

Thomson Data Analyzer 在机构专利情报分析中的应用研究

冯浩然^{1,2},方 曙¹

(1.中国科学院 国家科学图书馆 成都分馆,四川 成都 610041;
2.中国科学院 研究生院,北京 100049)

摘要:使用 Thomson Data Analyzer TDA(TDA)软件对 1986-2005 年期间在 Derwent Innovation Index DII(DII)中收录的德国马普学会(MPG)、美国国立卫生研究院(NIH)、美国斯坦福大学(Stanford)、美国康奈尔大学(Cornell)、美国威斯康辛州立大学(Wisconsin)、日本东京大学(Tokyo)六家机构的专利数据,从专利数量、发明技术领域、机构对外合作情况、专利家族国家分布和技术领域相似度 5 个方面进行了比较分析。

关键词:专利分析;技术领域;合作情况;专利家族

中图分类号:G350 **文献标识码:**B **文章编号:**1007-7634(2008)12-1833-04

Research on the Application of Thomson Data Analyzer to Analyses the Patent Intelligence of Scientific Institutions

FENG Hao-ran^{1,2}, FANG Shu¹

(1. Chengdu Branch of National Science Library, Chinese Academy of Sciences, Chengdu 610041, China; 2. Graduate University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: This paper utilizes Thomson Data Analyzer to analyze the patent data, gathered in Derwent Innovation Index, of six organizations: Max Planck Society, National Institution of Health, Stanford University, Cornell University, University of Wisconsin-Madison and Tokyo University. Data of them are compared in patent quantity, area of the technical innovation, external cooperation, patent family country distribution and technical area similarity.

Keywords: patent analyze; technical area; cooperation; patent family

1 引 言

TDA (Thomson Data Analyzer) 是美国 Thomson 公司开发的一款专利分析软件,它具有强大的专利数据挖掘功能,能够对专利数据进行可视化操作。它提供了从数据清洗、比较矩阵、数据图谱到自动报告生成等一系列功能。它能够从专利授权组织、专利技

术领域分布、机构合作情况、专利家族、发明人产出情况等 10 多个不同方面对专利数据进行分析。

德温特世界专利创新索引 (Derwent Innovation Index, DII) 收录了自 1963 年以来全球 40 多个专利机构的 1300 万条基本专利,2000 万项专利。是最具有权威的专利数据库之一,其也是本文中数据采集的数据源。

本文选取了德国马普学会(MPG)、美国国立卫

收稿日期:2008-06-18

作者简介:冯浩然(1984-),男,四川内江人,硕士研究生,从事专利情报分析研究;方 曙(1957-),男,四川人,研究员,博士生导师,从事情报计量学理论与实践研究。

生研究院(NIH)、日本东京大学(Tokyo)、美国斯坦福大学(Stanford)、美国康奈尔大学(Cornell)和美国威斯康辛州立大学(Wisconsin)六家机构作为分析对象。其中 Stanford、Cornell、Wisconsin 均为 2005 年美国大学专利申请数量排名前十的大学^[1]。MPG 和 NIH 则是国际最为著名的科研机构,其地位不言而喻。而 Tokyo 则是作为亚洲机构的代表进行研究。本文检索的专利数据时间范围为 1985 年-2005 年,使用专利权人名称和专利权人代码在 DII 中进行检索,获得了本文的数据样本。样本采集时间为 2008 年 1 月 10-11 日^[2]。将检索得到的数据导入到 TDA 中以后,采用自动和人工的方式对数据进行了清洗,然后利用 TDA 对数据进行分析^[3]。

2 专利数据分析

本文主要选择了从公开专利数量、专利技术领域分布、机构相似度、对外合作情况和专利家族数量 5 个方面对上述六家机构的专利情况进行了分析^[4]。这 5 个方面可以比较全面的反映机构的专利情况和技术创新能力。

2.1 公开专利数量及其年增长率比较分析

如图 1 中所示,1986-2005 年间六家机构专利总量大致可以分为两类:专利公开数量最高的 NIH,达 3212 件;另外 5 家机构专利公开数量都处于 1000-2000 件之间。当中最低的为 Tokyo,有 1052 件。专利公开数量最高与最低机构之间相比为 3.05 倍。

