

# 2004年-2006年国内OA研究论文定量分析

吴凤仙<sup>1,2</sup>

(1. 武汉大学信息管理学院, 湖北武汉 430072; 2. 温州大学图书馆, 浙江温州 325035)

**摘要:** 选取《中国期刊全文数据库》与《维普中文科技期刊库》2004年-2006年3年间所发表的关于OA (open access, 开放存取) 出版的206篇研究论文作为分析数据, 采用文献定量分析方法进行统计, 从时间分布、空间分布、著者分布、主题分布和引文分布等角度分析OA出版的研究状况, 认为国内OA出版研究现状是: 主题分布广泛, 作者队伍分散, 引文载体与类型多样化, 期刊引文集中于核心期刊, 研究中大量使用网络引文。“开放存取”研究正处于蓬勃发展阶段。

**关键词:** OA出版; 文献分析; 定量分析

**中图分类号:** G230   **文献标识码:** A   **文章编号:** 1006-0375(2008)02-0047-08

20世纪以来, 随着计算机网络及数字技术的快速发展, 越来越多的研究人员采用网络交流学术信息、发布科研成果, 近几年在全球兴起了开放存取 (open access) 运动, 引起了各界科研人员的广泛兴趣。

开放存取出版, 或开放获取出版 (Open Access Publishing, OAP) 是在坚持期刊原有同行评议的质量审核机制的前提下, 在现有的法律框架内实施的新型学术出版模式, 任何读者通过互联网就可以随时随地免费地获取和使用期刊论文<sup>[1]</sup>。这是国际科技界、学术界、出版界、图书馆界为打破商业出版者对科学研究信息的垄断和暴利经营而采取的运动, 其目的是推动科研成果通过Internet 免费自由地利用<sup>[2]</sup>。OA自20世纪90年代被提出之后, 这种以自由、开放与共享为理念, 以网络技术为依托的学术出版和科学信息交流模式便在短时间内形成并迅速发展起来, 随着国际上对OA研究的不断深入, 国内研究者也对OA的研究与发展给予了积极关注。

## 1 文献来源与检索方法

以《中国期刊全文数据库》与《维普中文科技期刊库》作为检索工具, 统计了2004年-2006年在国内发表的OA论文。在两个数据库中分别采用关键词、题名、摘要三种检索途径, 以“开放存取”、“开放获取”、“开放内容”、“开放使用”、“OPEN ACCESS”为检索词进行检索, 同时以上述检索词从“主题”进行检索, 再以“图书馆”进行二次检索, 以保证查全率<sup>①</sup>。通过阅读检索所得的论文, 排除重复的、与主题不相关的论文以及会议报导等消息类文献, 最后共得出2004

收稿日期: 2007-11-02

作者简介: 吴凤仙(1968-), 女, 浙江绍兴人, 馆员, 研究方向: 图书馆参考咨询服务

① 最初检索时, 检索策略中只涉及“开放存取”、“开放获取”、“OPEN ACCESS”三个检索词, 在研究过程中, 因国内研究人员对“OPEN ACCESS”一词的不同阐述与理解, 所给的中文译名不同, 发现所统计文献仍有漏检, 因此调整检索策略, 将“开放内容”、“开放使用”加入检索词, 进行重新统计, 以求数据尽可能完整、准确。

年-2006年间关于开放存取研究的论文206篇。通过对它们进行定量统计分析,探究我国学术界对开放存取研究的概况与思路。

## 2 OA 论文的统计分析

### 2.1 论文发表时间分析

根据统计,2004年OA出版研究论文仅4篇,如:陈传夫的《开放内容的类型及其知识产权管理》<sup>[3]</sup>,统计数据时,凡是关于OA会议、动态介绍等消息类文献不计作研究论文,如2004年的《Thomson ISI公司公布关于开放获取期刊影响力的报告》<sup>[4]</sup>。也就是在这一年,国内学术界对国际上轰轰烈烈的OA出版运动做出反应。在接下来的两年中,国内学术界、图书情报界对OA研究给予了积极关注,在短短两年内,发表研究论文逾200篇。2005年为50篇,2006年的论文数迅速增加到152篇,是上一年的3倍之多。由此可见国内学者对OA出版的重视与支持,这对加强国内OA出版研究的深度与广度起到促进作用。

