

doi:10.3772/j.issn.2095-915x.2016.02.012

数据资源集成技术专利分析

——以八国两组织专利数据为例

朱新超¹, 潘薇², 刘会景²

(1 中国科学技术信息研究所 北京 100038; 2 北京万方数据股份有限公司 北京 100038)

摘要: 随着互联网、移动互联网、云计算技术的发展,数据资源的重要性越来越凸显。本文以八国两组织范围内的相关技术专利数据为依据,通过对数据资源集成技术专利信息的数据挖掘和数据分析,从专利申请趋势、重点专利申请人的专利申请趋势及专利技术生命周期等指标,分析数据资源集成技术在八国两组织范围内的发展现状及未来的发展趋势;结合专利申请地域分布、主要专利申请人专利申请区域分布等要素分析八国两组织的专利竞争情况及重点专利区域布局。

关键词: 专利分析, 数据资源, 数据集成

Patent Analysis of Data Resources Integration Technology

——Via Eight Countries and Two Organizations' Patent Data

ZHU Xinchao, PAN Wei, LIU Huijing

(1. Institute of Scientific and Technical Information of China, Beijing 100038, China;

2. Wanfang Data Co., Ltd, Beijing 100038, China)

Abstract: With the development of internet, mobile internet and cloud computing technology, the data resources are become more important. This study used the data from eight countries and two organizations within the scope of related technology patent, by employing the data resource integration technology of data mining and

基金项目: 本文为科技部国家科技支撑计划“面向科技创新的专利信息加工与服务关键技术研究与应用示范”(项目编号:2013BAH21B01)的研究成果。

作者简介: 朱新超(1990-),男,专利分析师。研究方向:专利分析、信息资源服务等;潘薇(1980-),女,湖南长沙人,博士,从事人工智能、信息技术研究;刘会景(1984-),女,硕士,专利分析师。研究方向:专利分析、专利数据挖掘、信息资源服务研究等。

analysis. From the aspects of patent application trend, the application trend of key applicants and the life cycle of patent technology, we analyzed the current status and future development trend of data resource integration technology of the eight countries and two organizations. The regional distribution of patent application and the main applicants are also considered for discussing the patent competition and key regional layout of the two organizations and eight countries.

Key words: Patent analysis, data resource, data integration

1 前言

随着互联网、移动互联网、云计算技术的发展,信息化浪潮已然兴起,数据资源的重要性越来越明显。在我国,数据资源建设与应用技术近年来有了长足的发展,但是也存在一定不足,制约了数据资源的进一步发展,数据资源集成技术是突破现状的唯一途径。集成是一种创造性的融和过程,只有构成一个系统要素,经过主动的优化、选择搭配、相互之间以最合理的结构形式结合在一起,形成一个有适宜要素组成的、优势互补的有机体^[1]。数据资源集成不是数据资源的堆积或数据载体的物理堆积。数据资源集成其本质是一种竞争性的互补关系,即各种要素通过竞争冲突,不断寻找、选择自身的最优功能点,在此基础上进行相互匹配^[2]。数据资源集成的目的,是对若干数据资源来源数据,利用计算机技术、数据通信技术、网络技术、云计算技术及数据库存储技术进行统一管理,实现数据资源的资源共享,同时保证数据资源集中的数据资源的冗余最小。

专利是评测某机构技术和创新能力的重要指标,有助于未来的研究与实际开发的研究成果类文献信息^[3-4]。本文的数据资源集成技术专利分析基础数据,来源EPO(欧洲专利局)专利数据中的八国两组范围内。通过对数据资源集成技术八国两组范围内的专利信息的数据挖掘和数据分析,从专利申请趋势、重点专利申请人专利申请趋势及专利技术生命周期等指标,分析数据资

源集成技术在八国两组组织范围内的发展现状及未来的发展趋势;结合专利申请地域分布、主要专利申请人专利申请区域分布等要素分析八国两组织的专利竞争情况及重点专利区域布局。为企业、高等院校或科研机构的数据资源集成的技术研发提供有益的参考和启示。

