

· 业务研究 ·

基于 CSSCI 计量分析我国学科信息门户的发展 及语义网在学科信息门户中的应用研究

胡 珊

(南京大学信息管理学院, 江苏 南京 210093)

〔摘 要〕论文以 CSSCI 为统计源, 采用文献计量的方法, 通过对 2002—2010 年间我国学科信息门户研究论文的时间分布、期刊分布、著者分布、著者机构分布、引证分布等进行了统计分析, 对我国学科信息门户的发展历程进行有效地研究, 并就语义网技术应用于学科信息门户进行一定的探讨。

〔关键词〕学科信息门户; 主题网关; 学科门户; 主体门户; SIG; CSSCI; 定量分析; 文献计量统计; 语义网; 本体

DOI: 10.3969/j.issn.1008-0821.2012.07.028

〔中图分类号〕G250.252 〔文献标识码〕A 〔文章编号〕1008-0821(2012)07-0126-05

Research on Development Situation of Domestic Subject Information Gateway Based on Statistical Data of CSSCI and Sematic Web Applied in SIG

Hu Shan

(School of Information Management, Nanjing University, Nanjing 210093, China)

〔Abstract〕Taking the CSSCI Database as the source origin and using the bibliometrics methods, this paper made a statistical analysis about subject information gateway study published from 2002 to 2010 in China, including time, periodicals, authors, institution, indexing and citation distraction and study the development history of SIG in China. Then give a brief introduction about sematic web applied in SIG.

〔Key words〕subject information gateway; SIG; CSSCI; quantitative analysis; bibliometrics; sematic web; ontology

自网络技术诞生以来, 网上信息的数量呈现出几何倍数地增长, 搜索引擎和网络目录的出现虽大大方便了人们检索和利用信息, 但其搜索信息数量的过剩以及质量的参差不齐, 都给人们带来了诸多不便, 尤其是无法满足科研和学术机构人员对高质量学术信息的需求。学科信息门户的建立是解决这一问题的有效措施之一, 它借用了商业信息门户的概念和技术, 并借鉴和改进了一系列传统文献信息处理理论与技术, 对特定学科领域的网上信息进行深层组织^[1], 有效解决了网络信息资源在体积、质量和可信度等方面存在的突出问题, 为科研工作者提供权威和可靠的学科信息和服务。

1 学科信息门户

学科信息门户 (Subject Information Gateway, 以下简称 SIG), 又有学科门户、主题网关、主题门户等多种提法。其概念最早的提出者 Traugott Koch 认为: “SIG 是支持系统化资源发现的因特网服务, 通过因特网提供对资源 (文献、对象、网站或服务) 的链接。该服务建立在资源描述的基础之上, 可以通过主题结构浏览资源是其重要特征。”^[2]

SIG 的概念在之后的发展中不断地完善, 鉴于当前的发展情况, 笔者认为: SIG 是针对特定学科或主题领域, 按照一定的资源选择和评价标准以及规范的资源描述和组

收稿日期: 2012-04-03

基金项目: 本文的研究受教育部人文社会科学研究规划基金项目“数据关联的语义数字图书馆研究”(项目号: 10YJA870014) 资助。

作者简介: 胡珊 (1989-), 女, 研究生, 研究方向: 语义网框架研究。

织体系,对具有一定学术价值的网络资源(网站、文献、工具或服务)进行搜集、描述和组织^[3],并提供检索、浏览、导航等基本功能,此外,SIG的扩展功能应包括提供一些自建资源(如全文文献、特色馆藏库等)、RSS、学术博客、学术论坛等增值服务以及个性化服务(如信息推送、信息定制等)。

本文以CSSCI为统计数据源,通过对我国学科信息门户相关论文在时间、期刊、著者、著者机构分布及引证关系等方面进行统计分析,以了解我国学科信息门户的发展特点与趋势,为进一步研究提供参考。

2 学科信息门户结果统计分析

本文的统计数据源来自于CSSIC(中文社科引文索引来源期刊)。CSSCI由南京大学中国社会科学研究评价中心主持评选,其遵循文献计量学规律,采取定量与定性评价相结合的方法从全国2700余种中文人文社会科学学术性期