### 2.2 论文发表空间的统计分析

#### 2.2.1 期刊类别分布

表1显示,206篇论文分布于64种期刊中,有41种图书情报类期刊为OA的研究提供了广阔的学术研究空间,共载文181篇,占总数的87.86%,论文期刊分布符合布拉德福的文献集中与分散定律<sup>[5]111-113</sup>,其他论文分散在非图书情报类期刊中,主要

表1 期刊类别分布情况

期刊类别	期刊数目(篇)	载文数量(篇)	比例(%)
图书情报信息类	41	181	87.86
高校学报类	7	7	3.40
编辑出版类	5	7	3.40
医学类	2	2	0.97
计算机及其他科技期刊类	9	9	4.37
合计	64	206	100

有高校学报类、编辑出版类、医学类、计算机及其他科技类期刊。一方面,反映出科学研究的交叉渗透,另一方面,说明了高校及其他各研究领域的科研人员对学术研究成果的出版、获取与使用的关注。

#### 2.2.2 高载文量期刊分析

表2 3年来高载文期刊统计(5篇及以上)

期刊名称	载文数量	比例(%)	期刊名称	载文数量	比例(%)
图书情报工作*	23	11.17	大学图书馆学报*	7	3.40
中国科技期刊研究	15	7.28	图书馆杂志*	7	3.40
数字图书馆论坛	11	5.34	情报科学*	6	2.91
现代情报*	10	4.85	图书情报工作动态	6	2.91
中国图书馆学报*	9	4.37	情报理论与实践*	5	2.43
图书馆学研究*	9	4.37	情报资料工作*	5	2.43
图书馆建设*	8	3.88			
图书馆论坛*	8	3.88	合计	129	62.62

注:带\*号的属于国内图书馆学情报学类的核心期刊

如果将发文5篇以上期刊视作高载文期刊,其具体信息如表2。从表2可知,在2004年-2006年三年间,关于OA出版研究载文最多的14种期刊均为图书情报信息类期刊,共载文129篇,

占总数的 62.62%。说明 OA 出版与图书馆学、情报学息息相关, 同时它也是该领域的一个研究重点。尤以《图书情报工作》最为突出, 三年内共发表 OA 研究论文 23 篇。

### 2.3 论文主题的分析

三年来, 开放存取研究论文所涉及的主题内容非常广泛。根据研究论文偏重可将 OA 研究划分为九个主题: 依次为 OA 出版模式与发展趋势、OA 模式研究与图书馆工作及其影响、OA 版权、知识产权与出版政策、OA 资源与管理研究、OA 研究综述、质量控制与评价、运行机制与经营模式、OA 案例研究、e 印本文库与内容管理等 (见表 3)。

由表 3 可知, 2004 年只有 4 篇 OA 研究论文, 分布在“OA 出版模式与发展趋势”等三个主题下, 但因这些论文发表在《中国图书馆学报》等国家级重点期刊上, 因此逐渐引起国内学术界对 OA 研究的关注, 这也说明国内图书情报领域已开始重视对 OA 的研究。

根据产生于 2001 年的“布达佩斯开放存取先导计划 (BOAI)”, 实现开放存

取出版的两种措施是建立“自我存档” (Self-Archiving) 和创办“开放存取期刊” (Open-Access Journals) [6], 那么出版模式便是真正实现开放存取的主要问题。从主题分布看, 三年中研究 OA 出版模式及发展趋势的论文最多, 共 75 篇, 占总数的 36.41%, 该主题论文的数量 2006 年是 2005 年的两倍多。在 OA 出版模式及发展趋势主题下, 集中介绍开放存取的概念、特征, 阐述 OA 实现模式与发展情况, 有的文献对 OA 出版的两种模式, 即 OA 期刊与 OA 仓储作了详细阐述, 提出 OA 在国内的发展建议。

