Digital Library Forum 数字图书馆论坛
http: // www.dlf.net.cn

近20年世界数字图书馆研究论文的计量分析：基于SCI-E和SSCI

□ 陈娟 / 厦门大学图书馆 厦门 361005

摘要：文章运用文献计量学的相关理论和方法，对近20年SCI-E和SSCI收录的有关数字图书馆的研究论文进行分析，揭示该主题论文的增长情况、核心作者及其机构分布、核心期刊。统计反映近20年数字图书馆研究热点的高频关键词56个，应用共词聚类法描述数字图书馆领域当前的8大研究热点，通过绘制战略坐标图分析各研究热点的发展趋势，以期为该领域研究提供参考。

关键词：洛特卡定律，K-S检验，布拉德福分散定律，共词聚类分析法，战略坐标图
DOI：10.3772/j.issn.1673—2286.2010.11.008

1 导论

文献计量学是1969年由英国学者普里查德首先提出的，指采用数学、统计学方法，对各类文献的诸计量特征进行统计分析，进而揭示和研究文献情报规律、文献情报科学管理以及学科发展趋势的一门科学。本文以“数字图书馆”为主题，采用文献计量分析的方法，从发文增长规律、作者分布规律、来源文献分布规律以及高频关键词等方面，较系统深入地分析近20年来世界数字图书馆研究领域的文献增长、核心作者、核心期刊以及研究重点。

1.1 数据来源

Web of Science（简称WOS）是美国汤姆森科技信息集团基于WEB开发的产品，是大型综合性、多学科、核心期刊引文索引数据库，包括三大引文数据库[Science Citation Index（简称SCI）、Social Sciences Citation Index（简称SSCI）和Arts & Humanities Citation Index（简称A&HCI）]和两个化学信息事实型数据库（Current Chemical Reactions，简称CCR和Index Chemicus，简称IC），以ISI Web of Knowledge作为检索平台。本文选取WOS的SCI-E（SCI-Expanded）和SSCI作为数据来源，以"digital library " OR "digital libraries"为检索词在主题检索字段进行词组检索，选择文献类型为ARTICLE OR PROCEEDINGS PAPER的文献，共得到1935篇论文。

1.2 数据处理及方法

从ISI Web of Knowledge检索平台导出1935篇相关论文的题录，将题录信息导入Excel。首先运用文献计量分析的方法，拟合文献量的增长曲线并计算文献量翻倍时间以了解该领域文献信息的增长规律。其次将计量分析与内容分析相结合，依次对作者、机构、文献来源、关键词进行分析：尝试性地应用洛特卡定律分别选择文献第一作者及所有作者进行计算，拟合其洛特卡曲线，并运用K-S检验验证洛特卡定律对该学科的适用性，尝试性地应用普赖斯定律分别选择文献的第一作者及所有作者进行计算，确定活跃作者发文量的上下限，分析排名前15的活跃作者及其所在机构的分布特征；分析该领域的文献信息集中与分散分布规律，绘制布拉德福分布图，应用埃格1986年提出的布拉德福核心区数量计算法计算各区期刊数量，从而确定该领域的核心期刊；结合内容分析法，对高频关键词进行共词分析，运用SPSS、Netdraw等软件绘制该领域研究热点与结构。

2 文献增长规律

根据SCI-E和SSCI数据（1987年至今），1990年出现数字图书馆方面的文献，1995年文献量出现了剧增，2004年达到近20年至高点，2007年开始进入稳定期。根据20年数据，描绘出文献量的普赖斯曲线$F(t)=2.168e^{0.276t}$，数字图书馆文献量

表1 年度发文量

年份	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
发文量（篇）	1	1	1	3	3	51	64	82	66	105
年份	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
发文量（篇）	143	110	146	153	220	211	187	116	125	111

