

引用格式:Yang Mai,Wang Xuemei. A Bibliometrics Analysis of World Libraries' Papers based on GIS[J]. Remote Sensing Technology and Application,2015,30(4):819-824. [杨麦,王雪梅. 基于 GIS 的图书馆文献计量分析[J]. 遥感技术与应用,2015,30(4):819-824.]
doi:10.11873/j.issn.1004-0323.2015.4.0819

基于 GIS 的图书馆文献计量分析

杨 麦¹,王雪梅²

(1. 甘肃省张掖市甘州区图书馆,甘肃 张掖 734000;2. 西南大学图书馆,重庆 北碚 400715)

摘要:图书馆在提供文献收集、整理和查阅等功能的基础上,开展了大量研究工作并发表了丰富的研究成果。对发表的文献进行信息挖掘,可为相关研究提供参考和借鉴。通过趋势线分析获取各个国家图书馆在不同时期的发文趋势,采用质心法可以刻画出全球图书馆国际发文数量随着时间变化其重心发生迁移的特征。分析结果显示,这两种方法成功地应用到了文献计量分析中,为文献计量学研究提供了新的方法。

关键词:图书馆发文;文献计量;趋势分析;质心法;GIS

中图分类号:P 208.2 **文献标志码:**A **文章编号:**1004-0323(2015)04-0819-06

1 引言

图书馆是对图书、期刊、地图和典章等文献资料进行搜集、整理和收藏并供人阅览、检索和参考的机构^[1]。图书馆具有保存人类文化遗产的职能,很多图书馆历史悠久,早在公元前 2600 年就出现了最早的图书馆。随着现代计算机和网络技术的发展,图书馆电子化资源越来越多,读者通过网络就能很方便地获取信息资源。为支持图书馆服务功能的发展,许多图书馆针对图书馆建设、查询检索模式和文献计量学等方面开展了相关研究工作。通过对全球各图书馆发表的论文进行信息挖掘,分析这些开展研究工作的图书馆具有什么样的时空分布特征,随着新技术的出现呈现怎样的发展态势,以期对相关研究提供参考和借鉴。

文献计量学被定义为用数学和统计学的方法定量地分析一切知识载体的交叉科学^[2]。它是集数学、统计学和文献学为一体,注重量化的综合性知识体系,通过大量文献计量指标来实现发文和被引的数量和质量的定量化分析^[3]。20 世纪以来,迅速发

展的计算机技术和网络技术被广泛应用于文献计量学研究中,使其逐渐趋向于可视化、网络化^[4-5]。文献计量学对论文作者地址信息的分析多局限于单位名称及其发文数量,此外,对各单位所在城市的空间地理位置信息也可以进行分析,展示其时空分布特征^[6]。地理信息系统(Geographical Information System,GIS)是获取、处理、管理和分析地理空间数据的重要工具、技术和学科,近 10 a 来 GIS 在文献计量学研究和空间展现中得到广泛应用,如何将 GIS 的空间分析功能应用到文献发文和引文的时空分布特征中仍需做更深入的研究^[7-8]。

为了实现空间分析和展示,原始的空间信息必须从大量的科研文献中提取。一些信息具有规则结构,能够被快速地自动提取,例如文章的作者或合作者的单位、城市和国家等信息通常被规范地记录在元信息里,能够快速地检索。而摘要和全文中的地名、坐标和研究区等信息为不规则结构,提取相对困难^[9]。在地名信息提取方面,近年一些研究致力于发展更高效和自动的算法以实现地名信息识别、消歧和可视化

收稿日期:2015-03-19;修订日期:2015-07-11

基金项目:中央高校基本业务费专项资金项目(XDJK2015C146),中国科学院文献情报能力专项项目“基于知识流地理扩散的科研活跃中心动态变化监测”和西南大学图书馆科研项目(SWULIB2015A02)资助。