刊中精选出学术性强、编辑规范的期刊作为统计源期刊。

笔者于2011年10月25日,以“学科信息门户”、“学科门户”、“主题网关”、“主题门户”等为检索词,对1998—2011年间的数据库进行检索,共检出SIG研究领域相关文献79篇,并据此进行文献计量学分析。

2.1 年代分析

某一时期的论文数量的变化速度和数量,在一定程度上说明了这一领域学术研究的理论水平和发展速度,也能在一定程度上反映研究者和实践者对该领域的重视程度。

表 1 年代分布

| 年 份 | 发文量(篇) | 年 份 | 发文量(篇) |
|------|--------|------|--------|
| 2002 | 1 | 2007 | 10 |
| 2003 | 1 | 2008 | 7 |
| 2004 | 12 | 2009 | 7 |
| 2005 | 17 | 2010 | 4 |
| 2006 | 19 | | |

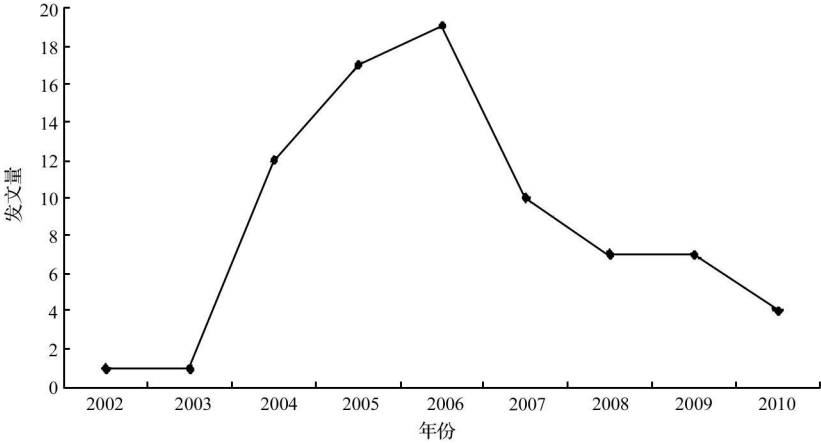


图 1 研究论文随时间分布趋势图

从表1和图1可以看出,按照各年的发文量进行统计,我国学科信息门户研究大致可以分为3个小阶段。

2.1.1 2002—2003 年

虽然笔者选取了1998—2011年间的所有数据库,但我国关于SIG研究的第一篇文章出现在2002年,由张晓林撰写关于国家科学数字图书馆分布式SIG及学科信息导航系统建设的研究。可以说这篇文章开启了我国对SIG的研究进程,对以后的研究起到了非常深远的影响。但在第1篇文章发表后的1年时间里,并没有很多学者对于SIG开展有效的研究,从以上数据可以看出,2003年也仅有1篇关于SIG的文献,由张智雄撰写对主要目录协议在主题网关中的应用进行了探讨^[4]。虽然国外从20世纪90年代就开始了SIG的建设与使用,但引进我国的时间是21世纪初。张晓林虽然对SIG进行了介绍,但是由于国内还没有正式

开始规划和组建SIG的项目,所以,没有引起学界的重视。

2.1.2 2004—2006 年

从2004年起,连续3年呈现出快速增长趋势,在2006年发文量已达到19篇,创年度发文量最高纪录。从2004年开始,由国家科学数字图书馆规划建设5个SIG,并聚集大量人力、物力等资源进行具体的落实,因而激起了全国范围内建设SIG的热潮,其中包括武汉理工大学、中国科学院武汉文献情报中心、国家科技图书文献中心等机构也纷纷加入建设队伍。在建设过程中,对于SIG的相关理论与实践研究也快速发展起来。2004年,国科图的5个SIG以及其他机构建设的SIG正式启用,研究人员在对SIG的评价,与国外SIG的对比,对国内SIG的调研等研究也陆续开展,这些原因都促使2004—2006年期间文献量迅速增长。