2 国内外数据资源集成技术发展现状

数据集成存在异构性、分布性和自治性等特点,数据集成技术一直受到人们的关注。数据集成是对各种异构数据提供统一的表示、存储和管理。屏蔽了各种异构数据间的差异,通过异构数据集成系统进行统一操作。因此集成后的异构数据对用户来说是统一的和无差异的,数据集成技术的研究涉及到多种计算机技术,如分布式对象技术、XML技术、面向对象技术、数据库技术等。目前,国内外已经研制出来的数据集成系统也有很多,本文列举其中的部分数据集成系统来简要介绍当前数据集成系统的发展现状。

LORE (A Lightweight Object REpository for Semistructured Data) 是美国斯坦福大学设计的一个专门用于管理半结构化信息的数据库管理信息系统。在知识模型中使用的数据模型是一个轻量级的数据称之为OEM(Object Exchange Model) OEM是面向对象数据库之间的半结构化数据交换模型。LORE可以对OEM和XML数据进行存储和管理。LORE具有标准数据库的一些特征,支

持多用户、日志和恢复功能以及查询和更新语言，同时 LORE 也提供了其他一些操作数据库的工具。但是，目前，LORE 在查询过程中并没有提供事务管理、并发控制或恢复等功能^[5]。

Ozone 是美国斯坦福大学设计的一个集成模型，该模型对面向对象数据库的 ODMG 模型进行了扩展，可以对结构化和半结构化数据进行集成，Ozone 系统特别适合处理杂合性数据尤其是广泛应用于 Web 上的数据^[6]。

Versatile 是由东南大学开发的一个基于 CORBA 的可扩展的异构数据源集成系统原型。该模型在 IONA 公司的 Orbix 产品上对 SQL Server Versant 文件系统超文本数据进行包装和集成。该系统不仅能集成上述数据源而且能集成随时插入的新数据源中的数据，在上述的三种模型中 LORE 储 XML 数据的一个专门数据库系统，它是异构数据集成系统代表性的例子。该系统的体系结构设计查询处理和优化方式为后来的异构数据集成系统提供了很好的借鉴，通过对 LORE 系统的分析可深入了解异构数据集成系统^[7]。

3 专利检索策略和分析工具

本文使用的数据资源集成技术专利分析数据主要来源于 EPO 专利数据，专利检索的国家/组织

范围包括中国 (CN)、德国 (DE)、日本 (JP)、美国 (US)、英国 (GB)、法国 (FR)、瑞士 (CH)、俄罗斯 (RU)、欧洲专利局 (EP) 及世界知识产权组织 (WO)。数据覆盖范围为 1967 年至 2014 年。经检索得知，瑞士在数据资源集成技术领域无相关专利，其他国家/组织共检索到 4150 条相关专利。

本文数据资源集成技术检索策略主要采用的是主题词检索的方式，其主题词主要为“Data integration”，同时使用国际专利分类号 (IPC) 予以限定，其 IPC 分类号主要为“G06F 17/30”、“G06F 7/00”。对初步检索的专利信息进行人工筛选，去除不相关的噪音数据；其次，将筛选结果进一步清洗、规范，最终得到用于本次分析的专利基础数据。其检索策略如图 1 所示。