作者简介:杨 麦(1964-),女,陕西泾阳人,图书馆员,主要从事图书馆学与 GIS 研究。E-mail:421133488@qq.com。

通讯作者:王雪梅(1976-),女,重庆人,博士,副研究员,主要从事图书情报学与 GIS 研究。E-mail wxm@lzb.ac.cn。

等功能^[10]。总的来说,目前的地名识别率能够达到 80% 以上。将形式和内容丰富多样的地名信息与地图上的空间坐标、位置点和区域如何关联,是实现文献地名信息空间展现和分析的前提。虽然各国建有不同详细程度的地名空间数据库,但在书写格式、重名和详略程度上还存在很多问题,针对专题研究对空间数据的制备也是十分重要的步骤。

本文以全球图书馆发表的论文为研究对象,首先提取作者单位的国家信息,制备针对英文文献的全球国家地名数据库和空间数据,两者关联形成基础空间信息。在此基础上基于 GIS 空间分析功能挖掘发文的时空动态变化特征,为相关领域的研究人员和管理者提供科学参考依据。

2 数据与信息提取

2.1 文献数据

美国汤姆森路透(Thomson Reuters)公司 Web of Science(WOS)数据平台收录了世界各学科领域优秀的科技期刊和会议论文集,其收录的论文能在一定程度上及时反映科学前沿的发展动态。在 Web of Science 核心合集(Science Citation Index Expanded, SCI-E; Social Sciences Citation Index, SSCI; Arts & Humanities Citation Index, A&HCI; Conference Proceedings Citation Index-Science, CPCI-S; Conference Proceedings Citation Index-Social Science & Humanities, CPCI-SSH)中检索 1900~2014 年出版的通讯作者或第一作者(没有通讯作者情况下)发文地址中包含 Library 的文章,类型包括 Article、Proceedings Paper、Review 和 Letter,共有 42 054 篇。统计结果显示,在 20 世纪上半叶以前,全球图书馆的国际发文量很少,基本上在 100 篇以内,1973 年有一个较大的跃升,达到了 357 篇,1975 年达到 443 篇。图 1 显示了近 40 a(1975~2014 年)以来文章数量随着发表时间变化的情况,该时间段发文数量占整个时间段发文数量的 90.8%,因此本文重点针对这个时间段的论文进行分析。从中可以看出,前 10 a,发文数量处于持续增加的趋势,年均增长率在 53.6 篇/a,到 1985 年达到 1 030 篇。之后的 30 a 论文数量处于上下起伏的状态,总体有轻微上升的趋势,年均增长率为 3.2 篇/a。图书馆的国际发文量与一般研究领域的发文趋势有所不同,没有像其他领域一样在最近 30 a 出现快速的指数增长。分析其原因,现在各国图书馆的职能仍是以传播知识、服务当地或本机构的读者为

主,随着社会和科学的发展,大量图书情报研究在相关信息科学研究或评价机构中开展。



图 1 近 40 a 图书馆相关论文发表数量随时间变化的情况

Fig. 1 Variation with time of the libraries' papers in the last 40 years

2.2 国家地名库与信息提取

本文在国家的尺度上对发文作者地址信息进行提取,虽然 WOS 检索的文献记录作者单位地址书写相对规范,多数文献计量软件可以快速提取国家信息。但由于每个期刊的要求不同以及作者的书写习惯,出现各种可能性,需要进一步对国家地名信息进行规范。在地址信息提取前,需要建立用于规范地名的基础数据库。容易引起混淆的主要有 3 种情况,一种是有的作者地址把联合王国的下属地区作为国家名称写出来,例如苏格兰(Scotland)、北爱尔兰(North Ireland)和威尔士(Wales)都属于英国,荷兰安的列斯群岛(Netherlands Antilles)属于荷兰,在建国家地名库时,将这些地区名规范成所属国家名。第二种就是国家因统一或者解体等名称发生变化,例如西德(FED REP GER)和东德(GER DEM REP)合并,在建国家地名库时,用最终的国家名称德国(Germany),但解体的情况就比较复杂了,例如前苏联(USSR)和捷克斯洛伐克(Czechoslovakia),只能从地址中的城市信息来区分最终所属的国家名称,例如 Prague 和 Czechoslovakia 属于当前的捷克(Czech Republic),而 Bratislava 和 Czechoslovakia 则属于当前的斯洛伐克(Slovakia)。第三种主要是美国论文中的作者地址常没有国家名称,只有州名缩写和城市名称,例如 Berkeley CA 是美国加利福尼亚州伯克利市,Morehead KY 是美国摩海德市。本文建立了用于规范地名的国家地名数据库,从国家的空间尺度实现以上 42 054 篇论文发文单位的地址信息提取,在此基础上统计出各国图书馆在 1975~2014 年不同年份发文数量,形成属性表。