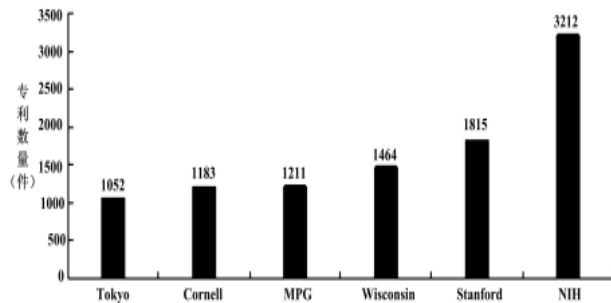


图 1 六家机构 1986-2005 年专利总量

从公开的专利总数看,1986-2005 年期间六家机构专利数量变化趋势呈现出多种态势。Cornell、Stanford、Wisconsin 和 MPG 处于一个缓慢上升通道中,尽管其间有些年份有负增长的情况,但是波动幅度不大,总体上还是呈现逐渐增长。NIH 的专利分布情况类似于双峰分布情况,在 1991 年和 1999 年达

到了两个峰值。而 Tokyo 则波动幅度非常剧烈,在 2003 年之前,年专利授权数量一直都维持在一个非常缓慢的上升水平。而从 2004 年开始急剧变化,较 2003 年增长了 150 件,2005 年又较 2004 年增长了 250 件,在两年之内一共增加了 400 件。

从整体上来看 MPG、Cornell、Stanford、Wisconsin、NIH 和 Tokyo 这六家机构在 1986-2005 年 20 年间,保持了平均每年 6.11%、3.89%、9.29%、6.85%、16.13%、64.75%的年度专利成长率,都呈现逐年增长的趋势。其中年均增长率最高的是 Tokyo,达到了 64.75%;年均增长率最低的是 MPG,为 3.89%。年均增长率最高与最低机构之间相比为 18.75 倍。尽管部分机构在其中有些年份专利公开数量有较大的波动,但是从长期来看还是保持了增长的趋势。

从公开专利数量及其年增长率的趋势中可以看出,Tokyo 正处在一个专利申请的快速成长期,专利公开数量急剧增加,表现出了一个强劲的增长态势。而欧美机构的专利公开数量总体维持在一个缓慢增长的水平,没有太激烈的变化。

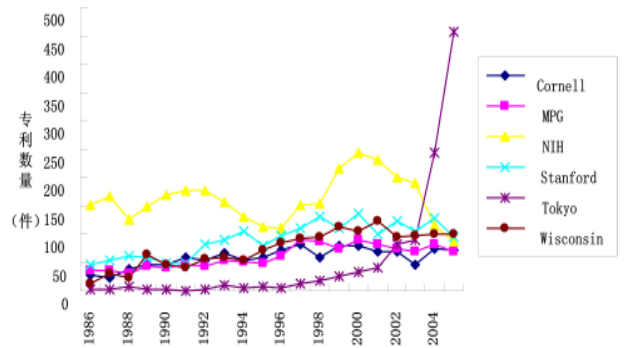


图 2 六家机构 1986-2005 年专利变化趋势

2.2 发明专利家族情况

运用 TDA 的比较矩阵计算得到了六家机构基本专利公开年的专利家族国家数统计数据,见图 3。在专利家族国家总数上,排在第一位的 NIH 有 39 个国家,第二位的 MPG 有 38 个国家,Wisconsin 和 Stanford 以 35 个国家并列第三位,最后两位为 Cornell 和 Tokyo,分别为 28 和 20 个国家。第一位的 NIH 约为最后一位的 Tokyo 的 2 倍。

结果表明,从 1986 年以来各个机构专利家族的覆盖范围不断扩大。欧美机构从 1986 年平均 13 个专利家族国家,到 2005 年达到了平均 25 个专利家族国家,数量增长接近了 1 倍。而 Tokyo 的专利国家数量在 20 年间提升接近了 2 倍,从 1986 年的 6 个达到了 2005 年的 17 个。相比欧美机构,Tokyo 在专