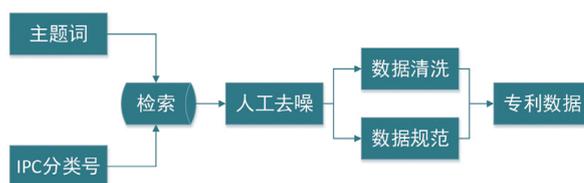


图 1 检索流程图

4 数据资源集成技术专利分析

4.1 专利申请趋势分析

图 2 为数据资源集成技术八国两组织的专利

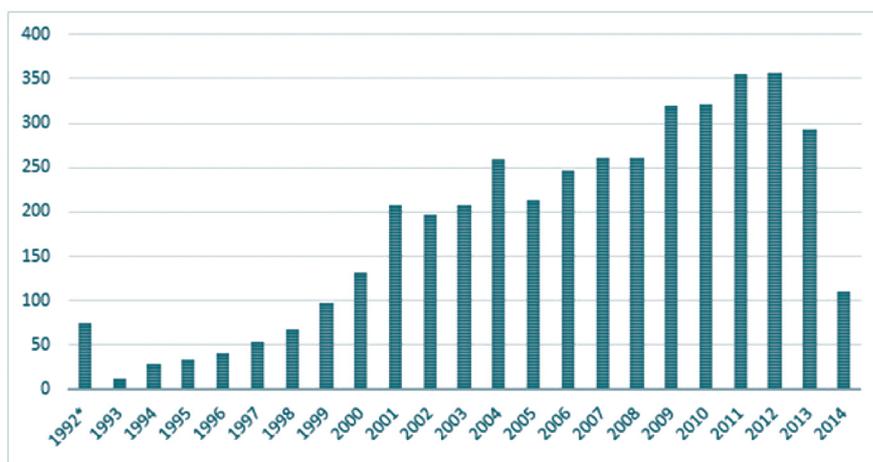


图 2 数据资源集成技术发展趋势图 (1992* : 1967 - 1992)

数量年度分布趋势图。第一件数据资源集成技术的专利申请于1967年，截止到1993年，数据资源集成技术一直处于缓慢发展的阶段，专利申请量变化不大，该阶段数据资源集成技术处于起步阶段。自1994年以后，数据资源集成技术领域专利的申请量呈逐年递增的态势，2012年达到专利申请量的顶峰，专利申请量高达356件。这与近来数据积累量及对多源异构数据资源集成的需求相一致。专利申请量在2013年以后出现下滑的趋势，这是由于经过多年的发展，数据资源集成技术不断趋于成熟，技术突破难度加大。

4.2 专利申请区域分布

从八国两组织的数据资源集成技术各国申请的专利情况来看，如图3所示，美国、中国、日本的专利申请量最大，占据专利总量的84%以上，

其中美国的专利申请量最多，可见美国、中国、日本在数据资源集成技术领域处于数据资源集成技术领先地位。德国、法国、英国、俄罗斯的数据资源集成专利相对较少，其数据资源集成技术的起步可能较晚，且数据资源集成技术可能处于技术萌芽阶段，技术发展缓慢。

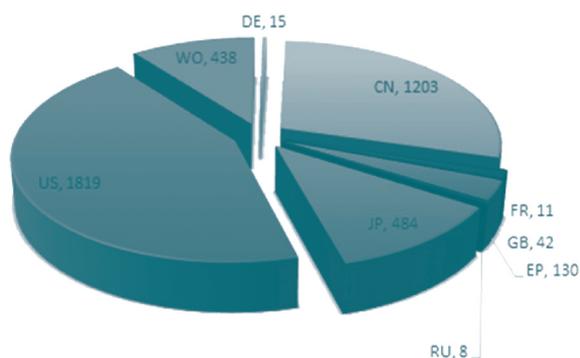


图3 八国两组织数据资源集成技术专利地域分布

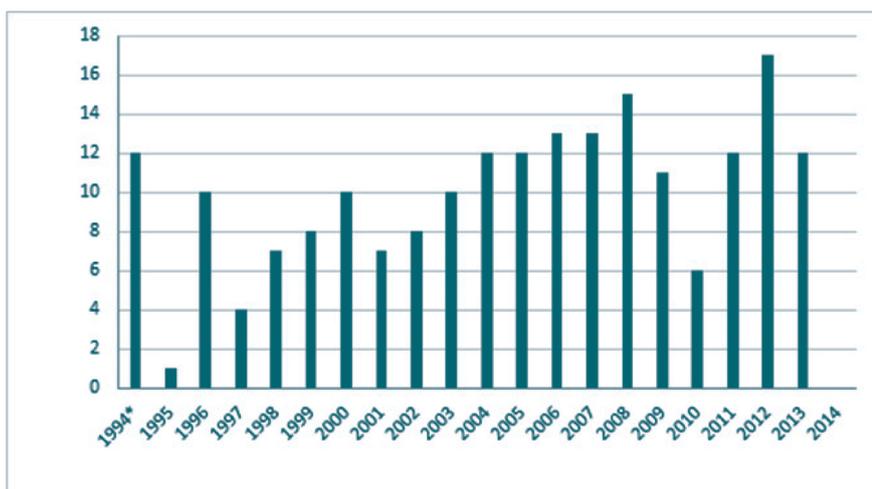


图4 IBM专利逐年申请趋势图（1994*：1976-1994）

4.3 主要专利申请人分析

4.3.1 申请量排名第一的专利申请人：IBM

4.3.1.1 专利逐年申请趋势图（图4）

IBM（International Business Machines Corporation），中文全称国际商业机器公司。公司总部位于美国，主要产品包括软件、服务器及