OA 模式研究与图书馆工作及相互影响是另一研究较多的主题, 共计 46 篇论文, 占论文总数的 22.33%。主要论述 OA 的定义、类型和优势, 归纳分析 OA 出版对图书馆的冲击与影响, 探讨大学图书馆在开放存取运动中应扮演的角色和应发挥的作用, 提出图书馆促进 OA 资源利用的对策等。这符合实际情况, OA 运动的兴起本来就是为了解决图书馆有限的经费与高价学术期刊之间的矛盾, 解决“学术期刊危机”问题, 构建一个真正服务于科学研究的学术交流系统[7]。该主题的论文对于图书馆支持国内 OA 的发展、帮助用户了解、利用 OA 资源具有一定的指导意义。

此外, 版权与出版政策、OA 资源的管理研究、OA 质量控制与评价等也是 OA 研究的主要内容。在 206 篇文献中, 共有 14 篇论文是对 OA 研究的综述, 主要是从总体上对 OA 出版的定义、特点, OA 出版的产生背景与发展前景及其实现模式等进行介绍研究, 提出国内外 OA 的整体发展趋势以及对学术界的影响。

除了以上九大主题外, 其他研究主题还包括: OA 中文译名探讨、档案界与 OA 研究及高校学术期刊与 OA 出版等, 但均只占 1-2 篇论文。

表 3 3 年来各种主题年度分布表

主题	2004 年	2005 年	2006 年	合计	比例 (%)
OA 出版模式 发展趋势	2	22	51	75	36.41
OA 模式研究 图书馆 影响	0	11	35	46	22.33
版权 知识产权 出版政策	1	2	17	20	9.71
OA 资源 管理研究	0	6	13	19	9.22
综述	1	4	9	14	6.80
质量控制 评价	0	2	7	9	4.37
运行机制 经营模式	0	0	8	8	3.88
OA 案例研究	0	0	6	6	2.91
e 印本文库 内容管理	0	0	5	5	2.43
其他	0	3	1	4	1.94
合 计	4	50	152	206	100

在所有主题当中,偏重于阐述OA的定义、概念、产生背景、OA实施模式、国外发展概况,以及国外OA实施政策与制度、OA版权保护与知识产权、OA对图书馆服务的影响及图书馆的对策等的研究.而对于国内发展OA的相关制度、政策、技术与经费支持等研究涉及较少,有待于今后对之作进一步的研究,以找到国内OA资源的开发、管理与利用的切入点.

#### 2.4 论文著者的统计分析

三年间共有213位著者发表了关于开放存取的研究论文206篇,其中独著论文有146篇,以第一著者身份发表论文的作者有136位(一位作者发表多篇以一位计).

##### 2.4.1 论文署名分析

表4 3年中论文署名统计<sup>①</sup>

年份	独著(篇)	2人合著(篇)	3人合著(篇)	4人合著(篇)	5-7人合著(篇)	总数(篇)
2004年	3	1	0	0	0	4
2005年	39	9	2	0	0	50
2006年	104	28	10	5	5	152
合计	146	38	12	5	5	206
比例(%)	70.87	18.45	5.83	2.43	2.43	100

表4数据表明,第一,参与开放存取研究的著者较多,但著者的平均发文量较低,人均不到一篇;第二,以独著为主,占有著作方式的70.87%;第三,合著方式中,俩人合著的比例最高,2005年只有2人或3人合著,共11篇论文,而2006年合著比例与人数均大为增加,俩人合著是前一年的近三倍,并且出现了4人、5人甚至更多人合著的情况,合著者最多达7人,说明该课题研究以个人为主,同时注重合作研究.3年内共有306人次参与OA出版研究.

##### 2.4.2 核心作者分析

通过统计分析论文著者人数,可以知道该课题研究领域是否已形成核心著者群.

表5 发表论文篇数的著者统计

发文量	1篇	2篇	3篇	4篇	5篇	6篇	7篇	13篇	合计
著者人数	173	24	5	2	2	3	3	1	213
比例(%)	81.22	11.27	2.35	0.94	0.94	1.41	1.41	0.47	100

从表5可知,我国关于此课题研究发表一篇论文的著者比例为81.22%,而且发表两篇论文的人数约是发表一篇论文著者的14%,与洛特卡定律的学科参考值差距较大<sup>[5]178-183</sup>.根据普赖斯定律:某一特定的领域中,“核心著者”的发文量应为总论文数的50%<sup>[5]188-190</sup>,按普赖斯定律的公式计算,此研究课题高产作者或“核心著者”的最低论文数m应为2.7篇,取整数,定为3篇,则核心著者为16人,占著者总数的7.51%,3年间共参与85篇论文的撰写,占论文总数的41.26%,离总论文数的50%尚有差距<sup>②</sup>.因此,无论从洛特卡定律出发,还是以普赖斯定律为基准,国内此课题的研究未形成核心著者群体.