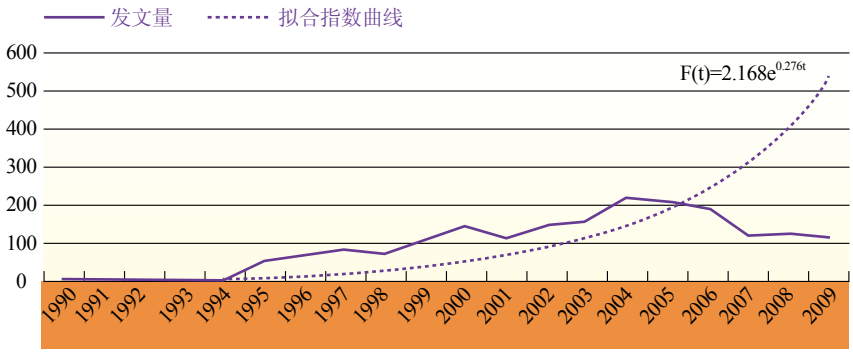

图1 文献增长量

翻倍时间为2.51[①]年。

3 作者分布规律

3.1 洛特卡定律

洛特卡定律是文献计量学三大定律之一，是描述论文与著者科学生产频率分布的重要定律。1926年美国统计学家A. J. 洛特卡在《华盛顿科学院杂志》上发表了《科学生长率的频率分布》，他研究了化学和物理学领域中作者数量与论文数量的关系，得出作者数量与论文数量的关系中遵循“平方反比定律”，其经典公式为$f(y_x)=\frac{c}{x^2}$[②]。此后其一般表达式扩展为$x^n y_x=w\cdot c$，w为论文总数。通过适当变换，得到$y_x=\frac{w\cdot c}{x^n}$，两边取自然对数进行转换，得到：

$$\ln y_x=\ln(w\cdot c)-n\ln x$$

从而得到 的估计量为：

$$\hat{n}=-\frac{N\sum\ln x\cdot\ln y_x-\sum\ln x\sum\ln y_x}{N\sum(\ln x)^2-(\sum\ln x)^2},$$

N为被观察数据对的数量。

3.2 洛特卡定律的应用

洛特卡在计算过程中仅选择了文献的第一作者，并排除了高产作者（极限下取值），只是对物理和化学领域抽样而导出的理论估计，并非精确的统计分布。洛特卡定律要求所研究的学科必须相对稳定，论文时间区间必须足够长。为了评价该定律对其他学科的适用性，F. J. 科尔于1977年提出用Kolmogorov-Smirnov进行检验，并对美国伊利诺伊大学图书馆和国会图书馆的文献书目和作者进行了统计检验，前者符合洛特卡的期望，后者则有所偏离。K-S检验的基本思想是观察数值的累积频率分布与理论值的累积频率分布之间的差异，选取最大的差值记为D_{max}，若D_{max}大于某显著水平（一般取0.01）下的临界值，则认为理论和实际有显著差异；若D_{max}小于临界值，则不能认为理论与实际有显著差异。

本文试运用该定律对SCI-E和SSCI收录的数字图书馆的研究论文进行分析。分别选择文献的第一作者和文献的所有作者进行计算，并拟合相应的洛特卡曲线，运用K-S检验法验证该定律对数字图书馆领域文献的适用性。

3.2.1 第一作者

（1）洛特卡曲线

选择文献的第一作者进行计算，作者共计1508人，共发表论文1935篇，发文量小于等于5的作者占99.34%，发文量大于5定义为高产作者，的估计量计算公式中涉及的项目如表2所示。

① $F(\Delta t)=e^{0.276\Delta t}=2$，$\Delta t=\frac{\ln 2}{0.276}\approx 2.51$

② y_x是发表x篇论文的作者数量，$f(y_x)$为发表x篇论文的作者数占作者总数的比例，由于$\sum_{x=1}^{x_{max}}\frac{1}{x^2}$收敛于$\frac{6}{\pi^2}$，且$\sum_{x=1}^{x_{max}}f(y_x)=1$，由此可以得出$c=\frac{1}{x^2}=0.6079$，即仅写一篇论文的人数占作者总人数的60.79%。

表2 洛特卡定律计算表（第一作者）

x	y_x	$x \cdot y_x$	lnx	lny_x	$lnx \cdot lny_x$	$lnx \cdot lnx$
1	1246	1246	0.00	7.13	0.00	0.00
2	177	354	0.69	5.18	3.59	0.48
3	49	147	1.10	3.89	4.28	1.21
4	19	76	1.39	2.94	4.08	1.92
5	7	35	1.61	1.95	3.13	2.59
$\sum_{x=1}^{5}$			4.79	21.09	15.08	6.20

$$\hat{n}=\frac{N\cdot\sum(\ln x\cdot\ln y_x)-\sum\ln x\sum\ln y_x}{N\sum(\ln x)^2-(\sum\ln x)^2}=\frac{5\times15.08-4.79\times21.09}{5\times6.20-4.79^2}\approx3.20$$

$$\hat{c}=\frac{1}{\sum_{x=1}^{20}\frac{1}{x^{3.2}}}\approx0.857$$，其洛特卡公式为$f(y_x)=\frac{0.857}{x^{3.20}}$。

（2）K-S检验

选择文献的第一作者进行计算，拟合的洛特卡曲线，其K-S检验结果如表3所示。

$D_{max}=0.0307<D_{临界}=\frac{1.643}{\sqrt{\sum y_x}}=\frac{1.643}{\sqrt{1508}}=0.0423$，因此显著水平0.01上认为SCI-E和SSCI收录的数字图书馆论文作者（第一作者）分布服从洛特卡分布$f(y_x)=\frac{0.857}{x^{3.20}}$。