3 空间分析方法

3.1 发文趋势分析

针对发文数量,利用线性趋势线模拟其年际变化特征,斜率 SLOPE 也就是年均增长的数量,采用的方法为一元线性回归法,公式为:

$$\text{SLOPE} = \frac{n \times \sum_{i=1}^n i \times C_i - (\sum_{i=1}^n i)(\sum_{i=1}^n C_i)}{n \sum_{i=1}^n i^2 - (\sum_{i=1}^n i)^2} \quad (1)$$

其中: n 为模拟的时间段年数, i 为 1 至 n 的年序号, C_i 为第 i 年发文数量。

3.2 质心法

针对地理要素,几何中心(Geometric center)计算的是该要素形状所有点在二维空间上的算术平均,例如经常提到的中国的几何中心在兰州。基于此定义,可以先将世界各个国家的几何中心计算出来,将每个国家统计得到的图书馆年际发文数量关联到该中心上。为了分析这些几何中心发文量在全

球尺度上的总中心的迁移情况,引入物理学和统计学中常用的质心(Centroid)概念,计算全球图书馆发文的质心,并根据每年质心位置的变化来分析其迁移情况。地理要素空间分布的质心变化情况可以反映其在空间上的聚散与迁移随时间变化的过程^[11]。质心坐标的计算公式如下:

$$X_t = \sum_{i=1}^n (C_{ti} \times X_i) / \sum_{i=1}^n C_{ti} \quad (2)$$

$$Y_t = \sum_{i=1}^n (C_{ti} \times Y_i) / \sum_{i=1}^n C_{ti} \quad (3)$$

其中: X_t 、 Y_t 分别表示在 t 年全球发文的质心坐标, C_{ti} 表示在 t 年第 i 个国家的发文数量, X_i 、 Y_i 分别表示第 i 个国家的几何中心坐标。