2.1.3 2007 年一至今

从 2007 年开始，学界的研究热情逐渐减退，发文量也呈现出下滑趋势。笔者认为，首先，国内学者都喜欢跟风，一段时间内会出现对一个问题全方位、多角度的研究，研究成果很快覆盖了该问题几乎所有的内容，导致之后的研究难以找到突破口和创新点。其次，由于 SIG 研究的逐渐成熟，研究核心从介绍性的理论研究转向评价与维护、平台构建与技术创新等方面，这些内容要求研究人员对 SIG 有更深层次的探索。再者，我国 SIG 在建成之后的利用效果非常不令人满意，很多 SIG 从建成后几乎一直没有更新，用户访问量和使用率都极低，二者应该是互为因果的关系，所以，影响了学者的研究热情。

2.2 期刊分布

对某一领域研究论文的期刊分布情况进行分析，不仅可以确定该领域的核心期刊，为文献的收集管理者提供依据，为读者提供指导，还可以了解各刊物的办刊特点，指导刊物发展^[5]。经调查统计，2002 年来，研究人员在 17 种刊物上发表 SIG 论文共 79 篇，平均每种期刊发表约 4.6 篇。

从表 2 可以看出，图书情报工作、情报理论与实践、图书馆杂志与情报杂志发表文献量较多，这说明这 4 种期刊是研究 SIG 的核心期刊，为以后学者研究该课题或是发表该类文献提供了指南作用。同时，还能看出，除了科技进步与对策外，其余 16 种都属于图书馆、情报与文献类期刊。由于在 SIG 的建设过程中涉及到信息采集、信息组织以及质量控制等工作都是图情领域的核心价值，所以，得到了图情界学者的广泛关注。

| 表 2 期刊分布 | | | | | |
|----------|----------|-----|----|----------|-----|
| 序号 | 刊物名称 | 载文量 | 序号 | 刊物名称 | 载文量 |
| 1 | 图书情报工作 | 13 | 10 | 中国图书馆学报 | 3 |
| 2 | 情报理论与实践 | 9 | 11 | 图书馆工作与研究 | 3 |
| 3 | 图书馆杂志 | 8 | 12 | 图书情报知识 | 3 |
| 4 | 情报杂志 | 8 | 13 | 图书馆论坛 | 2 |
| 5 | 现代图书情报技术 | 6 | 14 | 图书馆 | 2 |
| 6 | 大学图书馆学报 | 5 | 15 | 图书馆论坛 | 2 |
| 7 | 情报科学 | 5 | 16 | 科技进步与对策 | 1 |
| 8 | 情报资料工作 | 4 | 17 | 情报学报 | 1 |
| 9 | 图书馆理论与实践 | 4 | 总计 | | 79 |

2.3 著者分析

2.3.1 著者发文分布

对著者群的统计分析，能反映出文献作者与文献量的关系，同时能预测、估计和揭示科学研究人员的研究能力，对进一步了解 SIG 研究领域的现状和发展趋势都具有重要

的意义。自 2002 年以来，共有 108 位作者发表了有关 SIG 的文献，而文献总量为 79 篇，可以看出合著率较高。发表 1 篇文章的作者占总数的 82.4%，发表 2 篇文章的作者占总数的 10.2%。

从表 3 可以看出，共有 8 位作者发文量在 3 篇及 3 篇以上。发文量最多的是 5 篇，分别是中国科学院资源环境科学信息中心的祝忠明、肇庆学院图书馆的李育嫦以及武汉大学信息管理学院的胡明玲。从发文时间以及发文内容分析可以得出，2003 年，国科图开启了资源环境学科信息门户建设项目，祝忠明、吴新年等中国科学院资源环境科学信息中心的研究人员参与其中，伴随着项目的推进而合著了多篇有关 SIG 的文献。同样，胡明玲和武汉理工大学图书馆员王建涛也背靠武汉理工大学建设 SIG 的机遇，深入研究 SIG，合著了多篇文献。侯汉清是我国图情界著名的专家学者，现任南京农业大学信息科技学院教授，在表 2 中，提及的惟一一篇被情报学报收入的《学科门户智能搜索引擎的设计与构建——农业史门户搜索引擎的实验》即出自侯教授笔。

| 表 3 著者分布 | | | | | |
|----------|-----|-----|----|-----|-----|
| 序号 | 作者 | 发文量 | 序号 | 作者 | 发文量 |
| 1 | 祝忠明 | 5 | 5 | 王建涛 | 4 |
| 2 | 李育嫦 | 5 | 6 | 侯汉清 | 4 |
| 3 | 胡明玲 | 5 | 7 | 雷菊霞 | 3 |
| 4 | 吴新年 | 4 | 8 | 韩丽 | 3 |

