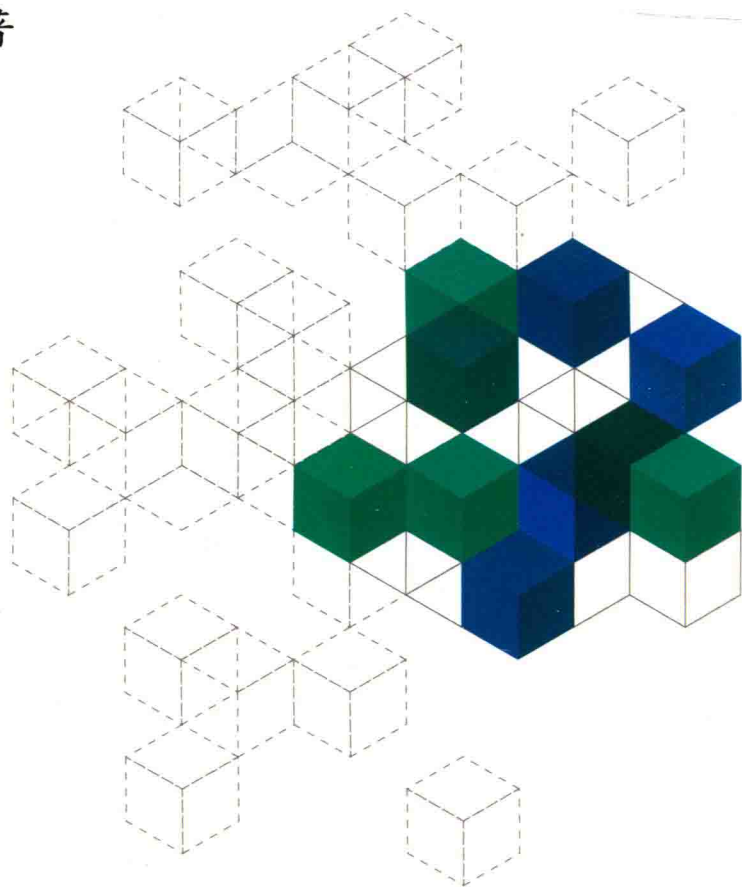


Methods, Technology, and Application
of Resource Monitoring of International
Scientific Cooperation based on Data Mining

基于数据挖掘的 国际科技合作资源监测 方法、技术及应用

刘 云 等◎编著



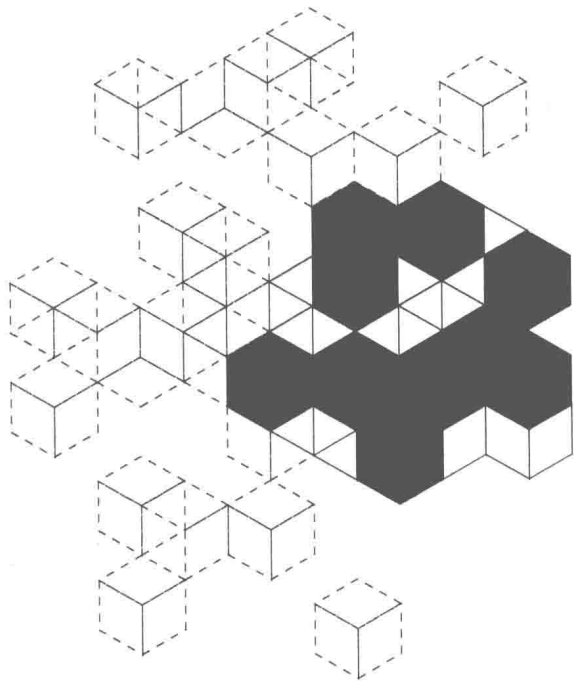
科学出版社

本书的研究工作得到国家自然科学基金重点项目
(编号: 71033001) 的资助

Methods, Technology, and Application
of Resource Monitoring of International
Scientific Cooperation based on Data Mining

基于数据挖掘的 国际科技合作资源监测 方法、技术及应用

刘 云 等◎编著



科学出版社

北 京

内 容 简 介

本书首先基于结构化国际科技资源信息数据库及非结构化国际科技资源信息,研究国际科技资源信息的数据抽取技术、学科与技术分类及数据库构建方法;其次,基于文献计量、专利计量、文本挖掘、研究价值图谱、技术路线图方法,研究构建国际科技资源监测分析方法和模型;再次,基于聚类分析、社会网络分析、地理信息系统等方法,研究国际科技资源信息监测分析的可视化方法和模型;最后,介绍了国际科技合作资源监测系统和国际科技资源监测地理信息系统的开发及应用。