4 结果分析

4.1 图书馆国际发文时空分布

图 2 显示了近 40 a 国家或地区图书馆发文量的空间分布情况,从中可以看出,图书馆国际发文的

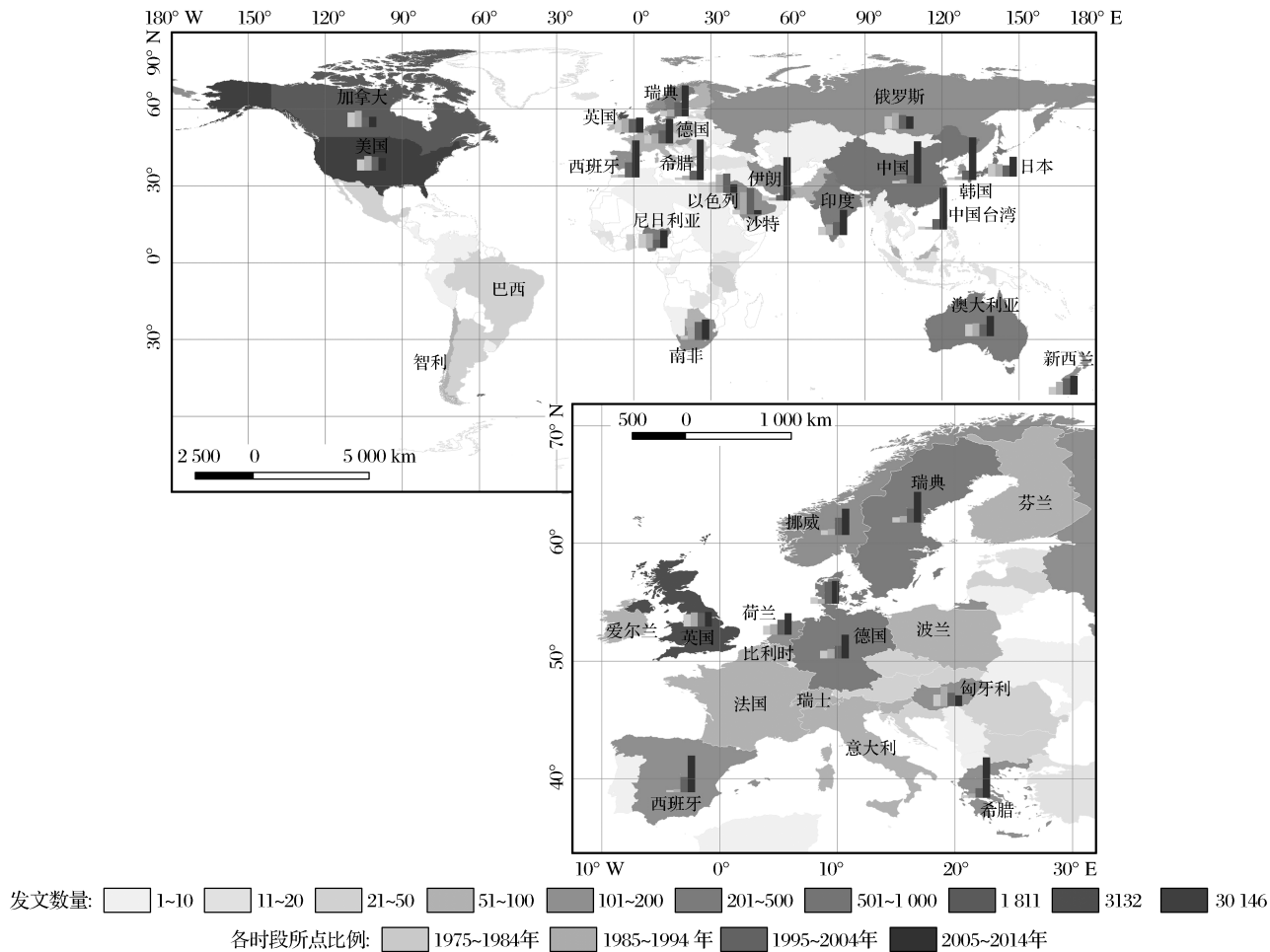


图 2 近 40 a 国家或地区图书馆发文空间分布图

Fig. 2 Spatial distribution map of the libraries' papers of the countries or regions in the last 40 years

空间分布差异很大,北美的美国和加拿大是主要的高值分布区。其次是欧洲地区,从图 2 的右下角放大图可以看出,英国、丹麦、德国、瑞士、俄罗斯、匈牙利、希腊、荷兰和西班牙等国家都有较高的发文数量。其次是亚洲地区,中国、日本、印度、中国台湾、韩国、伊朗、以色列和沙特等是发文数量较多的国家。澳洲的澳大利亚和新西兰也有较多的发文量。南美洲和非洲国家的图书馆发文量比较少。对发文较多的国家每 10 a 的论文量进行合计,分析发文较多国家在最近 40 a 的 4 个时间段里的发文比例的变化,从图 2 可以看出美国、加拿大、俄罗斯和匈牙利等国家 4 个时间段的发文比例接近,近年呈下降的趋势。而其他国家则出现上升趋势,其中,英国、澳大利亚、新西兰和南非等国家上升趋势相对缓慢,而德国、挪威和瑞典等国家上升趋势比较明显。中国及中国台湾、韩国、伊朗、希腊和西班牙等国家和地区,前期发文量很少,但在后两个时期或者最后一个时期发文比例非常高。

