

专利审查研究系列丛书

(第十辑 检索特辑)

专利审查研究

国家知识产权局专利局专利审查协作北京中心◎组织编写

郭 雯◎主编



知识产权出版社

全国百佳图书出版单位

专利审查研究系列丛书

(第十辑 检索特辑)

专利审查研究

国家知识产权局专利局专利审查协作北京中心◎组织编写

郭 雯◎主编



知识产权出版社

全国百佳图书出版单位

图书在版编目 (CIP) 数据

专利审查研究. 第十辑, 检索特辑/郭雯主编. —北京: 知识产权出版社, 2019. 1
ISBN 978-7-5130-5999-2

I. ①专… II. ①郭… III. ①专利—审查—研究—中国 IV. ①G306.3

中国版本图书馆 CIP 数据核字 (2018) 第 287680 号

内容提要

文献检索是获取知识、信息的基本手段, 是从事文献调研、科学研究、发明创造、科技查新的重要前提, 也是专利审查程序中的关键步骤和专利审查工作的重要组成部分。国家知识产权局专利局专利审查协作北京中心在专利审查工作中, 持续开展检索研究, 不断总结检索经验, 近年来加快推进检索智能化工作, 取得了较好的成效。为了推广和分享检索学术研究成果, 本书精选了 40 篇高质量检索论文结集成册, 内容涉及专利检索理论、检索实践、检索信息化及工具等方面的研究。可为从事相关工作的人员的文献检索实践提供帮助、借鉴和指导。

责任编辑: 王祝兰

执行编辑: 周 也

责任校对: 潘凤越

责任印制: 孙婷婷

专利审查研究 (第十辑)

——检索专辑

国家知识产权局专利局专利审查协作北京中心 组织编写

郭 雯 主编

出版发行: 知识产权出版社有限责任公司

社 址: 北京市海淀区气象路 50 号院

责编电话: 010-82000860 转 8555

发行电话: 010-82000860 转 8101/8102

印 刷: 北京建宏印刷有限公司

开 本: 720mm×960mm 1/16

版 次: 2019 年 1 月第 1 版

字 数: 540 千字

ISBN 978-7-5130-5999-2

网 址: <http://www.ipph.cn>

邮 编: 100081

责编邮箱: wzl@cnipr.com

发行传真: 010-82000893/82005070/82000270

经 销: 各大网上书店、新华书店及相关专业书店

印 张: 28.25

印 次: 2019 年 1 月第 1 次印刷

定 价: 120.00 元

出版版权专有 侵权必究

如有印装质量问题, 本社负责调换。

《专利审查研究（第十辑）》编委会

主 编 郭 雯

副主编 夏国红 崔 峥 朱晓琳

编 委 （按姓名拼音排序）

陈玉华 郭震宇 李意平

刘新民 马秋娟 聂春艳

石贤敏 田 虹 王晓峰

张 蔚 仲惟兵

《专利审查研究（第十辑）》编写组

审 稿 （按姓名拼音排序）

曹 旭	陈 辉	陈韦杰
范肖凌	付 佳	甘 雨
葛加伍	韩 冰	胡敬东
江定国	焦永涵	孔 昕
李 闻	刘 昶	马 然
马艳林	任志安	宋 超
孙晓明	田 远	王 婷
谢幸初	严 华	杨 明
杨 倩	于冶萍	张 宇
张颖浩	赵 良	

编 辑 宋 超 王秋丽

前 言

当前围绕国家经济社会发展需要，国家知识产权局正在深入实施知识产权战略，奋力开拓新时代知识产权强国建设新局面。为了提升专利审查质量和效率，国家知识产权局专利局专利审查协作北京中心（以下简称“专利审查协作北京中心”）持续加强业务能力建设，努力探索世界一流专利审查机构建设路线，并积极推进学术研究工作，取得了丰硕成果。

文献检索是获取知识、信息的基本手段，是从事文献调研、科学研究、发明创造、科技查新的重要前提。从某种意义上说，国家之间的科技能力的差距，归根结底是国民吸收信息能力及从事创新工作能力的差距，而文献检索能力正是衡量信息吞吐能力的重要标识。文献检索也是专利审查程序中的关键步骤，是专利审查工作的重要组成部分，专利审查协作北京中心在专利审查工作中，持续开展检索研究、不断总结检索经验，近年来加快推进检索智能化工作，取得较好的成效，以解决文献检索这一制约审查质量和效率提升的问题。

为了推广和分享检索学术研究成果，本书精选 40 篇高质量检索论文结集成册，内容涉及专利检索理论、检索实践与经验、检索工具运用等方面的研究。我们衷心地希望《专利审查研究（第十辑）——检索特辑》的出版能够为文献检索实践提供帮助、借鉴和指导。