本书可供相关专业师生阅读,也可供科技管理人员使用。

图书在版编目(CIP)数据

基于数据挖掘的国际科技合作资源监测方法、技术及应用 / 刘云等编著.
—北京:科学出版社,2014

ISBN 978-7-03-041528-3

I. ①基… II. ①刘… III. ①国际科技合作-资源管理-数据采集-研究
IV. ①F1*

中国版本图书馆 CIP 数据核字 (2014) 第 127688 号



责任编辑: 乔艳茹 / 责任校对: 鲁 素
责任印制: 徐晓晨 / 封面设计: 铭轩堂

编辑部电话: 010-64035853

E-mail: houjunlin@mail.sciencep.com

科学出版社 出版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

北京京华虎彩印刷有限公司印刷

科学出版社发行 各地新华书店经销

*

2015 年 7 月第 一 版 开本: 720×1000 1/16

2015 年 7 月第一次印刷 印张: 16 3/4

字数: 309 000

定价: 88.00 元

(如有印装质量问题, 我社负责调换)

目 录

1	绪论	1
1.1	国际科技合作的重要性	1
1.2	国际科技合作资源挖掘的意义	3
1.3	国内外研究现状	4
2	国际科技合作资源信息来源及特点分析	6
2.1	结构化国际科技资源信息数据库	6
2.2	非结构化国际科技资源信息	21
3	国际科技资源信息的数据抽取与分类技术	42
3.1	主要软件技术	42
3.2	学科分类和技术分类	72
3.3	数据抽取、加工与建立本地数据库	77
4	国际科技资源监测分析方法和模型	84
4.1	文献计量	84
4.2	专利计量	98
4.3	文本挖掘	104
4.4	研究价值图谱	113
4.5	技术路线图	118
5	国际科技资源信息监测分析的可视化方法和模型	131
5.1	数据空间	131
5.2	空间映射	137



5.3	多维数据可视化方法	141
5.4	聚类分析方法	164
5.5	社会网络分析软件工具	174
5.6	关联规则挖掘	183
5.7	基于地理信息系统的可视化技术	185
6	国际科技合作资源监测系统简介	216
6.1	系统简介	216
6.2	系统功能	216
6.3	操作指南	219
7	国际科技资源监测地理信息系统简介	245
7.1	启动登录界面	245
7.2	主要功能	247
7.3	导航功能	249
7.4	查询功能	250
7.5	制图输出功能	255
7.6	系统管理功能	257
	参考文献	259
	后记	264

绪论

1.1 国际科技合作的重要性

科学是人类共同的财富，科学技术是在人类的共同努力、相互交流中发展起来的。目前，科学技术研究正朝着全球化的方向发展，国际化已成为当代科技研究的重要特征之一。科学研究所涉及问题的空间尺度、投资强度和复杂程度都标志着科学研究已进入一个全球化时代。例如，从事科学研究的科学家正更多地在全球化和网络化的开放环境中相互竞争、交流与合作；跨国公司在不同的国家建立研发机构开展合作研发等活动，充分利用全球科技资源，获得更强大的竞争力，从而成为推进科技全球化的主要力量。今天，科学研究的全球化趋势正在不断扩大，全球化将对科学研究的对象、方向、范围、水平，科学家之间的学术交流与合作方式，以及学科交叉研究的发展产生重大而深远的影响。随着科学技术活动范围的日益扩展和科学信息交流工具的不断发展，科学研究的全球化趋势在未来将会越来越扩大，成为科学研究的一种普遍方式。因此，我国的科学研究必须顺应全球化的潮流，抓住全球化带来的机遇和挑战，把国际合作交流放在一个重要的位置。

多年来，国内外的实践证明，积极开展国际科技合作，在技术引进、人才引进的基础上，实施消化吸收再创新，是实现科技进步的一条捷径，是发展中国家和落后地区实现技术跨越的有效手段。经过 30 多年的改革开放，我国国际科技合作进入了新的战略转型期，国际科技合作对促进我国科技的跨越发展、增强我国科技自主创新能力的的作用日益明显，在推进科技、经济、社会发展中的战略地位日益提升。《“十一五”国际科技合作实施纲要》明确指出，未来我