2.3.2 著者机构分布

自 2002 年来，共有 58 个机构发表了有关 SIG 的文献，发表 1 篇文献的机构占总数的 72.4%。发表 2 篇文献的机构占总数的 13.8%。从表 4 可以看出，武汉大学信息管理学院发文量高达 16 篇，其次是中国科学院资源环境科学信息中心，发文量 5 篇。紧跟其后的是肇庆学院图书馆、武汉理工大学图书馆、武汉大学信息资源研究中心，发文量均为 4 篇。由此可知，武汉大学和中国科学院系统是研究 SIG 的重要机构，值得关注。

2.3.3 著者机构类型分布

学界一直在探讨，SIG 的建设应该由图书馆牵头还是由高校院系牵头。在对著者机构类型进行统计，结果显示有 53 篇文献出自高校院系，有 49 篇文献出自图书馆，因此，可以看出高校院系和图书馆都对 SIG 开展了有效的研究，所以，SIG 的构建不应该由一个机构单独完成，而应该是院系与图书馆合作开展，分别发挥自身的优势，实现一加一大于二的成效。

表 4 著者机构分布

| 序号 | 单位名称 | 发文量 |
|----|-----------------|-----|
| 1 | 武汉大学信息管理学院 | 16 |
| 2 | 中国科学院资源环境科学信息中心 | 5 |
| 3 | 肇庆学院图书馆 | 4 |
| 4 | 武汉理工大学图书馆 | 4 |
| 5 | 武汉大学信息资源研究中心 | 4 |
| 6 | 中国科学院文献情报中心 | 3 |
| 7 | 南京农业大学人文学院 | 3 |
| 8 | 北京师范大学图书馆 | 3 |

2.4 引证分析

2.4.1 被引情况分析

由于本文的统计源是 CSSCI，CSSCI 的来源期刊质量均较高，故相关文献被引次数较高是普遍存在的特点。本文研究的这 79 篇文献的被引次数也符合以上规律。从统计数据可知，被引次数在 10 次以上的文献占总数的 41.8%。从表 5 可以看出，在 79 篇文献中，有 14 篇文献的被引量达到 20 次及以上。被引次数最高的是张晓林的《分布式学科信息门户中网络信息导航系统的规范建设》一文，高达 143 次，这正是笔者在年代分布部分提及过的，我国第 1 篇发表的有关 SIG 的文献，该文献不仅发表时间最早、质量高，而且文中描述了 SIG 中的资源选择、搜寻策略、元数据、知识组织体系、维护使用管理机制等内容^[6]，都是之后 SIG 研究的重点，所以，被引次数非常高。其次，近几年，web2.0 毫无疑问成为各领域都高度关注的话题，研究 SIG 的学者也不例外。虽然王晓光基于 web2.0 在 SIG 应用的文献发表于 2007 年，但被引次数高达 50 次。随后，王玮、韩丽以及王楠等文献被引次数都较高，这些文献大都得到了各类基金项目的赞助。因此可知，在基金项目的赞助下，不仅容易出成果，而且成果质量普遍较高、参考价值大。

2.4.2 引用情况分析

对发布文献的引用情况进行统计，结果显示对 SIG 研究的作者引用了大量的参考文献，79 篇文献一共引用了 985 篇相关文献，平均每篇都引用了 12.47 篇。这与以上统计的被引次数高有着密切的联系，因为引用与被引就是一对相互影响的关系体。

在 79 篇文献中，徐佳宁和赵洋合著的《国外学术信息主题网关特点分析》，引用 67 篇，引用篇数最多；其次，何琳和陈焕之合著的《我国学科信息门户研究综述》，引用 47 篇；随后，祝忠明、吴新年、孙成权等合著的《资源环境学科信息门户的研究与建设》，引用 35 篇。