表 1 显示了近 40 a 图书馆国际发文前 20 位的国家或地区的发文数量与趋势,从中可以看出,美国

发文数量占据了绝对优势,达到 71.68%。排名第 2 位的英国只占 7.45%。通过计算前 20 a(1975~1994 年)和后 20 a(1995~2014 年)的发文趋势可以看出,多数国家的年均增长量都很小,其值在±0.2 篇/a 以内。年均增长数量最大的是美国前 20 a,达到每年 20.32 篇/a,但其在后 20 a 也达到了最高的年均递减量,为一7.33 篇/a。发文总数排名第 2 位的英国也是前期处于上升趋势,后期处于下降趋势,只是幅度均较小。加拿大则是两个时段均处于上升趋势,前一个时段上升更快一些。中国同样是两个时段均处于上升的趋势,但第一个时间段上升趋势缓慢,而在第二个时段上升趋势十分明显,达到 5.63 篇/a,这也印证了图 1 中不同时间段的发文比例特征。日本在第一个时段处于轻微下降的趋势,在第二个时段处于较快上升的趋势。印度的趋势和中国相似,在第一个时段上升趋势很缓慢,但在第二个时段上升趋势明显。在发文量超过 100 篇的国家中,前后 20 a 的发文趋势,丹麦、伊朗和日本相似,澳大利亚、德国和印度相似,俄罗斯、匈牙利、以色列、沙特和英国相似。

表 1 近 40 a 图书馆国际发文前 20 位的国家或地区发文数量与趋势一览表

Table 1 Numbers and tendencies of the libraries' papers of the top 20 countries or regions in the last 40 years

排序	国家/地区	发文总数/篇	百分比/%	前 20 a 发文趋势(篇/a)	后 20 a 发文趋势(篇/a)
1	美国 USA	30 146	71.68	20.32	−7.33
2	英国 England	3 132	7.45	0.97	−0.38
3	加拿大 Canada	1 811	4.31	0.62	0.14
4	中国 China	768	1.83	0.20	5.63
5	日本 Japan	514	1.22	−0.11	0.88
6	印度 India	452	1.07	0.28	1.20
7	丹麦 Denmark	378	0.90	−0.10	0.14
8	澳大利亚 Australia	357	0.85	0.20	0.46
9	尼日利亚 Nigeria	350	0.83	0.00	0.52
10	中国台湾 Taiwan	269	0.64	0.00	1.64
11	德国 Germany	247	0.59	0.12	0.44
12	瑞士 Sweden	213	0.51	0.03	0.69
13	俄罗斯 Russia	190	0.45	0.15	−0.17
14	匈牙利 Hungary	189	0.45	0.19	−0.14
15	韩国 South Korea	169	0.40	0.04	1.23
16	伊朗 Iran	149	0.35	−0.06	0.98
17	希腊 Greece	145	0.34	0.04	0.84
18	荷兰 Netherlands	139	0.33	0.06	0.08
19	以色列 Israel	134	0.32	0.24	−0.25
20	西班牙 Spain	124	0.29	0.02	0.50
21	南非 South Africa	121	0.29	0.20	0.04
22	沙特阿拉伯 Saudi Arabia	107	0.25	0.43	−0.56
23	新西兰 New Zealand	105	0.25	0.17	0.08
24	挪威 Norway	101	0.24	0.02	0.21

4.2 图书馆国际发文质心动态变化

图 3 显示了近 40 a 全球图书馆发文在国家尺度上的质心时空变化特征,从中可以看出质心的变化主要体现在经度位置上,总体变化范围介于 28°~82°W 之间,而纬度的变化范围很小,在 36°~41°N 之间。其原因就如前面分析的结果一样,全球图书馆发文主要分布在北半球国家,使质心的纬度处于北半球且变化不大。进一步分析可以看出,在 2003 年以前,质心的经度变化范围相对较小,变化范围介于 69°~82°W 之间,从放大后的图 4 可以看出,质心在这个范围内

来回变化,但总体上均靠近美国本土。说明这个时间段主要发文国家的发文数量比例差不多,通过统计也可以看出,美国在 2003 年之前发文数量占全球比例基本都在 70% 以上,最高为 1992 年占 79.10%。这段时期质心在小范围内东西摆动主要源于美国和欧洲之间的发文数量上下浮动的结果。但在 2004 年以后质心快速向东移动,在 2013 年达到最东边的经度值 28°W。分析其原因,是因为近 10 a 来亚洲地区的发文数量快速增加,而美国发文数量占全球的比例在下降,在 2013 年降到 52.25%。