国国际科技合作要实现战略转变，树立全局观念，加强宏观指导和统筹协调，形成资源共享、信息沟通、政策配套、多方参与的“大国际科技合作”的新格局，充分利用全球科技资源，更好地服务于国家战略目标，在体制和机制上保障《国家中长期科学和技术发展规划纲要（2006—2020年）》的贯彻落实。

围绕国家科技发展战略，建立高层次、高水平的对外科技合作平台，以充分利用国际科技资源，增强我国科技自主创新能力，努力为国家经济建设和社会发展服务，为国家科技发展和建设中国特色国家创新体系服务，是我国科研管理的一项重要任务。

国家“国际科技合作专项”是国家层面上首个也是唯一的旨在通过整合、统筹，充分利用全球科技资源，提高自主创新能力的对外国际科技合作与交流平台。“国际科技合作专项”配合国家外交战略，有选择地资助一批政府间科技合作协定中的重点项目，以及一批有效利用全球科技资源，着力解决制约我国社会、经济、科技发展的重大科学问题和关键技术瓶颈，有利于实现“重点跨越”的重大国际科技合作项目。该计划自启动以来，每年针对我国重点领域及项目展开的国际合作进行资助，在逐年增长的国际合作与交流专项经费的基础上，整合 863 计划、973 计划、科技支撑计划等国家各大科技计划中的国际合作资源，形成了国际科技合作凝聚力。其中，2001~2006 年立项的国际科技合作项目共 424 项；而 2007 年和 2008 年分别立项 170 项和 233 项，合作项目分别涉及 27 个、41 个国家和地区组织，主要集中在欧洲、北美洲和亚洲，尤以美国、德国、日本、荷兰、澳大利亚、俄罗斯、英国等为主。中华人民共和国科学技术部（以下简称科技部）专项经费资助超过 4 亿元，集成国际科技合作经费超过 10 亿元，工作重点是政府间国际科技合作项目，以及围绕中医药、新能源、信息通信、生物医药、纳米技术等重点领域的合作项目。

在科学技术全球化的今天，国际科学技术合作已经成为国家科技发展战略中的一个重要组成部分。由于人类对未知世界的探索是无限的，而社会可以用于科学探索的资源却是有限的，所以无论是发达国家要保持 in 科学知识前沿领域的领先地位，还是发展中国家在原本落后的起点上实现局部突破、跨越式发展，都需要通过国际合作的途径，在全球范围内整合有限的资源，以求得本国科学技术最有效的发展。综上，我国的科研管理工作应深入研究，在充分考虑国家安全和国家利益的前提下，在国际科技合作的对象选择、资源配置等方面制定并采取有效措施，使国家的科学技术研究在科技全球化的大趋势下获得更高的效率。

1.2 国际科技合作资源挖掘的意义

随着计算机技术和网络技术的广泛使用,以及各类科技文献信息服务系统的涌现,科研用户要查找所需的基础文献已不再是难事,而对研究领域内的科技热点、前沿、发展趋势、领域内的知识结构及发展变化过程、领域内科研活动主体之间的各种关系等更加专业化和深层次的知识服务有着强烈的需求。当前科学技术日新月异,科研工作也广泛深入开展,新兴学科和交叉学科不断出现,科学研究要解决的问题也日益复杂,跨学科、跨机构、跨区域的科研协同已成为科学发展的必然趋势,未来的科研工作将更多地由分散在不同地域的研究团体通过合作的方式进行,便于交流和协作的科研信息环境是科研工作取得成功的重要保障。

此外,现代科学的发展为社会和科学家们提供了越来越多的科研机会,科学研究领域也越来越广阔。这就向现代的科技决策管理者提出了一个难以解决的问题,即任何一个国家和政府都无法满足科学家们无尽的探索能力和科研兴趣,无法给当代科学活动以全面的支持,而必须有所选择。因此,在科学资源和科技经费受限的情况下,我们开展的国际合作项目必须是高水平的国际合作项目。

在我国国际科技合作的宏观战略制定和计划项目管理中,决策者、管理者、战略专家和评审专家经常遇到无法准确判断合作伙伴的质量和水平的问题,中方寻找合作伙伴也存在一定的随意性,缺乏一个高水平的国际科技合作资源数据中心和监测分析系统为决策者、管理者和执行者提供行动参考,致使我国的国际科技合作难以把握战略上的主动权和针对性,一些低水平、低效率的合作活动削弱了我国国际科技合作应有的作用,不能适应新时期国际科技合作战略转型和国家财政对国际科技合作经费投入大幅提升的发展需求。因此,研究建立高水平的国际科技合作资源监测平台,对提升我国国际科技合作战略管理的科学化水平和实现合作中的效益最大化意义重大。