3.2.2 所有作者

（1）洛特卡曲线

选择文献的所有作者进行计算，作者共计3720人，共发表论文1935篇，发文量小于等于10的作者占99.62%，发文量大于10定义为高产作者，其计算原理同3.2.1之（1），故略去其洛特卡定律计算表。

$$\hat{n}=\frac{N\cdot\sum(\ln x\cdot\ln y_x)-\sum\ln x\sum\ln y_x}{N\sum(\ln x)^2-(\sum\ln x)^2}=\frac{10\times43.05-15.10\times37.57}{10\times27.65-15.1^2}\approx2.83$$

$$\hat{c}=\frac{1}{\sum_{x=1}^{20}\frac{1}{x^{2.83}}}\approx0.808$$，其洛特卡公式为$f(y_x)=\frac{0.808}{x^{2.83}}$。

（2）K-S检验

选择文献的所有作者进行计算，拟合的洛特卡曲线，其K-S检验方法同3.2.1之（2），故略去。

$D_{max}=0.0247<D_{临界}=\frac{1.643}{\sqrt{\sum y_x}}=\frac{1.643}{\sqrt{3720}}=0.0269$，因此显著水平0.01上认为SCI-E和SSCI收录的数字图书馆论文作者（所有作者）分布服从洛特卡分布$f(y_x)=\frac{0.808}{x^{2.83}}$。

表3 K-S检验（第一作者）

x	y_x	实际累计频率	理论累计频率	D
1	1246	0.8263	0.8570	0.0307
2	177	0.9436	0.9503	0.0066
3	49	0.9761	0.9757	0.0004
4	19	0.9887	0.9859	0.0028
5	7	0.9934	0.9909	0.0025
6	3	0.9954	0.9936	0.0017
7	2	0.9967	0.9953	0.0014
8	2	0.9980	0.9964	0.0016
9	2	0.9993	0.9972	0.0022
11	1	1.0000	0.9976	0.0024

3.3 活跃作者群的分布

1963年美国文献计量学家D. 普赖斯在洛特卡定律的基础上提出普赖斯定律，确立了筛选科学界核心研究力量的有效方法。普赖斯定律的基本内容是：就某一学科而言，核心科学家中最低产的那位科学家发表论文数等于最高产的科学家发表论文数的平方根的0.749倍，其数学表达式为：$m=0.749\sqrt{n_{max}}$（m表示核心著者中发表论文数的最低值，n_{max}表示核心的著者中发表论文数的最高值）。

选择文献的第一作者计算，其m值为：$m_{第一}=0.749\sqrt{n_{max}}=0.749\sqrt{11}\approx3$；选择文献的所有作者计算，其m值为：$m_{所有}=0.749\sqrt{n_{max}}=0.749\sqrt{26}\approx4$。

按第一作者计算的活跃作者其

表4 Top 15活跃作者排名

按第一作者计算			按所有作者计算		
位次	作者	篇数	位次	作者	篇数
1	Witten, IH	11	1	Fox, EA	26
2	Borgman, CL	9	2	Witten, IH	22
3	Chen, HC	9	3	Goncalves, MA	18
4	Fox, EA	8	4	Goh, DHL	17
4	Theng, YL	8	4	Theng, YL	17
6	Adam, NR	7	6	Chen, HC	15
6	Candela, L	7	6	Blandford, A	15
8	Kim, H	6	8	Lim, EP	13
8	Bollen, J	6	9	Nelson, ML	12
8	Ding, H	6	9	Maly, K	12
11	Agosti, M	5	11	Zubair, M	11
11	D'Alessandro, DM	5	11	Shen, R	11
11	Nottelmann, H	5	11	D'Alessandro, DM	11
11	Chen, CM	5	11	Marchionini, G	11
11	Yang, CC	5	15	Buchanan, G	10
11	Shiri, A	5	15	Xing, CX	10
11	Frias-Martinez, E	5	15	Bollen, J	10
			15	Fuhr, N	10
			15	Bainbridge, D	10

发文量范围为[3,11]，共计核心作者85人，占作者总数的5.64%，核心作者发表的论文335篇，占论文总数的17.31%；按所有作者计算的活跃作者其发文量范围为[4,26]，共计核心作者160人，占作者总数的4.30%，核心作者发表的论文987篇，占论文总数的18.50%。表4分别列出按第一作者和所有作者计算的排名前15的活跃作者。

表5 共同活跃作者所在机构

国家/地区	所在机构	院系	作者
英格兰	Middlesex Univ（密德萨斯大学）	Sch Comp Sci, Interact Design Ctr	Theng, YL
印度	Virginia Tech（弗吉尼亚理工大学）	Dept Comp Sci	Fox, EA
新西兰	Univ Waikato（怀卡托大学）	Dept Comp Sci	Witten, IH
新加坡	Nanyang Technol Univ（南洋理工大学）	Sch Commun & Informat, Div Informat Studies	Theng, YL
美国	Univ Arizona（亚利桑那大学）	Dept Management Informat Syst	Chen, HC
	Los Alamos Natl Lab（洛斯阿拉莫斯国家实验室）	Res Lib	Bollen, J
	Univ Iowa（爱荷华大学）	Coll Med, Dept Pediat	D'Alessandro, DM