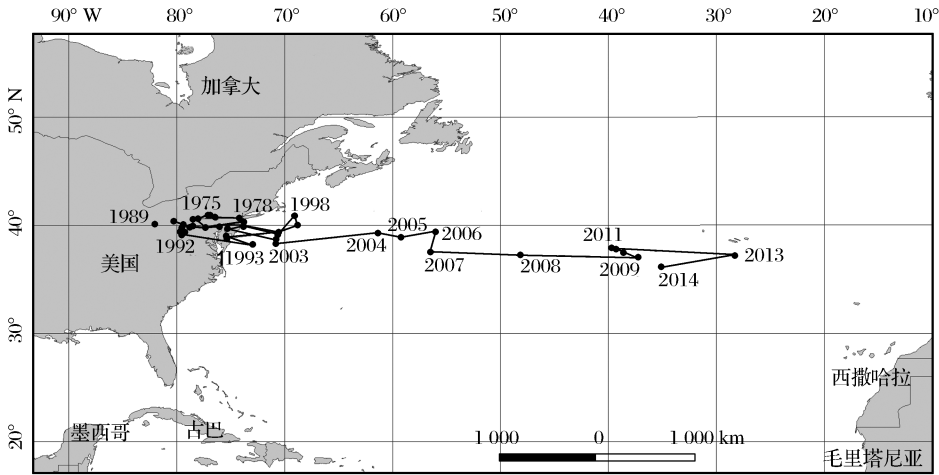


图 3 近 40 a 全球图书馆发文质心时空变化分布图

Fig. 3 Spatial and temporal changes of the centroids of the world libraries' papers in the last 40 years

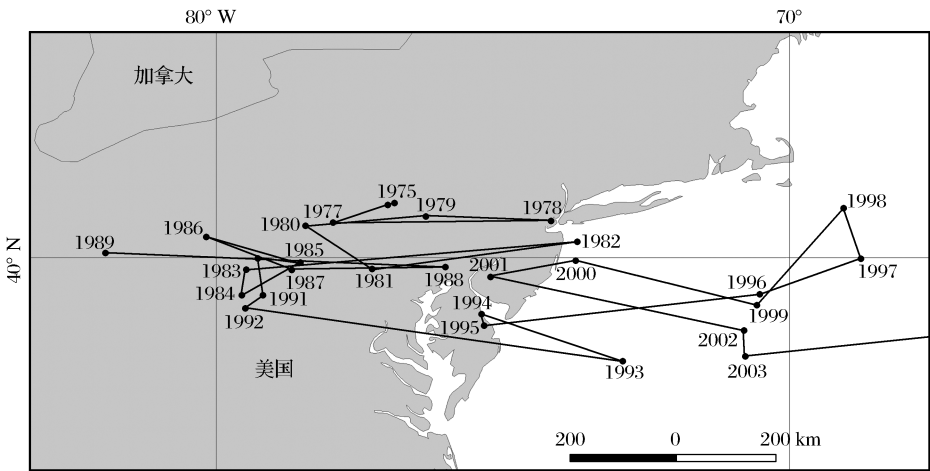


图 4 1975~2003 年国家 and 地区图书馆发文质心时空变化分布图

Fig. 4 Spatial and temporal changes of the centroids of the world libraries' papers during 1975 to 2003

5 结 语

从以上分析可以看出,本文将趋势线分析和质心法成功地应用到了文献计量学研究中,实现了世

界各国图书馆发文数量近 40 a 来的时空分布及变化特征的刻画,为文献计量学研究提供了新的方法论。通过趋势线分析,可以获得统计单元(此处为国家)在不同时间段发文数量的增长或者减少的趋势,

并定量地计算出年均增长或者减少的数量。而质心法可以细致刻画全球图书馆发文数量随着时间其重心发生迁移的特征,可以看出在前30 a美国和欧洲之间发文量的上下浮动导致质心小范围内东西摆动,但到了近10 a,由于亚洲国家发文数量快速增加,导致质心快速向东移动。

