的重复建设,也容易实现各个系统间的资源共享,同时也方便了外部访

问者通过统一的接口访问系统。

参考文献

[1] Raym Crow. The case for Institutional Repositories a SPARC Position Paper[EB/OL]. [2006/2010-04-08]. http://www.arl.org/sparc/lm~doc/ir_final_release_102.pdf

[2] OpenDOAR. Proportion of Repositories by Country - Worldwide[EB/OL]. [2010-03-04]. <http://www.opendoar.org/onechart.php>

[3] Cybermetrics Lab. Ranking Web World Repositories Top 400 Institutional Repositories[EB/OL]. [2010-04-07]. http://repositories.cybermetrics.info/top400_rep_inst.asp?offset=200

[4] 厦门大学图书馆. 厦门大学学术典藏库简介[EB/OL]. [2010-04-17]. <http://dspace.xmu.edu.cn/dspace/bitstream/2288/2480/1/厦门大学学术典藏库简介.pdf>

[5] 中科院力学研究所. 力学所 R (MECH-R) 开通通知[EB/OL]. [2009-09-07]. <http://tszcm.ecn.ac.cn/info/detail.asp?infono=12533>

[6] 北京高教学会图书馆工作研究会数字图书馆专业委员会. R 内容服务形式变革的探索与实践 - 中国农业大学[EB/OL]. [2009-11-16]. <http://lgw.lib.tsinghua.edu.cn/case/152>

[7] 厦门大学图书馆. libpedia - Dspace[EB/OL]. [2010-04-04]. <http://libfonm.xmu.edu.cn/wiki/Dspace>

[8] Fred Friend. 科学信息开放获取战略与政策国际研讨会 - 为什么欧洲的大学和科研资助机构热衷于开放获取[EB/OL]. [2005/2010-04-19]. <http://www.open-access.net/cn/5f00653e83b753d64f1a8bae/79d15b664fe1606f5f00653e83b753d6621875654e0e653f7b5656fd964578148ba84fla/>

[9] 于佳亮等. IR 资源建设中的产权策略研究[J]. 情报理论与实践, 2008(3): 353-355 (责编:王平军)

(上接第 62 页)

在着显著的促进作用,而信息生产能力、信息检索能力与信息应用能力之间并没有产生有效的支持。因此,我国企业信息能力系统具有一定的运作绩效,但仍存在若干不足之处,有待进一步整合。

本研究结论的产生基于我国企业信息能力实施的实践性检验,具有较高的现实性与客观性,从而为我国企业信息能力的改进和提高提供了可靠的理论借鉴。企业不仅可以依据信息能力开发的普遍性规律来分析自身信息能力应用的优势与不足,也可以依据本文的理论体系对自身的信息能力进行定量分析,还可以结合本文的研究结论进行企业能力研究的相关性探讨。

参考文献

[1] Kizner IM. The Driving Force of the Market Essays in Austrian Economics[M]. London, Routledge 2000

[2] Zurkowski P G. The Information Service Environment Relationship and Priorities[J]. National Commission on Libraries and Information Science 1974, 6(3): 325-346

[3] 张同健. 我国商业银行客户关系管理战略结构模型实证研究[J]. 技术经济与管理研究, 2009(6): 106-108 112

[4] Shirley Behrens. A Conceptual Analysis and Historical Overview of Information Literacy[J]. College & Research Libraries 1994 (3): 309-322

[5] 张同健, 孔胜, 李迅. 我国中小型企业技术创新战略模型研究[J]. 技术经济与管理研究, 2009(6): 28-30

[6] 汪向东. 信息化不等式: 能力 ≠ 应用 ≠ 有效[N]. 光明日报, 2006-07-20

[7] 邱均平, 粟大为, 余以胜. 企业信息能力综合评价研究[J]. 情报科学, 2006(10): 1446-1450

[8] 焦豪, 邬爱其, 张祥. 企业信息技术能力度量与功效——本土模型的构建与实证研究[J]. 科学学研究, 2008(6): 596-603

[9] 张海涛, 靖继鹏. 企业信息能力: 内涵、维度与结构模型[J]. 情报杂志, 2008(12): 109-111

[10] 莫力科. 北欧四国信息能力建设的发展及对我国的启示[J]. 未来与发展, 2009(7): 77-80

[11] 付睿臣, 毕克新. 企业信息能力到技术创新能力的传导机制研究[J]. 2009(10): 1576-1583

[12] John M ckean 著; 姚志明, 史莹如译. 信息大师: 客户关系管理的秘密[M]. 上海: 上海交通大学出版社, 2001

(责编:白燕琼)