表 5 被引情况

| 序号 | 篇 名 | 第一作者 | 被引次数 |
|----|-------------------------|------|------|
| 1 | 分布式学科信息门户中网络信息导航系统的规范建设 | 张晓林 | 143 |
| 2 | 基于 Web2.0 的学科信息门户设计与实现 | 王晓光 | 50 |
| 3 | 网络信息资源组织的新模式——主题网关 | 王 玮 | 32 |
| 4 | 学科信息门户中的信息推送服务 | 韩 丽 | 30 |
| 5 | 学科信息门户的资源选择与组织 | 常 唯 | 30 |
| 6 | 学科信息门户评价指标体系探讨 | 王 楠 | 29 |
| 7 | 学科信息门户：概念、结构与关键技术 | 孔 敬 | 27 |
| 8 | “看不见”的网站与学科信息门户的比较分析 | 马费成 | 25 |
| 9 | 学科信息门户信息组织的优化 | 黄如花 | 24 |
| 10 | 主题网关关键技术研究 | 李运景 | 24 |
| 11 | 学科信息门户元数据格式的选择与设计 | 祝忠明 | 23 |
| 12 | 面向用户的学科信息门户构建模式分析 | 向 菲 | 22 |
| 13 | 国外学科信息门户发展现状分析 | 张一芮 | 21 |
| 14 | 专业博客与学科门户的互动分析 | 孔 军 | 20 |

3 总 结

本文以 CSSCI 为数据源，对我国学科信息门户的发展研究进行了文献计量学统计工作，以了解我国学科信息门户的发展历史及进程。最终检索出 79 篇相关文献，对其进行统计得出，我国学科信息门户的发展主要分为 3 个阶段：2002—2003 年的萌芽期，2004—2006 年的快速成长期，2007—2010 年的缓慢发展期，其发展特点完全符合对一个新事物的认识发展过程。从期刊的载文量来看，呈现出集中与分散的特点，具体表现在，这 79 篇文献分散于 17 种不同的期刊上，但数量排名前 3 的期刊载文量占总量的 48%。从学科分布来看，在这 17 种期刊中，除了《科技进步与对策》外，其余的 16 种都是图书馆、情报与文献学类期刊，这表明图情领域是我国学科信息门户研究的中心。从著者发文中了解到，这 79 篇文献涉及 108 位作者，可见合著率较高，发表 1 篇及 2 篇的作者占比分别为 82.4%、10.2%，其中有 8 位作者发文量在 3 篇及 3 篇以上。从著者机构的统计结果看，共有 58 个机构发表了相关论文，其中发表 1 篇和 2 篇的机构占比分别为 72.4%和 13%，发表 3 篇及 3 篇以上的共有 8 个机构，这些机构基本是发文量较高的作者所在单位，比较突出的是武汉大学和中国科学院系统，说明这两家机构在我国学科信息门户的研究中占有重要地位，是该领域研究的主要阵地，代表着该领域研究的最高水平。由于 CSSCI 来源期刊的质量较高，被收录文献被引频次都较高，被引次数在 10 次以上的文献占总数的 41.8%。其中，有 14 篇文献的被引量达到 20 次及以上，而

张晓林在国内发表的第一篇有关学科信息门户的文章被引量达到了 143 次。其次，各类基金项目赞助的科研成果也成为被引热文。由于被引与引用是一对相互影响的关系体，在如此高的被引量下，这 79 篇文章的引用量也比较高，平均每篇引用 12.47 篇，引证关系的统计表明，在我国学科信息门户的研究中，借鉴较多、参考频繁。

总的来说，在近十年的发展中，我国学者对 SIG 的研究取得了一定的进展，不仅在理论上有所突破，也建成了一些颇具代表的 SIG。但从维护以及使用情况来看，这些 SIG 运行都不太理想。既然学界已经认识到 SIG 对研究人员来说的重要性，就应该加强对 SIG 的建设与维护。应该加强图书馆和高校院系的合作，搭建出优秀的 SIG，建立起有效的维护机制，加强宣传等举措，更好地为研究人员的研究工作服务。