3.4 排名前15的活跃作者机构分析

分别选择文献的第一作者和所有作者两种方式计算，来自英国密德萨斯大学的Theng, YL、印度弗吉尼亚理工大学的Fox, EA、新西兰怀卡托大学的Written, IH、美国亚利桑那大学的Chen, HC、美国洛斯阿拉莫国家实验室的Bollen, J和美国爱荷华大学的D'Alessandro, DM发文量均稳固位于活跃作者前20名内。他们主要来自高校计算机及信息系统管理系，D'Alessandro, DM则来自爱荷华大学医学院。

选择文献的第一作者计算，发文量排名前15的活跃作者主要来自在美洲的美国和加拿大，欧洲的英格兰、苏格兰、德国、意大利和挪威，大洋洲的新西兰，东南亚的印度、韩国、新加坡及中国香港。他们几乎来自高校，高校中以计算机、信息系统管理系所为多，图书馆和信息科学系也占部分比例，还有少数来自国家的研究机构，Shiri, A则来自斯特拉斯克莱德大学的数字图书馆研究中心，具体情况见表6。

选择文献的所有作者计算，得到的排名前15活跃作者的机构分布情况与按第一作者计算得到的结果类似。稍有不同的是，清华大学计算机与科学技术系的邢春晓进入了前15活跃作者，其发表有关数字图书馆方面的文献有10篇为SCI-E和SSCI收录。

表6 TOP 15活跃作者所在机构（第一作者）

国家/地区	所在机构	院系	作者
加拿大	Univ Alberta（艾尔伯塔大学）	Sch Lib & Informat Studies	Shiri, A
英格兰	Middlesex Univ（密德萨斯大学）	Sch Comp Sci, Interact Design Ctr	Theng, YL
	Brunel Univ（布鲁内尔大学）	Sch Informat Syst Comp & Math	Frias-Martinez, E
德国	Univ Duisburg Essen（杜伊斯堡—埃森大学）	Dept Informat	Nottelmann, H
香港	Chinese Univ Hong Kong（香港中文大学）	Dept Syst Engn & Engn Management	Yang, CC
印度	Virginia Tech（弗吉尼亚理工大学）	Dept Comp Sci	Fox, EA
意大利	CNR（意大利国家研究中心）	ISTI	Candela, L
	Univ Padua（帕多瓦大学）	Dept Informat Engn	Agosti, M
新西兰	Univ Waikato（怀卡托大学）	Dept Comp Sci	Witten, IH
挪威	Norwegian Univ Sci & Technol（挪威科技大学）	Dept Comp & Informat Sci	Ding, H
苏格兰	Univ Strathclyde（斯特拉斯克莱德大学）	Ctr Digital Lib Res	Shiri, A
新加坡	Nanyang Technol Univ（南洋理工大学）	Sch Commun & Informat, Div Informat Studies	Theng, YL
韩国	Yonsei Univ（延世大学）	Dept Lib & Informat Sci	Kim, H
美国	Los Alamos Natl Lab（洛斯阿拉莫斯国家实验室）	Res Lib	Bollen, J
	Rutgers State Univ（罗特格斯州立大学）	Ctr Informat Management Integrat & Connect	Adam, NR
	Univ Arizona（亚利桑那大学）	Dept Management Informat Syst	Chen, HC
	Univ Calif Los Angeles（加州大学洛杉矶分校）	Grad Sch Educ & Informat Studies, Dept Informat Studies	Borgman, CL
	Univ Iowa（爱荷华大学）	Coll Med, Dept Pediat	D'Alessandro, DM

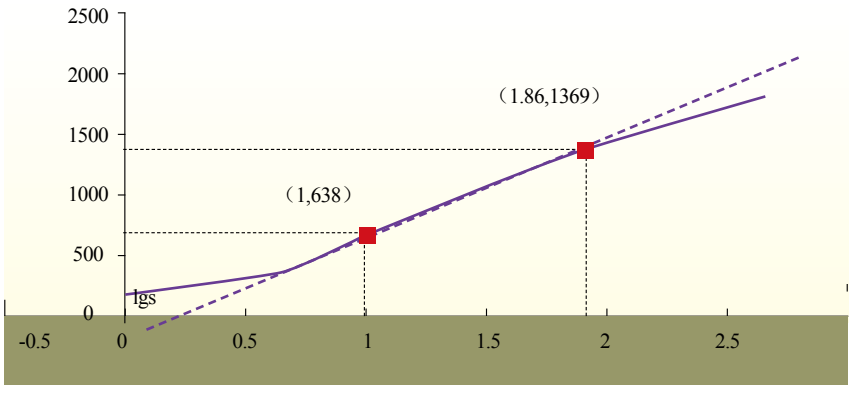

图2 布拉德福分布图

4 来源文献分布规律

4.1 布氏定律

布氏定律是英国文献学家布拉德福于1934年在《工程》杂志上发表的《专门学科的情报源》上提出的描述文献分散规律的经验定律。文字表述为：“如果将科技期刊按其刊载某专业论文的数量多寡，以递减顺序排列，则可分出一个核心区和相继的几个区域，每区刊载的论文量相等，此时核心期刊和相继区域期刊数量成1:n:n^2的关系。”