利家族国家数量增长速度更快,势头更加强劲。

从上面数据可以看出,各个机构的专利技术向国际市场发展的实力已较以往有了较大程度的提高,专利家族的国家覆盖率不断增加,基本包含了世界上主要的经济体。各个机构加强了在新兴经济体的专利保护力度,在新兴经济体的专利公开数量持续增长。六个机构在中国、巴西、印度、墨西哥的专利家族公开数量都有明显提高。而在台湾、韩国、南非等国专利家族公开数量呈现出增长的态势。这从一个方面反映出技术创新和经济全球化的趋势。

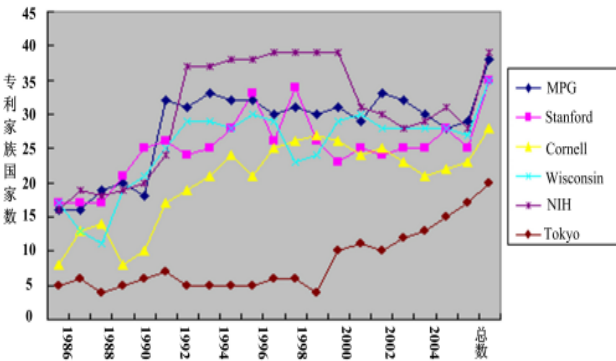


图3 六家机构 1986-2006 年专利家族国家数量变化图

2.3 公开专利合作机构数量比较分析

根据 1986-2005 年六家机构合作的公开专利总

数,表 1 中列出了 1986-2005 年间进入各机构合作对象(除个人)前十位的机构名称。从表 1 可以看出,在六家机构中 NIH 在合作方面表现的最为积极,总数达到了 877 件,较最后一位的 Wisconsin 多出了 813 件。

根据表 2 中的分析结果来看,与 Tokyo 合作的前十家机构中,企业占据六成,包括了日产汽车、丰田、富士、日立和三菱重工五家世界知名企业。欧美机构的合作机构则是政府部门或者其他科研机构占多数。NIH 前十机构中有六家都是美国政府部门,以及两家大学。这与其本身的政府背景也有关。但是也不乏像 Stanford 和 MPG 这样与企业合作密切的机构。Stanford 与六家企业保持联系,包括了 IBM、GE、HONDA 等;MPG 与五家公司共同合作,有 SONY、BASF 和 BAYER 等。上述六家公司也同样属于世界知名企业。

六家机构的共性点是:他们更倾向于选择与世界知名企业合作。这些企业本身具备雄厚的势力,自身有着强烈的技术需求。尽管在某些方面企业自身的研发体制能够满足一部分需求,但是还需要向科研机构寻求帮助。这些机构在各自领域内拥有着领先的科研实力,而企业则能为科研机构提供资金支持,与企业的合作能够相互补充。