##### 2.4.3 著者所在地区分布情况

著者地区分布情况可以帮助我们了解课题是否出现地区规模化(见表6、表7).

① 译作以译者为主计算,如原作为合著而译者为一,则认为是独著.

② 洛特卡定律认为:写一篇论文的著者人数应该占总人数的60.79%左右,而写两篇论文的人数应该为写一篇论文人数的1/4;写三篇论文的人数约为写一篇论文人数的1/9.

③ 普赖斯定律的计算公式为:  $m=0.749(n_{\max}^{1/2})$ ,  $n_{\max}$ =统计年限中最高产作者的论文数, m=确定“核心著者”的最低论文数.

表 6 著者区域分布表

类别	华中	华北	华东	华南	东北	西南	西北	国外	合计
篇数	58	58	43	21	8	7	2	9	206
比例 (%)	28.16	28.16	20.87	10.19	3.88	3.40	0.97	4.37	100

从统计结果看, 论文著者分布于全国 20 个省 3 个直辖市, 另外有 9 篇论文来自国外, 原因: 其中有 8 篇论文是张晓林在《图书情报工作》于 2005、2006 年组织了“科技信息的开放获取”和“科技信息开放获取的战略与政策”两辑专题中发表的, 此专题论文来自于中国科学院文献情报中心于 2005 年 6 月在北京召开的“科技信息开放获取战略与政策国际研讨会”。另有一篇是国外作者发表于《图书情报工作》2005 年第 6 期上的研究论文。这些来自于国外专家的研究论文, 虽由国内图书情报刊物组织编辑发表, 但不计入国内任一地区或城市的研究成果, 将其单独列出。

表 7 著者所在省份前 8 位分布表

类别	北京	浙江	湖南	河南	广东	湖北	江苏	辽宁	合计
篇数	53	26	21	20	20	17	7	6	170
比例 (%)	25.73	12.62	10.19	9.71	9.71	8.25	3.40	2.91	82.52

由表 6 著者区域分布可知, 从全国来看: 华中、华北和华东地区发文相对于别的地区多, 分别占总数的 28.16%、28.16% 和 20.87%, 这三个地区分布相对比较均匀, 也就是说没有一个地区是作者特别集中的, 但显示出分散中有集中。若按照各省份单独统计, 根据表 7, 发文作者最多的是北京, 有 53 篇, 接着是浙江、湖南等省, 最少的省只有一篇; 由此可见该课题研究作者分布范围广, 地区间不均衡, 没有形成规模的集中研究区域。

## 2.5 引用文献的统计分析

### 2.5.1 引文类型分布

笔者对 206 篇关于开放存取研究论文的参考文献进行统计, 有 11 篇没有参考文献, 总引文量为 2 407 篇 (见表 8)。

从表 8 可以看出, 国内关于开放存取研究论文的参考文献, 引用自互联网的文献最多, 占总数的 60.37%, 其次是中文期刊, 占 25.26%, 接着是外文期刊, 占 8.64%, 其他还有图书、学位论文、报纸文献、会议报告等, 可见引用文献的载体和类型的多样化。参考文献总数为 2 407 篇, 平均每篇论文的引文量为 11.68 篇, 这一方面说明大量引用网络文献提高了论文的引文数, 另一方面也说明国内研究者引用参考资料比过去丰富。

表 8 不同类型文献被引用情况<sup>①</sup>

引文类型	中文期刊	外文期刊	互联网	中文图书	其他文献 <sup>②</sup>	合计
被引频次	608	208	1453	49	89	2407
比例 (%)	25.26	8.64	60.37	2.04	3.70	100.00

### 2.5.2 印刷型文献与网络文献的引用对比

由表 9 可见, 研究论文中“网络引文”现象与日俱增, 不光打破了传统清一色印刷型引文的现象, 并且已成为又一种重要的引文类型。在此课题研究论文中, 网络引文的比率甚至高于印刷

① 同一论文中重复引用同一篇文献时, 以一篇计。

② 其他文献指被引用的学位论文、年度报告、报纸、研讨会或学术报告会论文、机构公报等文献。

型引文,占60.37%,成为研究者参考的首选资料来源,这也是OA出版作为网络资源并以网络为支撑的一个研究特点.