4.2 布拉德福分布图

通过绘制布拉德福分布图，能更为直观地反映文献分布的状况。布拉德福首先根据期刊发表论文数量降序排列（论文数量相同的期刊列入同一等级），以累计期刊量的对数lgC为横坐标，以相应的论文累积数R(n)为纵坐标绘制二维坐标图。由图可知，该主题文献分布基本符合布拉德福分布定律：布拉德福曲线首先是一段上升的曲线，在进入直线部分以后出现弯曲下垂。

由图可以看出，第I区与第II区分界线出现在点（1, 638），其累计期刊量为10，累计载文量为638篇，第II区和第III区分界线出现在点（1.86, 1369），其累计期刊量为72，累计载文量为1369篇。

4.3 核心期刊的确定

可利用情报学家埃格1986年提出的布拉德福核心区数量计算法计算各区期刊数量，计算公式为：$n=(e^{E}\bullet Y)^{1/p}$，n为布拉德福系数，p为分区数，E为欧拉系数，Y为期刊

表7 TOP 15 活跃作者所在机构（所有作者）

国家/地区	所在机构	院系	作者
巴西	Univ Fed Minas Gerais（米纳斯联邦大学）	Dept Comp Sci	Goncalves, MA
英格兰	Middlesex Univ（密德萨斯大学）	Sch Comp Sci, Interact Design Ctr	Theng, YL
			Buchanan, G
	UCL（伦敦大学学院）	Interact Ctr	Blandford, A
德国	Univ Duisburg Essen（杜伊斯堡—埃森大学）	Inst Informat & Interact Syst	Fuhr, N
印度	Virginia Tech（弗吉尼亚理工大学）	Dept Comp Sci	Fox, EA
新西兰	Univ Waikato（怀卡托大学）	Dept Comp Sci	Witten, IH
			Bainbridge, D
中国	Tsing Hua Univ（清华大学）	Dept Comp Sci & Technol	Xing, CX
新加坡	Nanyang Technol Univ（南洋理工大学）	Sch Commun & Informat, Div Informat Studies	Theng, YL
			Goh, DHL
		Sch Comp Engn, Ctr Adv Informat Syst	Lim, EP
美国	Los Alamos Natl Lab（洛斯阿拉莫斯国家实验室）	Res Lib	Bollen, J
	NASA（美国国家航空航天局）	Langley Res Ctr	Nelson, ML
	Old Dominion Univ（多米尼恩大学）	Dept Comp Sci	Maly, K
			Zubair, M
	Univ Arizona（亚利桑那大学）	Dept Management Informat Syst	Chen, HC
	Univ Iowa（爱荷华大学）	Coll Med, Dept Pediat	D'Alessandro, DM
	Univ N Carolina（北卡罗来纳大学）	Sch Informat & Lib Sci	Marchionini, G
	Virginia Tech（弗吉尼亚理工大学）	Dept Comp Sci	Goncalves, MA
			Shen, R
威尔士	Univ Coll Swansea（斯旺西大学）	Dept Comp Sci	Buchanan, G

最高载文量。

1935篇文献来源于443种出版物，最高载文149篇，因此：

$n=(e^{E}\cdot Y)^{1/P}=(e^{0.5772}\cdot 149)^{1/3}\approx 6.426$

第I区期刊数量的计算公式为：

第II区的期刊数量为：

$R=N(n-1)/(n^{p}-1)=443\cdot(6.426-1)/(6.426\text{^}3-1)=9.09\approx 9$

第III区期刊的数量为：

$R\bullet n\text{^}2=9.09\bullet 6.426\text{^}2=375.36\approx 375$

结合实际情况，将SCI-E和SSCI收录数字图书馆方面的论文之来源文献划分为三个区域，每个区域中的文献数量大致相同，分别为638、731和566，而期刊量之比为$10:62:371\approx 1:6:6^{2}$。