其他产品，是全球最大的信息技术和业务解决方案公司^[8]。IBM早在1976年就提出了第一件关于数据资源集成的专利申请，研发起点早。随着时间的推移，IBM的专利申请量总体来说呈现逐步增长的态势，2012年专利申请量已达到17件，说明IBM在不断加大其在数据资源集成领域的科研投入，并且取得了较大成果。

4.3.1.2 专利申请区域分布

IBM 公司总部位于美国，其专利申请主要针对美国市场。鉴于数据资源集成的全球化及技术保护需求，IBM 也在其他国家进行了专利布局。由图 5 可知，IBM 在 CN、JP、EP、GB 和 DE 均申请了相关专利，其中在中国申请的专利数量最多，说明 IBM 较为重视中国市场。

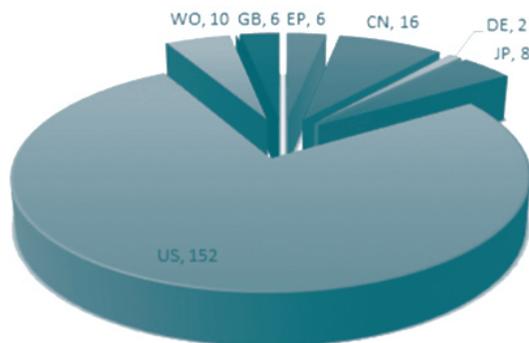


图 5 IBM 八国两组织专利申请量地域分布图

4.3.2 申请量排名第二的专利申请人：MICROSOFT CORP

4.3.2.1 专利逐年申请趋势图

MICROSOFT CORP (微软) 总部位于美国，是一家跨国电脑科技公司，也是世界 PC 机软件开发的先导^[9]。如图 6 所示，MICROSOFT CORP 在数据资源集成领域的专利申请始于 1996 年。1998 年 MICROSOFT CORP 提出 ADO.NET 数据提供程序，专门用来同特定类型的数据源进行通信。不同于 IBM 的申请趋势，

MICROSOFT CORP 申请的专利主要集中在 2004 年、2005 年和 2011 年，说明在 2004 年、2005 年和 2011 年，该公司掌握了某些关键技术，并申请了大量的专利予以保护。MICROSOFT CORP 在 2013 年申请的专利达 10 件，表明公司一直重视数据资源集成技术的研发，注重技术的延续性，这与市场需求一致。

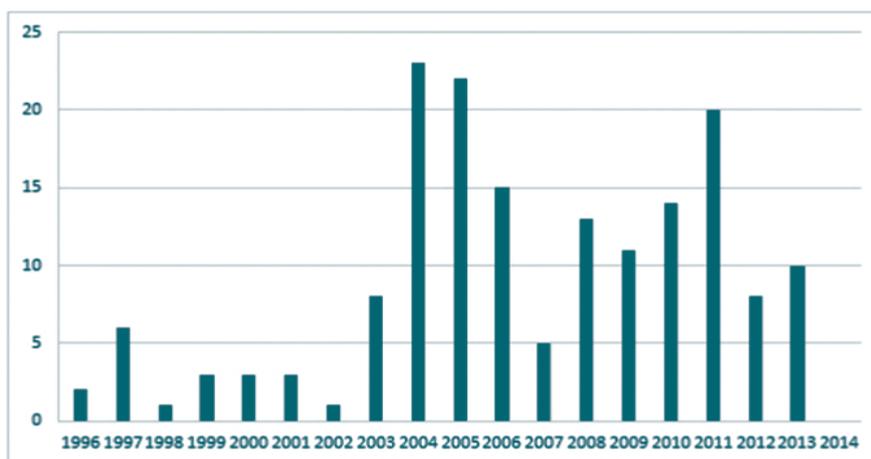


图 6 MICROSOFT CORP 专利逐年申请趋势图

4.3.2.2 专利申请区域分布

MICROSOFT CORP 总部位于美国，其专利申请主要针对美国市场。面对数据资源集成的全球化需求，MICROSOFT CORP 在 CN、EP、JP 及 RU 国家进行了专利布局。MICROSOFT CORP 在不同国家的专利申请量，体现出其对不同国家市场的重视程度，由图 7 可知，