表 1 六家机构 1986-2005 年前十位合作机构(除个人)分布及合作数量

机构	1	2	3	4	5	6	7	8	9	10	总数
NIH	合作机构 US SEC OF COMMERCE	US GOVERNMENT	US SEC OF NAVY	NAT INST HEALTH OFFICE TECHNOLOGY TRANSF	UNIV IOWA RES FOUND	COLEY PHARM GROUP INC	US SEC OF ARMY	UNIV JOHNS HOPKINS	US SEC OF AGRIC	GENELABS TECHNOLOGIES INC	877
	数量	347	169	160	56	36	28	27	20	17	877
Cornell	合作机构 SLOAN KETTERING INST CANCER RES	LUDWIG INST CANCER RES	NESTEC SA	US SEC OF AGRIC	UNIV DUKE	SEMINIS VEGETABLE SEEDS INC	UNIV TEXAS SYSTEM	UNIV ROCKEFELLER	US DEPT HEALTH & HUMAN SERVICES	UNIV LOUISIANA STATE & AGRIC & MECH COLL	97
	数量	22	17	10	10	8	7	7	6	5	97
Stanford	合作机构 HONDA GIKEN KOGYO KK	UNIV CALIFORNIA	HARVARD COLLEGE	INT BUSINES MACHINES CORP	MASSACHUSETTS INST TECHNOLOGY	GENERAL ELECTRIC CO	UNIV WASHINGTON	SASAHARA	SRI INT	APPLIED GENOMICS INC	119
	数量	20	18	18	12	10	10	9	8	7	119
Wisconsin	合作机构 VICAL INC	US SEC OF AGRIC	UNIV CALIFORNIA	UNIV NEW YORK STATE	PENN STATE RES FOUND	UNIV MINNESOTA	PIONEER HI-BRED INT INC	YEDA RES & DEV CO LTD	UNIV ILLINOIS FOUND	US DEPT HEALTH & HUMAN SERVICES	64
	数量	13	12	8	5	5	5	4	4	4	64
MPG	合作机构 BASF AG	BAYER AG	DEUT KREBSFORSCHUNGSZENTRUM	HOECHST AG	FRAUNHOFER GES FOERDERUNG ANGEWANDTE EV	PLANTTEC BIOTECHNOLOGIE GMBH	SONY INT EURO GMBH	SUGEN INC	UNIV GEORG AUGUST GOETTINGEN	YISSUM RES & DEV CO	119
	数量	29	14	12	12	10	10	9	8	8	119
Tokyo	合作机构 ONCOTH ERAPY SCI INC	DOKURITSU GYOSEI HOJIN KAGAKU GIJUTSU SH	NISSAN MOTOR CO LTD	UNIV TOKYO AGRIC & TECHNOL OGY NAT UNIV	TOYOTA JIDOSHA KK	FUJITSU LTD	HITACHI LTD	MATSUSHITA DENKI SANGYO KK	JAPAN SCI & TECHNOL OGY AGENCY	CELL CROSS CORP	294
	数量	69	60	33	33	16	15	14	13	12	294

2.4 机构公开发明专利技术领域分布

技术领域在不同时期专利数量的变化,反映了技术创新随时间的变化情况,显示了技术受重视程度。同一机构在不同技术领域的专利数量变化情况,显示了该机构的核心知识产权的技术形式。不同机构的技术领域专利数量情况,说明了机构技术发展和研发行为状况。

2.4.1 专利技术领域及其公开专利数量的比较分析

按照国际专利分类号(International Patent Classification Codes),依 20 年公开专利总量排名,计算得到各机构在德温特专利数据库中位于前 20 位专利技术主题及其公开专利数量。表 2 为六家机构发明专利公开专利主要涉及的专利技术领域情况(忽略在专利总量中小于 3%的技术领域)。

表 2 六家机构 1986-2005 年 IPC 技术领域分布

	化学冶金	人类生活需要	物理	电学
MPG	54%	24%	12%	5%
Stanford	69%	18%	12%	<3%
Cornell	44%	24%	24%	8%
Wisconsin	53%	30%	11%	3%
NIH	52%	32%	13%	<3%
Tokyo	38%	26%	22%	10%

由此可见,化学、冶金领域已成为六家专利权属机构专利的主要技术领域,均超过了 30%,而 MPG、NIH 和 Wisconsin 更是超过了 50%。人类生活需要和物理两个领域也在各个机构中占据很大的份额。

相对而言,只有 Tokyo、Cornell 和 MPG 在电学领域更有优势,其在电学领域的专利数量占到了 5%以上。而其他三家机构在电学领域只有 3%或更低的份额。由此也可以看出,Tokyo、Cornell 和 MPG 在其整个机构技术部署中对电学领域较其他三家机构更为关注,投入了更多的力量。

2.4.2 专利技术创新领域及其公开专利数量的比较分析

德温特手工代码(Manual Codes)表明了该专利的技术创新领域,它揭示了专利技术的外部特征和应用领域^[5]。图 4 列出了按专利量排序在前 20 位的各机构专利技术创新领域及其数量情况。六家机构位于前 20 名的专利技术创新领域具有以下特征:

六家机构中排名第一位的创新领域完全相同,都为 D05-H09 试验与探测(外部细菌、真菌、病毒)。

六家机构共有的创新领域还有 B04-E08 载体、质粒、黏粒、转位子、抗癌;B04-G01 用抗原定义的抗体;B12-K04F 涉及 DNA 的试验等;B14-H01 抗

癌;D05-H08 细胞或组织培养;D05-H12A 野生型编码序列;D05-H12E 载体。

五家机构共有的创新领域有:B04-E01 核酸;B04-F0100E 细胞、微生物、转化株、宿主、细胞系、组织(基因工程);B04-K04A 诊断与测试;B12-K04E 测试物质以外的其他疾病。

四家机构共有的创新领域有:B12-K04 诊断与试验;D05-H14 重组细胞;S03-E14H4 免疫。

六家机构前 20 位专利技术创新领域中共有的达到了 8 个,表明这 8 个领域是当前各个机构研究的重点,也可以认为是如今热门的研究方向。

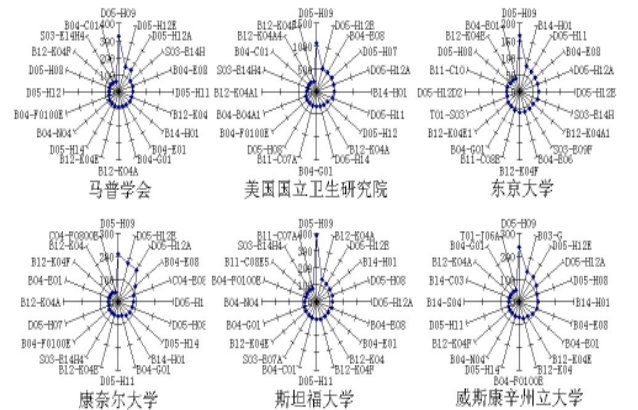


图 4 六家机构技术创新领域分布

2.5 机构专利技术相似度分析

机构专利技术相似度分析可以了解两个机构研究的专利技术领域分布重叠情况。两家机构间相似度越高则说明两家机构研究的专利技术领域分布重叠度越高,共有的专利技术领域越多;反之则两家机构之间各自拥有更多对方没有的研究领域,可了解各自的研究兴趣倾向。技术相似度指标可以作为专利技术领域分布指标的一个补充,二者结合可以更好地了解机构间的专利技术领域分布情况和强弱。

在进行相似度分析时,主要采用了以德温特手工代码来划分各个技术领域分布情况作为分析指标。相似度的取值分布区间为[0,1],0 代表无相似性,1 代表完全相似。

表 3 六家机构技术相似度矩阵

	MPG	Stanford	Cornell	Wisconsin	NIH	Tokyo
MPG		0.896	0.896	0.800	0.887	0.838
Stanford	0.896		0.825	0.855	0.886	0.817
Cornell	0.896	0.825		0.848	0.838	0.771
Wisconsin	0.86	0.855	0.848		0.813	0.761
NIH	0.887	0.886	0.838	0.813		0.776
Tokyo	0.838	0.817	0.771	0.761	0.776	

(下转第 1843 页)

tionoverloaded)引发的困惑和茫然,甚至产生了信息饥渴(information anxiety)、信息焦虑(informationphobia)等症候。调查发现大多数在校本科生(68.24%)面对海量而又杂乱无序的网络信息还不能进行有效的管理分类,这就给高校图书馆提出了更高的要求:加强对网络信息的组织,开展网络信息导航、学科信息门户、数字图书馆建设,为学生提供大量有序、优质、能使学生感到可以信任的信息。

另外我国许多高校图书馆都开展了数字信息咨询服 务,设立了“参考咨询馆员”、“学科馆员”等职位来专门负责解答读者的咨询问题,网络环境下,要更加深化图书馆的信息咨询服务,充分发挥网络快捷性的特点,通过各种途径包括:电子邮件、网络实时咨询、信息推送、个性化订制等为大学生提供网络信息服务,进一步提升图书馆的服务能力。