开展高水平的国际科技合作,首先要识别一流水平的国际科技合作资源。我们将国际科技合作资源界定为一流机构、一流人才和一流技术,那么,这“三个一流”国际资源在哪里?如何挖掘?如何评价?如何表征?这是本书要解决的关键问题,通过研究开发面向结构化和非结构化的国际科学技术信息的数据挖掘与可视化方法和技术,构建国际科技合作资源监测系统,建立可视化的国际科技合作资源多维分布地图,为国家的国际科技合作宏观战略管理和产学研寻找高质量的国际科技合作伙伴提供决策支持平台。



1.3 国内外研究现状

近年来,随着科技文献信息量的猛增、科技信息资源和科学研究朝着数字化方向的发展、计算能力提高和计算速度的迅速增长、存储容量的不断提高,以及一系列新的挖掘、分析、检索、可视化等技术的成熟应用,国内外在科技监测领域开展了一些理论研究和实践探索。

1.3.1 国外研究现状

在科技监测领域,国外已基于科技文献和互联网信息资源,在监测模型、相关算法和系统工具等方面开展了大量的理论研究与应用实践,具体如下。

德国 Transinsight 公司与德累斯顿工业大学 (Technische Universität Dresden, TU Dresden) 的生物信息集团合作研发的 Gopubmed, 基于语义分类工具 Gene Ontology 和 Mesh, 实现语义网络、生物医学信息检索和生物医学热点可视化分析的相互结合, 帮助用户迅速定位某一领域的热点术语、核心作者、核心期刊、年度分布、著者城市分布及国家分布; 美国科学信息研究所 (Institute for Scientific Information, ISI) 2004 年在对 Web of Knowledge 升级过程中新增的分析工具 Result Analysis, 使用户能够对通过检索获得的数据进行深入挖掘, 并将检索结果进行过滤, 揭示出隐含的趋势和模式, 从而达到掌握领域科技研究热点的目的; 美国利哈伊大学 (Lehigh University) 开发了分层分布式动态索引 (hierarchical distributed dynamic indexing, HDDI), 利用语义局部性区域方法进行概念的聚类, 并通过神经网络来识别新兴概念; 美国曼彻斯特大学研发的 TimeMines 系统可以从具有时间阈的自由文本集合中生成时间线, 该系统可以基于统计属性, 对语义特征进行监测、排序及分类, 从而确定科技研究热点; 美国德克赛尔大学研发的 Citespace 采用基于突发词的科技动态监测方法, 辨识和探测学科知识领域研究的热点, 预测知识领域发展的前沿趋势; Kleinberg 于 2002 年提出了突发监测算法, 它主要关注焦点词和相对增长率突然提高的词。突发词监测法更注重的是研究领域内研究活跃、有潜在影响研究热点的因素, 突发词监测有助于发现、推动学科 (或主题) 研究发展中的微观因素。

1.3.2 国内研究现状

国内在科技监测领域的研究开展得相对较晚, 处于起步阶段, 但也有一些

成果。

在理论探讨和科技监测实践方面,北京理工大学朱东华和袁军鹏(2004)提出了“科技监测”的概念和目标,并开展了基于智能化知识采掘的高新技术监测分析技术研究;中国农业科学院农业信息研究所进行了农业热点及重大事件监测与快速反应机制研究与实践;复旦大学周傲英教授(2001~2003年)开展了基于Web的信息主题的监测与跟踪研究,对互联网科技资源实现了科技监测技术的研究探索;中国科学院国家科学图书馆自2004年12月开始,每月出版《科学研究动态监测快报》,报道各科学领域的国际科技战略与规划、科技计划与预算、科技前沿与热点、重大研发与应用、科技政策与管理等方面的最新进展与发展动态;国家科技图书文献中心在2006年启动了“十一五”国家科技支撑计划重点项目“科技文献信息服务系统关键技术研究及应用示范”,开展了“科技热点动态监测技术研究与应用”的研究工作,由北京邮电大学与北京万方数据有限公司合作,对基于数据挖掘、文献计量学、复杂网络理论的科技热点监测和科技评价方法开展了探索性研究工作。科技热点动态监测技术研究与应用主要是针对互联网中文科技信息资源进行监测,尚未涉及对基于大型科技数据库的海量外文科技文献的动态监测与分析。从科技监测的领域应用上来看,中国医学科学院、中国农业科学院及中国科学技术信息研究所从知识组织、知识抽取、文本表示、聚类算法、知识演化等角度开展了相关的研究,并且在医药卫生领域、科技领域及农业领域开展了初步的科技监测应用。