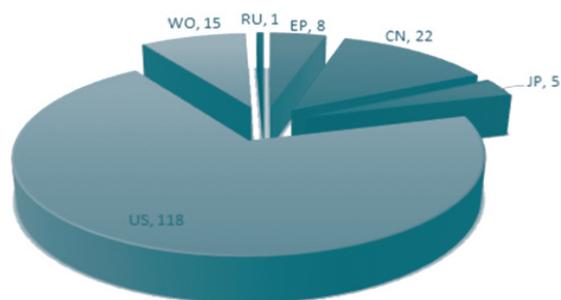


图 7 MICROSOFT CORP 八国两组织专利申请地域分布图

MICROSOFT CORP 较为重视中国市场，累积在中国提出 22 件专利申请。

4.3.3 申请量排名第三的专利申请人: FUJITSU LTD

4.3.3.1 专利逐年申请趋势图

FUJITSU LTD (富士通) 总部位于东京，是一家起源于日本的全球性信息通信技术 (ICT) 企业，提供着全方位的技术产品、解决方案和

服务^[10]。如今，富士通是日本排名第一的 IT 厂商，全球第四大 IT 服务公司。如图 8 所示，FUJITSU LTD 的专利申请趋势与 MICROSOFT CORP 相类似，2004 年和 2011 年的专利申请量明显高于其他年份，推测与该年的技术需求及技术突破有关。

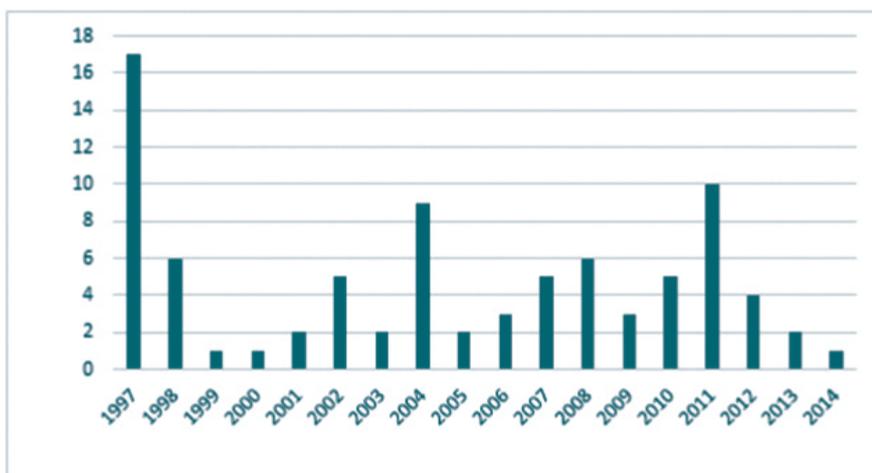


图 8 FUJITSU LTD 专利逐年申请趋势图 (1997* : 1980 - 1997)

4.3.3.2 专利申请区域分布

FUJITSU LTD 总部位于东京，其专利申请主要针对日本市场。由图 9 可以看出，FUJITSU LTD 在美国申请 23 件专利，将近本土专利申请量的一半，表明 FUJITSU LT 非常看重美国市场。另外，FUJITSU LTD 还在 CN、EP 及 GB 国家申请了专利，体现出其专利全球化布局的特点。