SCI-E和SSCI收录数字图书馆方面的论文其核心来源出版物[③]为RESEARCH AND ADVANCED TECHNOLOGY FOR DIGITAL LIBRARIES（Proceeding Papers）、ELECTRONIC LIBRARY（JCR Social Science Edition 2009其影响因子为0.544）、INFORMATION PROCESSING & MANAGEMENT（JCR Social Science Edition 2009其影响因子为1.783）、PROGRAM-ELECTRONIC LIBRARY AND INFORMATION SYSTEMS（JCR Social Science Edition 2009其影响因子为0.385）、DIGITAL LIBRARIES: INTERNATIONAL COLLABORATION AND CROSS-FERTILIZATION, PROCEEDINGS（Proceeding Papers）、JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY（JCR Science Edition 2009其影响因子为2.30）、RESEARCH AND ADVANCED TECHNOLOGY FOR DIGITAL LIBRARIES, PROCEEDINGS（Proceeding Papers）、ONLINE INFORMATION REVIEW（JCR Social Science Edition 2009其影响因子为1.423）、LIBRARY HI TECH（JCR Social Science Edition 2009其影响因子为0.272）、DIGITAL LIBRARIES: PEOPLE, KNOWLEDGE, AND TECHNOLOGY, PROCEEDINGS（Proceeding Papers）。

5 关键词分析

对所有文献关键词进行归并处理，得到4254个关键词，不重复关

③ 1935篇文献中，588篇是会议论文，其来源出版物为会议录及丛书（合计135种），这些在JCR里查不到影响因子。

表8 布拉德福分布图的计算项目

期刊量	累计期刊量C	lgC	载文量	累计载文量R(n)
1	1	0	149	149
1	2	0.30103	97	246
1	3	0.477121	59	305
1	4	0.60206	58	363
1	5	0.69897	55	418
1	6	0.778151	53	471
1	7	0.845098	44	515
3	10	1	41	638
3	13	1.113943	35	743
1	14	1.146128	32	775
1	15	1.176091	31	806
1	16	1.20412	27	833
2	18	1.255273	25	883
2	20	1.30103	22	927
1	21	1.322219	20	947
1	22	1.342423	19	966
1	23	1.361728	17	983
1	24	1.380211	15	998
1	25	1.39794	14	1012
1	26	1.414973	13	1025
4	30	1.477121	12	1073
4	34	1.531479	11	1117
1	35	1.544068	10	1127
4	39	1.591065	9	1163
7	46	1.662758	8	1219
4	50	1.69897	7	1247
12	62	1.792392	6	1319
10	72	1.857332	5	1369
18	90	1.954243	4	1441
38	128	2.10721	3	1555
65	193	2.285557	2	1685
250	443	2.646404	1	1935

键词1891个，选取出现频次超过10的关键词列入高频词，累计1508频次，占词频总数35.45%。

对高频关键词进行聚类分析，

表9 高频词分布

位次	关键词	数据总量
1	Digital library	349
2	System	72
3	Retrieval	66
4	Information	65
5	Information retrieval	57
6	Internet	53
7	Web	46
8	Design	46
9	Model	44
10	Library	38
11	Database	34
12	Search	28
13	Worldwide web	25
29	Interface	13
30	Service	13
31	Segmentation	13
32	Electronic publishing	13
33	Archives management	12
34	Search engine	12
35	Features	12
36	Information Seeking	12
37	Access	12
38	Framework	11
39	Project	11
40	Journals	11
41	Text	11

统计高频词组两两共同出现的次数建立共词矩阵，并应用Ochiia系数法对矩阵进行标准化处理。运用SPSS将标准化后的相异度矩阵载入，采用分层聚类法④，聚类结果见图3。