国内在科技监测方面的理论、方法及系统建设方面尚处于起步、探索阶段,在高新技术监测、基于网络数据的科技热点监测,以及医药卫生领域的科技监测等方面取得了一定的成果,但是,对面向国际科技合作的科技资源监测体系的研究尚属空白。

本书调研分析国际科技资源信息的分布状况及变化趋势,有效利用国外在数据挖掘、科学计量分析、专利数据分析、技术机会分析、科技评价、科技政策、高技术评估等领域的优势,引进国外先进技术和人才,合作研究面向结构化和非结构化国际科技资源信息的数据挖掘、计量分析和可视化技术,开发具有自主知识产权的国际科技资源信息监测分析系统,为国家制定国际科技合作战略和实施国际科技合作计划管理提供决策支持,为产学研各界选择国际合作伙伴、有效利用全球科技资源、开展高水平的国际合作提供信息支持。

国际科技合作资源 信息来源及特点分析

2.1 结构化国际科技资源信息数据库

2.1.1 科学引文索引数据库

科学引文索引 (Science Citation Index, SCI), 创办于 1961 年, 是国际上收集科学论文相互引证最为完备的数据库, 由引文索引、来源索引和轮排主题索引三个主要部分组成。

经过 50 多年的发展, SCI 逐步形成了印刷版、磁带版、光盘版、联机版和网络版等多种版本相结合的国际科技论文检索体系。其中, 网络版科学引文索引 (Web of Science) 是 SCI 发展中的最新代表。1997 年, ISI 在光盘版科学引文索引的基础上发布了 Web of Science (其网址为 <http://www.isinet.com>), 提供网络 ISI 引文数据库检索。网络版科学引文索引涵盖了世界上最重要的自然科学、社会科学、艺术和人文期刊引文数据库, 提供多学科、高质量的研究信息。Web of Science 的出现, 是现代网络技术发展和信息技术更新需求相结合的产物, 其信息交流快捷、准确、全面的特点, 不仅导致国外用户数量的激增, 也引起了我国高等院校、科研单位及信息服务机构的重视。

Web of Science 作为一种综合性文献检索工具, 数据来源于世界上 40 多个国家和地区的学术期刊、专题文集和会议录, 以及大量的专利文献和科技图书等。涉及的学科包括生物学、农学、医学、化学、物理学、地球科学、生命科学等。其中, 引文索引数据库主要包括:

- (1) SCI 3500 种期刊;
- (2) 社会科学引文索引 (Social Science Citation Index, SSCI) 1700 种期刊;
- (3) 艺术与人文引文索引 (Art & Humanities Citation Index, A&HCI) 1150 种期刊;
- (4) 其他专业引文索引。

2.1.1.1 SCI 的选刊原则及要求

SCI 只收录其选定的来源期刊, 而且对有的来源期刊也不是全部收录。据统计, 全世界每年出版的科技期刊多达 10 万种以上。SCI 编者根据“加菲尔德文献集中定律”及“费用-效果”原则选择入选期刊。情报学家加菲尔德在研究文献引用情况后得出的结果表明, 大量 (75% 左右) 的被引用文献出自少数“核心期刊”, 而其余少数被引用文献则分散在大量期刊上, 入选 SCI 的期刊不及全世界科技期刊总量的 6%。

SCI 对入选期刊有一定的要求, 主要包括以下几个方面。

(1) 基本的期刊出版标准: 要求期刊做到及时出版, 遵循国际编辑惯例提供足够的英文信息 (包括文献标题、摘要、关键词、目录), 包括作者的全部联系信息, 以及期刊编辑、审稿人的同行审阅制度等。

(2) 期刊内容: 收录期刊论文的内容能充实 ISI 的数据库, 强调在内容上要能反映新进展, 研究的是热点问题、领导或有特殊意义的领域。严格审稿, 确保内容的科学性。