4.4 加强大学生的网络信息道德教育

高校可以通过各种途径加强大学生的网络信息道德教育,如:在“大学生思想道德修养”课中加入信息道德教育的内容,包括信息安全、信息传播法规、

信息共享、知识产权、信息的合理利用等等。信息道德教育可以促使大学生在信息活动中遵循一定的信息伦理与道德标准,规范自身的信息行为;使他们树立遵守信息法规的意识,不断提高自身的信息道德修养,进而能够增强鉴别与抵制不良信息的能力,有效地抵制信息污染。

参考文献

- 1 Behrens S J. A conceptual analysis and historical overview of information literacy [J]. College & Research Libraries, 1994, 55(4): 309-322.
- 2 陈兰杰,等. 网络环境下高校教师信息素质调查分析[J]. 情报科学, 2007, (3): 404-410.
- 3 符绍宏,等. 国外信息素质教育研究动向[J]. 大学图书馆学报, 2006, (3): 68-72.
- 4 吴 钢,等. 大学生信息素养现状调查与对策分析[J]. 图书馆理论与实践, 2006, (3): 39-40.
- 5 符绍宏. 论当代信息环境下中国的信息素质教育[J]. 情报学报, 2003, (4): 238-249.
- 6 王 琼. 网络环境下高校教师信息素质调查与分析[J]. 中国图书馆学报, 2002, (5): 70-74.

(责任编辑:徐 波)

(上接第 1836 页)

如表 2 所示六家机构当中, 两两机构之间技术相似度最高的是 MPG/Stanford 和 MPG/Cornell 两组, 达到了 0.896; 其次为 MPG/NIH 的 0.887; 最低的为 Tokyo/Wisconsin 的 0.761。六家机构之间的平均技术相似度为 0.8378。从整个矩阵的值来看 MPG、Stanford、Cornell、Wisconsin 和 NIH 五家机构之间的技术相似度均超过了 0.813 以上; 而 Tokyo 和 Cornell、Wisconsin、NIH 之间的技术相似度平均不足 0.770。从中可以看出 Tokyo 与其他五家机构的技术重叠程度较五家机构之间更低。Tokyo 关注的技术领域与 Cornell、Wisconsin、NIH 的差别较大, 研究重点选择不同。这里也印证了之前专利技术领域分布分析结果当中 Tokyo 在电学领域和其他机构之间的差别, 同时还有更多细分方面的差别因为篇幅就不再赘述。

3 结 语

在以机构作为对象的分析中, 使用 TDA 可以从多维度对该机构的研究方向和技术领域进行深刻揭示。特别是对世界方位内具有代表性的机构分析, 可以把握当前的技术热点发展方向

当分析时间跨度比较长时, 需要对机构的历史沿革有一个清晰的了解, 才能提高检索得到的专利信息查全率和查准率。此外在数据清洗阶段还需自动和人工干预的方式相结合, 这样才能减少误差。

从分析过程可以看出 TDA 是一款强大的专利数据挖掘工具, 是情报分析工作强有力的助手, 可以大大的提高情报人员的工作效率, 得到更多更深层次的信息。使用 TDA 对专利数据从多个方面和层面进行解读, 可以为技术趋势预测、机构发展态势、研发重点等诸多领域提供有力的参考。

参考文献

- 1 USPTO Releases List of Top 10 Universities Receiving Most Patents in 2005[EB/OL]. <http://www.uspto.gov/web/offices/c-om/speeches/06-24.htm>, 2007-11-30.
- 2 Derwent Innovations Index[DB/OL]. <http://isiknowledge.com>, 2008-01-11.
- 3 王永红. 定量专利分析的样本选取与数据清洗[J]. 情报理论与实践, 2007, (1): 93-96.
- 4 张秋菊. 德温特分析工具在机构比较研究中的应用[J]. 图书情报工作, 2006, (5): 26-28.
- 5 Manual Code Lookup[EB/OL]. <http://scientific.thomsonreuters.com/mcl/>, 2008-01-15.

(责任编辑:徐 波)