本文分析结果表明,图书馆的国际发文量没有像其他领域一样在近年出现快速增长趋势,国际图书馆研究的发文主体是美国,全球质心一直在美国大陆及其附近。可以考虑将该方法论应用到其他研究对象中,分析其时空变化特征。同时,针对图书馆的发文,本文正在尝试建立城市数据词典,提取发文地址的城市信息,这样可以更加深入地对某一个国家的发文动态进行分析,例如此处的美国可以作为分析的重点,如何高效地提取城市地址的信息将是待开展工作需要重点突破的地方。

参考文献(References):

- [1] Allen R E. The Pocket Oxford Dictionary of Current English [M]. Oxford: Clarendon Press, 1984.
- [2] Pritchard A. Statistical Bibliography or Bibliometrics[J]. Journal of Documentation, 1969, 25(4): 348-349.
- [3] Wang Xuemei, Zhang Zhiqiang, Ma Mingguo, *et al.* Bibliometric Indicators and Their Spatial Display Using Geographic Information System[J]. Library and Information Service, 2014, 58(3): 62-67. [王雪梅, 张志强, 马明国, 等. 文献计量学指标的 GIS 空间展现[J]. 图书情报工作, 2014, 58(3): 62-67.]
- [4] Börner K, Chen C, Boyack K W. Visualizing Knowledge Domains[J]. Annual Review of Information Science and Technology, 2003, 37: 179-255.
- [5] Tomas C A, Peter I. Informetric Analyses on the World Wide Web: Methodological Approaches to 'Webometrics'[J]. Journal of Documentation, 1997, 53(4): 404-426.
- [6] Frenken K, Hardeman S, Hoekman J. Spatial Scientometrics: Towards a Cumulative Research Program[J]. Journal of Informetrics, 2009, 3: 222-232.
- [7] Allen R S. Interdisciplinary Research: A Literature-based Examination of Disciplinary Intersections Using a Common Tool, Geographic Information System (GIS) [J]. Science & Technology Libraries, 2001, 21(3-4): 191-209.
- [8] Wang X M, Ma M G, Li X, *et al.* Applications and Researches of Geographic Information System Technologies in Bibliometrics[J]. Earth Science Informatics, 2014, 7(3): 147-152.
- [9] Wang Xuemei, Li Xin, Ma Zhicheng, *et al.* Development of the Bibliometric Geographic Information System on Open Source Platform[J]. Remote Sensing Technology and Application, 2014, 29(5): 891-896. [王雪梅, 李新, 马志程, 等. 基于开源平台构建文献计量地理信息系统[J]. 遥感技术与应用, 2014, 29(5): 891-896.]
- [10] Bensalem I, Kholadi M K. Toponym Disambiguation by Ambivalent Relationships [J]. Journal of Computer Science, 2010, 6: 653-659.
- [11] Bansal A R, Anand S P, Mita Rajaram, *et al.* Depth to the Bottom of Magnetic Sources (DBMS) from Aeromagnetic Data of Central India Using Modified Centroid Method for Fractal Distribution of Sources[J]. Tectonophysics, 2013, 603(9): 155-161.

A Bibliometrics Analysis of World Libraries' Papers based on GIS

Yang Mai¹, Wang Xuemei²

(1. Gansu, Zhangye Ganzhou District Library, Zhangye 734000, China;

2. Southwest University Library, Chongqing 400715, China)

Abstract: The basic functions of the libraries are literature collection, sorting and retrieval. Furthermore, the libraries have also made a lot of researches and obtain abundant achievements. We can get beneficial references by making the information mining from these literatures. The change tendency of the paper publications can be obtained for different countries by using trend line analysis. The spatial and temporal migration characteristics of the center of the world libraries' papers can be described by using the centroid method. The results indicate that the two methods are successfully used in the bibliometrics analysis, which afford the new methodologies for the bibliometrics studies.

Key words: Libraries' papers; Bibliometrics; Trend analysis; Centroid method; GIS