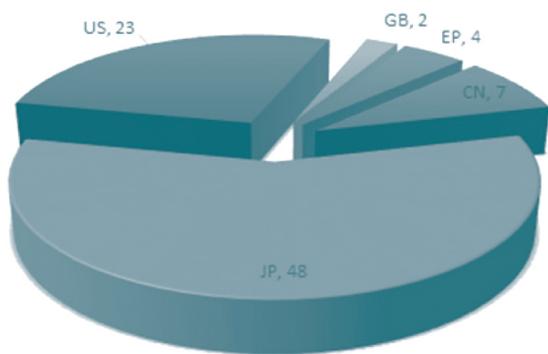


图 9 FUJITSU LTD 八国两组织专利申请量地域分布图

4.4 技术生命周期分析

4.4.1 八国两组织生命周期分析

数据资源集成技术的第一件专利申请于 1967 年的英国。随着数据资源集成技术的出现，数据资源集成技术开始进入技术发展缓慢的技术萌芽期，专利申请量及专利申请人数量变化不大；自 1994 年以后，数据资源集成技术的专利申请量和专利申请人数量逐年增加，呈现逐年递增的发展态势，直至 2012 年，专利申请量和专利申请人数量达到顶峰，该阶段的数据资源技术处于高速发展的技术发展期；2012 年之后专利申请量与专利申请人数量开始出现下滑现象，这是由于经过多年的发展，数据资源技术进入技术成熟期，技术突破难度大，近年来，更是初步呈现技术衰退的迹象，如图 10 所示。

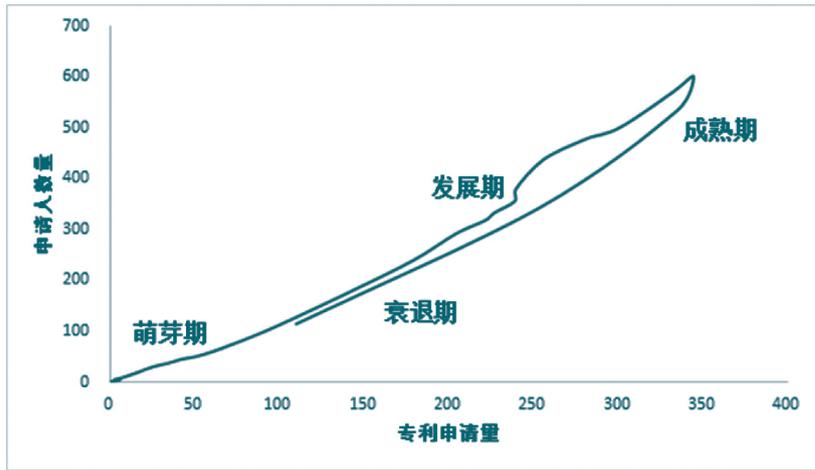


图 10 八国两组织专利技术生命周期图

4.4.2 各国 / 组织技术发展现状分析

八国两组织数据资源集成技术虽最早于英国申请，但是，数据资源集成技术在英国国内未得到重视，以至于英国的数据资源技术尚处于低速发展的技术萌芽期，如图 11 所示。同时，法国、德国、俄罗斯、欧洲专利局也处于技术发展的初期，其专利申请量相对较少，且专利申请人相对集中，技术发展缓慢。日本、中国的专利申请量与专利申请人数量基本成正比，可见日本、中国的数据资源集成技术尚处于技术高速发展

阶段，并开始趋于成熟。世界知识产权组织的专利申请数量大大高于专利申请量，由此可见，世界知识产权组织的数据资源集成技术已经从技术成熟期开始进入技术衰退期，数据资源集成技术突破难度加大。美国的专利申请量和专利申请人数量都远超其他国家 / 组织，可见美国的数据资源集成技术处于技术领先地位，同时，美国的专利申请人数量也高于专利申请量，可见美国的数据资源集成技术开始趋于成熟，并初步呈现衰退的迹象。