从图可以看出近20年国外数字图书馆研究热点主要集中在以下8

④ 聚类时选用组间连接法作为聚类方法，并选用欧几里德距离平方和进行距离测度。

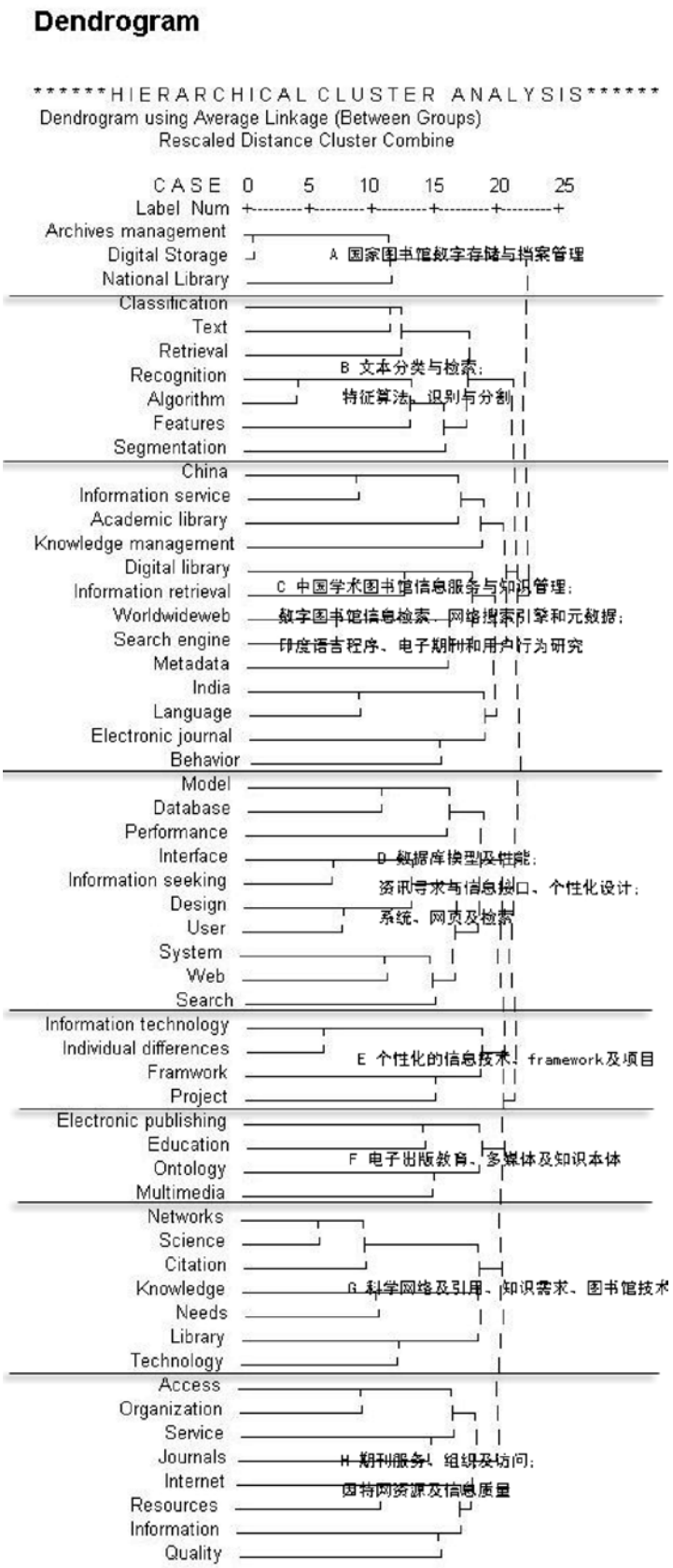

图3 龙骨图

方面：

（1）国家图书馆的数字存储与档案管理

（2）文本的分类与检索、文本特征、算法、识别以及分割

（3）中国的学术图书馆、信息服务及知识管理；数字图书馆信息检索、网络搜索引擎和元数据；印度计算机语言、电子期刊和用户行为研究

（4）数据库模型及性能；资讯寻求与信息接口、基于用户的设计；系统、网页和检索

（5）个性化的信息技术、framework及项目

（6）电子出版及教育、多媒体和知识本体

（7）科学网络及引用、知识需求、图书馆技术

（8）期刊服务、组织及访问；因特网资源及信息质量

根据各类团外部链接及其内部链接数，绘制如下类团关系图：点越大，其内部链接数越多，内部联系越紧密；连线越粗，类团间联系越密切。热点类团C和D是整个领域的研究核心，它们几乎与所有的类团都有着或强或弱的联系。

以向心度和密度为参数绘制战略坐标图[5]能更好地描述各研究热点类团的发展情况。类团C和D处于第I象限，其密度和向心度都处于较高水平，密度高代表研究主题内部联系紧密，研究趋向成熟，向心度高说明该类团代表的研究热点与其他各热点有广泛的联系，处于研究网络的中心；类团B处于第II象限，其密度较高，向心度较低，这些领域的研究已经形成了一定的研究规模，但是与其他类团联系不密

[5] 图5中，X 轴为向心度(Centrality)，Y 轴为密度(Density)，以向心度和密度的均值为原点。向心度用来测量一个类团和其他类团相互联系的程度，以每个类团与其他类团的链接的和作为该类团的向心度。密度用来测量类团内部词语之间的关联强度，以类团中每一对叙词在同一篇文献中同时出现的次数的平均值作为该类团的密度。