(3) 国际化程度: 被选期刊能正确反映科学研究的全球性, 表现为多国家的作者群, 引用刊物的多国性等。

(4) 引文分析: 引文数据包括被引频次、影响因子、即时指数、引用半衰期等各项指标, 利用这些指标对期刊进行定量评价。其中, 影响因子是一个很重要的衡量标准, 影响因子高的期刊就是重点期刊。SCI 对入选的刊物进行动态管理, 每年评出有前景的新刊, 淘汰利用率不高的旧刊。

2.1.1.2 SCI 的主要应用

SCI 作为一种大型的、综合性的检索工具, 也是一套比较科学的检索体系。通过揭示引证与被引证的关系, 可以反映当年、几年前、几十年前甚至上个世纪与当前作者或学科之间的关系。这就是“引证分析法”作为一种科学计量方法的独特之处, 该方法已被越来越多的人所认识和采纳。目前, 用引证分析法评价学科的发展状况、预测学科可能的发展趋势和突破, 评价科学家的影响或成就, 评价科学期刊甚至某些重要论文的学术价值等问题早已不是尤金·加菲尔德和他的科学情报研究所所能预料到的, 而是作为一门学科在世界范围内被



采用,成为文献计量学的一个重要组成部分。我国也不例外,广大科技情报人员、科技管理者在理论研究和工作中使得 SCI 及引证分析法得到了广泛应用,具体表现为以下五个方面。

(1) 测定核心期刊和评价期刊、图书、论文质量等方面的应用。核心期刊又被称为“重点期刊”“常用期刊”“基本期刊”,是指那些信息密度大、刊载的文章重要、具有权威性、代表着某一学科的水平、受读者重视的期刊。测定核心期刊是在期刊评价的基础上进行的。测定核心期刊是图书情报系统的一项业务工作,对于文献情报的选择、管理和利用等都有着重要的意义。尤其是在当今期刊数量激增和文献经费日趋紧张的情况下,这项工作显得更为重要。测定核心期刊的关键是要选择和采用较为科学的、具有定量分析的合适方法。目前,常用的方法有布拉德福定律和引文分析法。

(2) 评价作者学术水平的重要参考。引文索引的出版,为文献计量学在人才评价方面的应用开拓了极为广阔的前景。利用引文分析法评价人才主要是根据科学文献与科技人才之间的内在联系来评价。具体来说,科技人员的成就和学术造诣不仅与其发表的文献数量有关,而且与其文献的被引数量有着一定的联系。文献的被引用次数越多,不仅在一定程度上反映了文献的质量和价值,同时也能说明该文献作者在学术界的影响和地位及其对社会的贡献。而且,对于某一作者来说,发表的文献越多,被引用的次数也可能越多。因此,作者文献被引频率既从质量上也从数量上反映了文献价值,因而可以作为衡量和评价作者学术水平高低的一种重要参考依据。

(3) 评价科研机构学术水平的参考。统计某个科研机构的成员发表文章的发文数和被引用的总次数是评价该机构学术成就的一种参考依据,可以从某种程度上证明该科研机构的整体科研能力和科研产出情况。

(4) 科学学研究的重要工具。引文索引具有独特的结构,它既揭示了文献的外部特征,又反映出有关文献在学科内容上的联系。这种由文献引证与被引证关系体现的学科相关性,使大量科学文献以学科形式构成“引文网络”或“引文链”。利用引文索引提供系统资料,对“引文网络”和“引文链”进行分析,便可找出文献间的学科联系,追溯学科知识的进化,从而探讨学科横向的内在结构和纵向的发展变化规律。通过学科间的相互引证率和学科内的“自引率”的研究,可以反映出某学科利用情报源的分布结构,从而为组织专科情报、制定文献比例、合理分配经费等情报管理工作的科学化提供依据。

(5) 预测学中的应用。依据引文索引中“引文链”和“引文网络”,可以预测某一学科的未来发展方向及其应用,可以促进边缘学科或交叉学科的产生,还可以展望国家之间、科研机构之间、科学家之间的相互协作。依据“引证率”和“被引证率”也可以预测未来的学术带头人及其发展潜力,为人才培养提供