4 语义网技术在学科信息门户中的应用

虽然学科信息门户能够有效解决网络信息资源在体积、质量和可信度等方面的问题，但要想提供更高效的检索效率，就需要把语义网技术应用于学科信息门户。近年来，语义网逐渐成为 WWW 领域的研究热点以及智能化网络服务和应用开发中的关键技术之一。利用语义网技术可以为信息创建元数据，这些元数据可以被机器理解，然后利用本体对不同信息源中的元数据进行查询和推理，找出用户需要的信息^[9]。

2006 年我国有 2 篇文献分别从信息组织和信息查询两个维度把语义网技术应用于学科信息门户。冯玉和雷菊霞在论文中提及，由于学科信息门户的分类法和主题词表的局限性，所以需要扩展多变量、多层次、多功能的关系列表体系来组织学科信息，以强调表现概念之间的关系，语义网就是关系列表体系的代表。通过语义网来构建本学科领域的本体，利用其对学科信息资源内容进行语义标注或语义挖掘，形成基于语义的资源元数据，用于表述十分复杂的事物间的相互关系的特定概念模型^[10]，以实现门户知识的共享。但文章仅仅是提出此概念，并没有对如何实现或者实际案例进行阐释，有很大的局限性。吴浩东和刘强提出，用户查询信息有两种情况：能清楚表达需求信息和不能清楚表达需求信息。针对这两种情况，提出一种基于本体的信息查询模型，对于能清楚表达需求的用户，采用基于领域本体的改进搜索方法；对于不能清楚表达需求的用户，采用基于模糊本体的逐步精确查询方法^[9]。文章对

两种方式具体的实现过程进行了阐释。胡明玲、王建涛以宁波市数字图书馆服务外包产业信息门户关联数据应用为例，研究了关联数据发布与应用的架构，并以专业人才关联数据发布为例分析了关联数据发布和应用的实现方法^[11]。关联数据作为语义网的一个分支，其不仅能在系统内部很好地与对象信息资源有关的相关数据进行统一展示，还使企业能在自己系统中应用这些数据，拓展数据的利用方式的同时，支持企业的信息化建设，这对于企业或者图书馆来说，都具有重大的意义。

随着语义网研究的深入，语义网技术被应用到多个学科领域，从学科信息门户来看，融入语义网技术后，能够提供更有效的检索，能为用户提供更好的服务，但就我国目前的情况来看，还没有太多成熟的语义网应用，所以，应该重视语义网技术，并从理论研究上升到实践中，更好地为用户提供优质、有效的服务。

参 考 文 献

[1] 张芮. 学科信息门户建设研究 [D]. 成都: 四川大学, 2006.

[2] Quality—controlled subject gateways: definitions typologies empirical overview [EB/OL]. <http://chinesesites.library.ingentaconnect.com/content/mcb/264/2000/00000024/00000001/art00002/>, 2011—10—26.

[3] 王美琴. 我国学科信息门户可持续发展研究 [D]. 南京: 南京农业大学, 2007.

[4] 张智雄. 目录及其在分布式主题网关中的应用 [J]. 现代图书情报技术, 2003 (5): 41—44, 88.

[5] 李利, 唐淑香. 基于文献计量的国内文献传递服务研究综述 [J]. 新世纪图书馆, 2011, (1): 38—41.

[6] 张晓林. 分布式学科信息门户中网络信息导航系统的规范建设 [J]. 大学图书馆学报, 2002, (5): 28.

[7] 陈晓毅. 学科信息门户研究的文献计量分析 [J]. 现代情报, 2007, (4): 164—165, 167.

[8] 何琳, 陈焕. 我国学科信息门户研究综述 [J]. 图书馆工作与研究, 2010 (12): 9—13, 28.

[9] 吴浩东, 刘强. 一种信息门户中基于本体的信息查询模型 [J]. 计算机工程, 2006, (16): 46—47.

[10] 冯玉, 雷菊霞. 网络时代学科信息门户的信息组织 [J]. 图书馆理论与实践, 2006, (2): 61.

[11] 胡明玲, 王建涛. 关联数据特点及发布研究 [J]. 图书馆界, 2011, (6): 4—6.

(本文责任编辑: 孙国雷)