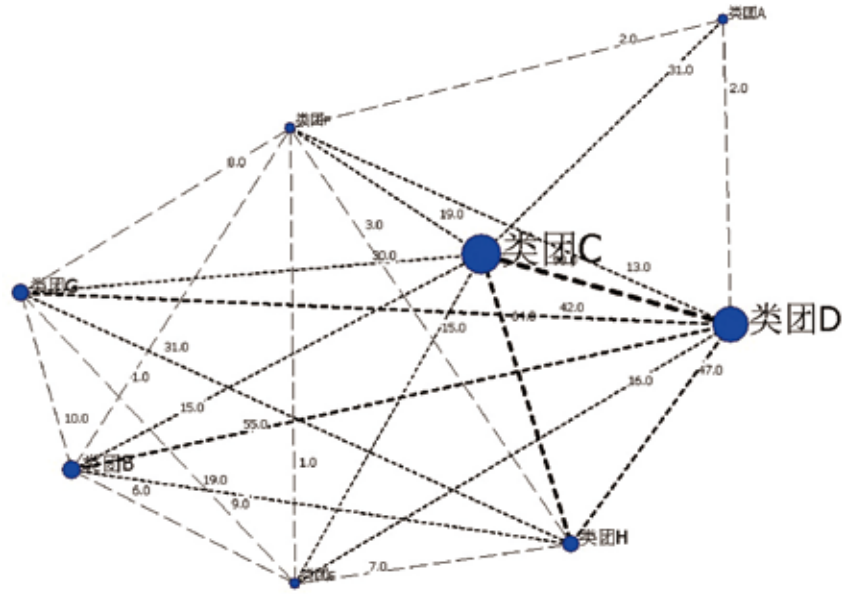

图4 类团关系图

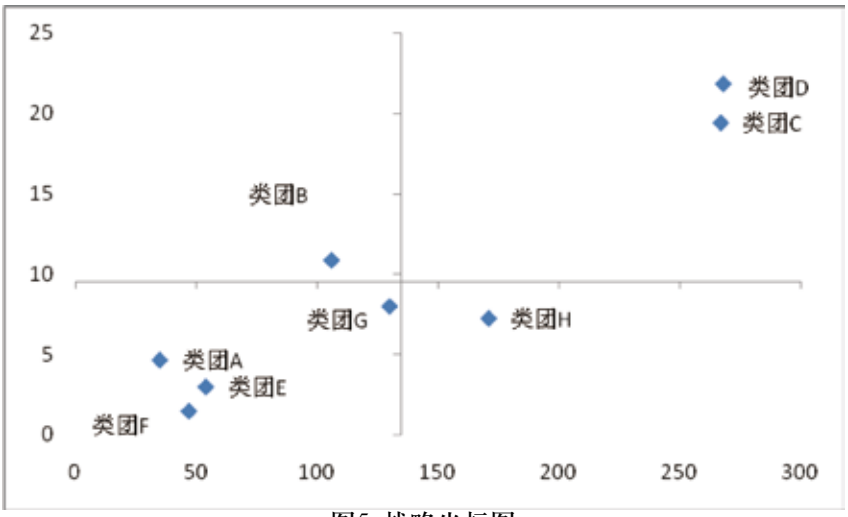

图5 战略坐标图

切，在整个研究网络中处于边缘位置；类团A、E、F和G处于第III象限，其密度和向心度都较低，是整个领域的边缘主题，内部结构比较松散，研究尚不成熟；类团H处于第IV象限，具有较高的向心度，处于核心领域，但密度较低，主题内部结构不够紧密，该领域的主题具有潜在的发展趋势。

6 不足

采用文献计量的方法较系统地分析了1987年以来SCI-E和SSCI收录数字图书馆领域文献的特征，未对论文被引情况进行分析。

参考文献

[1] 罗式胜. 文献计量学概论[M]. 广州:中山大学出版社,1994:41-86,296-325.
[2] 邱均平,王明芝. 1999-2008年国内数字图书馆研究论文的计量分析[J]. 情报杂志,2010,29(2):1-5.
[3] 刘金立,邵征翌,张健. 关于布拉德福定律的海洋科学学术论文分布研究[J]. 安徽农业科学,2009,37(14):6797-6798,6802.
[4] 钟旭,阎永胜. 洛特卡定律在合著者及全体著者中的验证研究[J]. 情报科学,2000,18(6): 564-565.
[5] 邱均平,丁敬达. 1999-2008年我国图书馆学研究的实证分析（下）[J]. 中国图书馆学报,2009(11):79-87.

作者简介

陈娟（1982-），女，馆员，厦门大学图书馆采访部工作，研究方向：数据分析。通讯地址：厦门大学图书馆采访部 361005。E-mail: chenjuan@xmu.edu.cn

A Metrological Analysis of Digital Library Research Papers in the World in Recent 20 Years: Based on SCI-E and SSCI of WOS

Chen Juan / Xiamen University Library, Xiamen, 361005

Abstract: The paper makes a metrological analysis of digital library research papers included by SCI-E and SSCI in recent 20 years and reveals the increase rule, core author list with its affiliation distribution, and core journal list. It collects 56 high-frequency keywords that reflect the focal points of digital library research and summarizes eight research structures in this field by co-word cluster analysis. Finally, it analyzes focal points trend through plotting strategic diagram so as to provide reference for research in the future.

Keyword: Lotka's Law, K-S Test, Bradford's Law, Co-word Cluster Analysis, Strategic Diagram

（收稿日期：2010-07-20）