依据，为预测科学奖励对象提供参考。

2.1.1.3 SCI 的检索举例

1) 数据库选择

进入 <http://isiknowledge.com>，选择 Web of Science 数据库（图 2.1）。



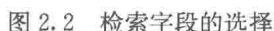
图 2.1 数据库的选择

2) 检索字段选择

如图 2.2 所示，进入数据检索界面后，根据检索项，可以选择从论文的主题、标题、作者、出版物名称等多个角度进行检索。图 2.3 给出了一个具体的检索实例。

3) 检索结果显示

根据检索式得到结果后，可以选择被引频次、相关性、来源出版物等方式进行重新排序，并可利用左侧的工具栏，对检索结果进行精炼和分析，如图 2.4 所示。



2.1.2 美国专利数据库

美国专利商标局 (United States Patent and Trademark Office, USPTO) 成立于 200 多年前, 是美国专利和商标事务的专责行政机构。它提供的服务包括处理专利、商标应用、发布专利和商标信息。USPTO 负责美国专利申请案件, 同时也提供免费查询所有的美国专利数据检索服务。该网站针对不同的信息用户设置了专利授权数据库、专利申请公布数据库、法律状态检索、专利权转移检索、专利基因序列检索、撤回专利检索、延长专利保护期检索、专利公报检索及专利分类等。

2.1.2.1 USPTO 专利数据库收录范围

美国专利授权数据库收录了 1790 年至最近一周 USPTO 公布的全部授权专利文献。该检索系统中包含的专利文献种类有: 发明专利、设计专利、植物专利、再公告专利、防卫性公告和依法注册的发明。

按照专利资料的不同类型, USPTO 收录的美国专利信息分为已授权专利文献和申请中的专利文献两部分: 已授权专利文献 (issued patents) 包括自 1976 年起已公开的专利全文及 1790 年起的专利全文影像; 申请中的专利文献 (published applications) 包括自 2001 年 3 月 15 日起已公开的专利全文及其专利全文影像。USPTO 专利数据库在网上可免费使用, 并且提供近两年美国专利申请件查询。数据库每周二更新一次。

2.1.2.2 USPTO 专利数据库的功能与特色

USPTO 专利数据库主要有以下三个特色:

(1) 免费提供的专利全文与全文影像资料, 提供 References Cited 连接功能, 可以立即链接到引用此专利及此专利引用的美国专利。

(2) 提供专利分类号辅助检索界面 (<http://www.uspto.gov/go/classification/>), 以协助使用者检索复杂的专利分类号系统。

(3) 双向超链接的方式, 取得某篇专利所引用的专利与引用此专利的其他相关专利

2.1.2.3 检索数据库及检索方式

1) 专利授权数据库

该检索系统中包含的专利文献种类有: 发明专利、设计专利、植物专利、再公告专利、防卫性公告和依法注册的发明。其中, 1790~1975 年的数据只有



图像型全文，可检索的字段只有专利号、美国专利分类号和授权日期 3 个；1976 年 1 月 1 日以后的数据除了图像型全文外，还包括可检索的授权专利基本著录项目、文摘和文本型专利全文数据，可通过最多 31 个字段进行检索。

2) 专利申请公布数据库

可供用户从 23 种检索入口检索 2001 年 3 月 15 日以来公布的美国专利申请公布文献，同时提供文本型和扫描图像型全文美国专利申请公布说明书，可供公众进行美国专利申请公布的全文检索及浏览。其专利申请公布说明书的起始号为 20010000001。

3) 专利公报检索

可浏览 Patents OG（最近 52 期电子形式美国专利公报全部内容）和 OG（Official Gazette，最近 10 年美国专利公报中的“Notices”）内容。

4) 专利分类检索

可检索最新版本的美国专利分类表中的相关主题的分类号，并直接浏览该分类号下所属专利文献全文。

5) 法律状态检索

通过查找专利缴费情况确定专利是否提前失效，通过查找撤回的专利确定专利是否在授权的同时被撤回，通过查找专利保护期延长的具体时间确定专利的最终失效日期，通过查找继续数据确定专利是否有继续申请、部分继续申请、分案申请等相关联的情报，还可以查看专利的审批情况及与该专利有关的其他信息等。

6) 撤回专利检索

收录至最新公布日撤回的所有专利的目录。

2.1.2.4 专利检索举例

1) 选择 USPTO 专利检索功能

进入 USPTO 官方网站 (<http://www.uspto.gov>)，点击加框区域进入专利检索页面，如图 2.5 所示。

2) 选择专利检索数据库与检索方式

在 USPTO 的专利全文与图片数据库（USPTO Patent Full-Text and Image Database, PatFT）中，提供了三种检索方式：快速检索，高级检索，专利号检索。点击加框区域，选择高级检索方式，如图 2.6 所示。