### 2.5.3 引文情况分析

表9 印刷型文献与网络文献引用比较

引文载体	2004年	2005年	2006年	合计	比例(%)
印刷型文献	5	225	724	954	39.63
网络文献	70	434	949	1453	60.37
合计	75	659	1673	2407	100

从被引用的87种中文期刊中列出了引文频次靠前的16种期刊(见表10).根据统计,中文期刊引用频次靠前的16种期刊中有13种是图书情报类的期刊,其中有10种是属于图书馆学情报学类的核心期刊,在不到20%的引用期刊中引用率超过了半数,占73.36%,特别是《中国图书馆学报》,被引用达92次,由表10可以看出,开放存取研究所引用的期刊来源集中于核心期刊.此现象符合加菲尔德引文集中规律,即大部分引文来自于本学科的少数核心期刊<sup>[5]323-325</sup>.

表10 引文期刊类别分布

序号	刊名	引用次数	比例(%)	序号	刊名	引用次数	比例(%)
1	中国图书馆学报	92	15.13	10	图书馆论坛	16	2.63
2	中国科技期刊研究	81	13.32	11	现代图书情报技术	16	2.63
3	大学图书馆学报	43	7.07	12	图书与情报	13	2.14
4	图书情报工作	40	6.58	13	图书馆杂志	11	1.81
5	情报资料工作	23	3.78	14	情报杂志	11	1.81
6	数字图书馆论坛	23	3.78	15	图书情报知识	10	1.64
7	图书情报工作动态	22	3.62	16	科技情报开发与经济	10	1.64
8	编辑学报	18	2.96				
9	出版经济	17	2.80	合计		446	73.36

另外,在统计中发现,李武、李春旺等少数核心作者的论文被引用频次极高,其中李武与刘兹恒合著的《一种全新的学术出版模式:开放存取出版模式探析》一文单篇论文被引用最高,达43次,其次是李春旺的《网络环境下学术信息的开放存取》一文被引用28次,接着是任胜利、乔冬梅等的文献,这些引用频次高的论文全部来自核心期刊.这一方面说明国家级核心期刊仍然是用户具有较高参考价值的首选来源,另一方面又说明高质量的研究论文是国内学者参考研究的首选对象,这些高引用频次文献有利于提高期刊总被引频次和期刊的影响因子,但也易导致研究成果的重复、雷同.

## 3 结 语

从2004-2006年的三年间OA出版研究论文的定量分析中可见,研究呈现出如下特点:

### 3.1 主题分布广泛

开放存取研究的主题分布广泛.出版模式及其发展趋势和其与图书馆工作关系的研究是大家比较感兴趣的话题,但总体看来,大部分论文都集中于介绍OA的定义、概念、产生的原因、实施模式、发展概况、OA出版的影响以及图书馆的对策等方面的研究,而针对OA质量评价方法、技术与经费支持以及OA在国内发展的相关制度、政策等方面主题的则涉及较少,有待于今后进一步的研究,以找到国内OA资源的开发、管理与利用的切入点.

### 3.2 作者队伍分散

从 206 篇论文来看, 共有 136 位作者是以第一著者的身份发表论文的, 但共有 213 位作者参与了此课题的研究, 也就意味着平均每位作者发文不到 1 篇, 高发文量作者仅占著者总数的 7.51%, 这表明国内此课题的研究还未形成核心著者群. 213 位作者分布在全国 20 省及 3 个直辖市, 但发文最多的地区仅占有 28.16% 的发文量, 最多的省或城市也只达到 25.73% 的发文量, 说明该课题研究作者分布范围广, 地区间不均衡, 尚未形成规模的集中研究区域, 但仍显示分散中有集中.