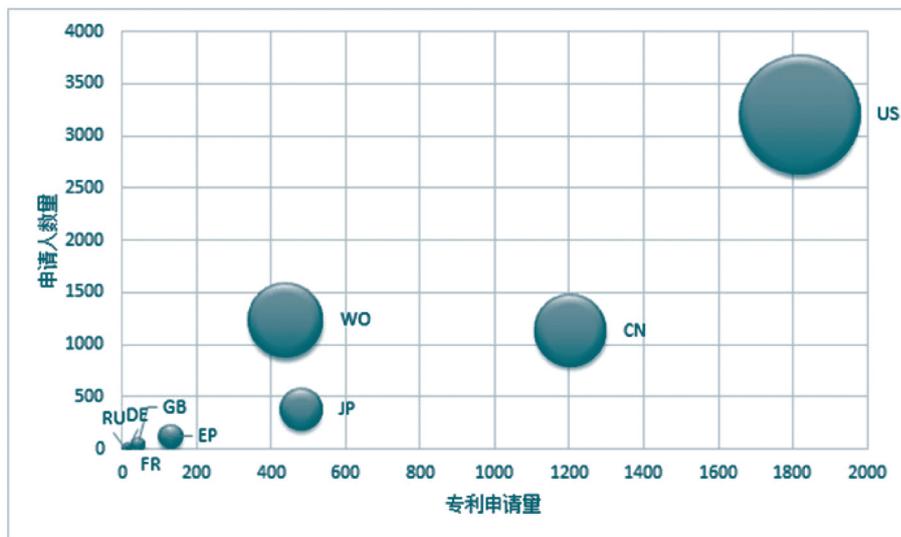


图 11 八国两组织技术发展分析

5 总结

数据资源的集成、整合,可以最大限度的实现资源共享,尽可能的减少中间环节,缩短信息传递与反馈的速度,实现数据资源的可持续发展^[11-13];减少数据处理过程中的费用,减低数据服务成本,充分发挥数据资源的整体效益。

上文通过对数据资源集成技术领域的八国两组织范围内的专利进行信息检索和数据统计,结合专利申请趋势和重点专利申请人专利申请趋势、专利申请区域分布等方面,分析八国两组织范围内的数据资源集成技术整体发展态势及其区域分布。从专利技术生命周期角度出发,对比了

各国/组织数据资源集成技术的发展态势及现状。同时,从重点专利申请人申请趋势等角度出发,对比了八国两组织范围内主要数据资源集成技术企业的研发实力。

通过对数据资源集成技术领域的分析和研究表明,数据资源集成技术产业自诞生以来经过长时间的缓慢发展,目前,数据资源集成技术已经进入缓慢发展的技术成熟期。同时,该技术领域在八国两组织范围内存在一定的技术发展不平衡,其中德国、英国、法国、俄罗斯数据资源集成技术尚处于技术萌芽期,其他地区或组织的数据资源集成技术开始进入技术衰退期。

参考文献

- [1] 百度百科.集成[EB/OL].http://baike.baidu.com/link?url=167rKK1BJ75QMtQQ_RQ1dsGF8ESSrOsdbFGQygGfZxH-4HcrMevZcZ29oeNMYQtNsJnLePwgKYkDARLr14xUMq, 2016-03-30.
- [2] 马大川,杨红平.信息资源的集成整合研究[J].中国图书馆学报,2004,30(3):36-40.
- [3] 朱新超,凌峰,潘薇,等.我国近场通信技术现状专利分析[J].中国科技信息,2015(19):124-126.
- [4] 朱新超,霍翠婷,刘会景.不同来源的中国专利信息资源整合研究[J].情报理论与实践,2014,37(11):90-93.
- [5] Quass D, Widom J, Goldman R, et al. LORE: A Lightweight Object REpository for Semistructured Data.[J]. Acm Sigmod Record, 1996, 25(2):549--4.
- [6] 勒强勇,李冠宇.异构数据集成技术的发展和现状[J].计算机工程与应用,2002,38(11):112-114.
- [7] 方茂华.企业资源整合与查询优化[D].华中科技大学,2004.
- [8] 王勇.IBM的人力资源管理[J].中国人才,2004(4):76-77.
- [9] 微软公司(中国).微软-中国[EB/OL].<http://www.microsoft.com/zh-cn>, 2016-03-30.
- [10] 富士通公司(中国).富士通-中国[EB/OL].<http://www.fujitsu.com/cn/>, 2016-03-30.
- [11] 马费成,赵红斌,万燕玲,等.基于关联数据的网络信息资源集成[J].情报杂志,2011,30(2):167-170.
- [12] 李军怀,张景,吕林涛,等.异构环境下信息资源集成与访问模型研究[J].大连理工大学学报,2003,43(S1):96-99.
- [13] 刘会景,凌峰,朱新超,等.多源专利信息资源整合研究--以美国专利数据为例[J].数字图书馆论坛,2015(9):22-26.