### 3.3 引文特点

主要表现在: 第一, 大量引用网络引文, 占总引文数的 60.37%, 说明网络引文已成为研究者的一种重要引文类型, 但网络文献本身的多变性和引用者标注的不规范和随机性(如未标注访问时间与文献类型标识), 增加了研究本身的不确定性, 易影响研究论文的评价效果和准确性; 第二, 引文来源除了期刊和互联网, 还有图书、报纸、学位论文及会议报导等, 显示国内研究引用文献的载体和类型多样化趋势; 第三, 期刊引文统计显示, 引文高度集中于权威的核心期刊以及单篇文献, 导致研究论文内容出现雷同、重复现象.

总之, 开放存取在国内学术界虽已引起极大关注, 但是, 研究尚不够注重与实践相结合. 目前国内虽然有了奇迹文库、中国科技论文在线、中国预印本服务系统等本国的“开放存取”知识库, 但是这三个“开放存取”知识库只为用户提供上传、浏览、下载学术论文, 为研究者提供免费、方便、快捷的 E-print 平台, 既不对文献作评审、修改、编辑, 更无相关领域的专家对论文作质量评审. 如, 中国预印本服务系统在其网站明确申明: 系统不对文章进行学术审核, 文章仅仅代表作者个人的观点, 不代表中国预印本服务系统的观点, 在系统中存储的文章, 作者可以自行以任何方式在其他载体上发表<sup>[8]</sup>. 因此, 就这三个知识库来说, 其规模和影响力尚小, 主要是缺乏相关的政策支持, 国内学者也大多抱观望态度. 可见, 开放存取还是一项新的事业, 需要科研人员 and 科研机构、学术团体、科研资助机构、大学师生、图书馆员、政府机构以及每个公民的积极参与和鼎力支持, 并在其中发挥应有的作用<sup>[9]</sup>. OA 研究在我国方兴未艾.

### 参考文献

- [1] 李麟, 初景利. 开放获取出版模式研究[J]. 图书馆论坛, 2005, 25(6): 88-93.
- [2] 肖冬梅. 开放存取运动缘何蓬勃兴起?[J]. 图书情报工作, 2006, 50(5): 128-131.
- [3] 陈传夫. 开放内容的类型及其知识产权管理[J]. 中国图书馆学报, 2004, 30(1): 9-13.
- [4] 本刊记者. Thomson ISI 公司公布关于开放获取期刊影响力的报告[J]. 图书情报工作动态, 2004, 11: 27-28.
- [5] 邱均平. 文献计量学[M]. 北京: 科学技术文献出版社, 1988.
- [6] 李武. 一种全新的学术出版模式: 开放存取出版模式探析[EB/OL]. [2007-07-21]. <http://openaccess.bokee.com/220202.html>.
- [7] 李武. 基于开放存取的学术期刊出版模式研究[EB/OL]. [2007-07-21]. <http://openaccess.bokee.com/3504299.html>.
- [8] 中国预印本服务系统. 中国预印本服务系统: 系统介绍[EB/OL]. [2007-07-21]. <http://preprint.nstl.gov.cn/eprint/help.jsp#check>.
- [9] 初景利. 开放获取的发展与推动因素[J]. 图书馆论坛, 2006, 26(6): 238-242.

# Literature Quantitative Analysis on Open Access Research in China from 2004 to 2006

WU Fengxian<sup>1,2</sup>

(1. School of Information Management, Wuhan University, Wuhan, China 430072;

2. Wenzhou University Library, Wenzhou, China 325035)

**Abstract:** On the basis of the CNKI Full Text Database and Chinese VIP Science and Technology Full Text Database, the paper retrieves 206 articles on “Open Access” published from 2004 to 2006. Using the method of literature quantitative analysis, this paper studies them from five aspects —— the distributions of time, space, author, contents and the quotation of the articles. With the research, author intends to produce a framework of the Open Access research in the past three years and puts forward the characteristics of the Open Access research, such as the universality on contents distribution, the wide distribution of the authors, and the variety of quoted literatures types. The discussion leads to the conclusion that the Open Access research is at its flourishing seedtime.

**Key words:** Open access; Literature analysis; Quantitative analysis

(编辑: 刘慧青)