由于编者水平有限，书中不当之处，敬请广大读者批评指正。

目 录

检索理论研究

- 程式化检索方法的提出与实现/ 赵 良 刘以成 (3)
- 基于技术特征频次统计的检索方法研究/ 李瑶琦 赵 良 刘以成 (17)
- 基于最优搜索理论的文献排序策略/ 刘 昶 范成博 池 娟 (29)
- 在 S 系统中建立高鲁棒性相似检索模式的初步研究/ 胡敬东 (44)
- 基于语义相似原理程式化扩展关键词/ 黄 蕾 赵 良 刘婉姬 (55)
- 利用层级统计分析方法实现分类号程式化扩展/ 胡敬东 赵 良
任志安 (64)
- “块检索”理论下的文献排序与浏览效能提升研究/ 陶 冶 赵 良 ... (77)
- 一种从专利数据中进行技术信息挖掘的方法
——基于 TFTA 模型的检索/ 张然兮 韩 冰 张 蔚 张超然 ... (88)
- 日本分类体系程式化扩展的初探/ 卞志家 赵 良 (95)
- 基于语义相似程式化扩展 IPC 分类号的实证研究/ 杨 明 赵 良 (104)

检索实践与经验

- 基于众筹网 Kickstarter 的关于创意产品的实用新型专利检索
新思路/ 梁 磊 靳红蕾 (119)
- 关于有效提高外观设计检索效率的分析研究/ 孙晓璐 陶海琴 (131)
- 专题专利文献数据库平台的构建及其商业化运营/ 张 博 (143)
- 公式检索策略研究/ 王晓燕 王雪婷 孙国辉 苏 菲 (150)
- 生活快消品领域数据抓取技术在融合检索中的应用初探/ 张 蔚
孙 敏 韩 冰 张然兮 (161)

区块链技术的专利检索研究/ 张颖浩 张 雪	(172)
浅谈互联网检索在实用新型明显新颖性审查中的应用/ 王 娟 张剑云	(185)
追本溯源,把握实质,基于评判预期的高效检索/ 陈正军 杨 玲	(197)
根据逻辑推理进行检索策略的调整/ 葛加伍 白芳芳	(210)
关于思维导图在专利审查检索中的应用研究/ 黄海云	(217)
浅谈检索流程的优化/ 李 闻	(225)
探讨在专利库中检索表达式如何体现发明构思/ 严 华	(232)
通信领域非专利技术标准及相关检索方法探讨/ 傅 颖 王 健 孙 洁	(243)
从裁决漏检案例看补充检索/ 孙红花	(253)
化学产品的制药用途发明检索中的几种扩展应用/ 刘军政	(262)
CPC 分类体系在空调领域 (F24F) 的检索应用/ 武利媛 曹斌宏 靳艳梅	(268)

检索工具运用

“一站式”检索工具提升检索效能的探讨/ 李瑶琦 赵 良	(283)
检索中基于人工智能快速站位本领域技术人员的途径/ 李 博 王 蕊 陈正军	(296)
在 STN 中检索化学反应的技巧/ 于冶萍	(308)
开源平台在计算机领域检索中的应用/ 欧晓丹 唐 娜 魏 峰 唐 嫣	(330)
基于 Patentics 的专利分析多维度碎片去噪/ 方 蕾 王 东	(338)
“一站式”检索工具提升医药领域检索效能探析/ 涂海华 赵 良	(352)
Patentics 检索全策略 ——智能语义结合人工干预/ 刘婉姬	(364)
合理选取语义排序基准 提高 Patentics 检索的准确性/ 李 安 赵 良	(376)
巧用 Patentics 实现块检索和二维检索结合的检索方法/ 刘时雄 张春伟 袁 野	(388)
基于 Patentics 中“N 阶搜索”功能对组合物相关申请进行分组检索的 初探/ 陈 皓 赵 菁	(397)

方剂识别软件在中药复方专利辅助检索中的应用/ 吕茂平 师晓荣	(405)
CNKI 数据库在中药领域的检索实用技巧/ 段 炼	(413)
Patentics 和“一站式”检索在农药组合物检索中的应用/ 李 姮 杨 杰	(418)
利用 Patentics 检索外文专利文献的技巧/ 李发喜 吴 冰	(427)

检索理论研究

一、引言

检索理论是信息科学的重要组成部分，也是信息检索工作的理论基础。随着信息技术的飞速发展，信息检索工作的重要性日益凸显。本文旨在探讨检索理论的研究现状、存在的问题以及未来的研究方向。首先，本文回顾了检索理论的发展历程，从早期的文献检索到现代的信息检索，再到当前的网络检索，检索理论不断演进。其次，本文分析了当前检索理论研究中存在的主要问题，如检索效率、检索精度、检索个性化等方面。最后，本文提出了未来检索理论研究的几个方向，包括智能化检索、个性化检索、多源异构数据检索等。通过本文的研究，希望能够为检索理论的研究者和实践者提供一些有益的参考。

关键词：检索理论；信息检索；网络检索；智能化检索；个性化检索

程式化检索方法的提出与实现

赵 良 刘以成

摘 要：本文提出程式化检索方法（Stylized Retrieval），并概述了专利审查协作北京中心前期所做的探索性工作。它是利用人、计算机与数据三要素相互关系，通过设置运算规则实现检索流程的智能化，包括检索要素的确定、关键词和分类号的扩展、检索式的构建和调整，以及文献排序和浏览。相比较于传统的检索方法，程式化检索充分发挥了计算机和关联数据的作用，而检索人员专注于发挥判断、调整和决策的作用，最终实现整体检索质量和效率的提升。本文从原理、规则、实现路径等方面全面阐述程式化检索的理论基础，为后续研究工作的推进提供理论依据。

关键词：程式化检索方法 S 系统 语义搜索智能化

一、引 言

深入贯彻习近平总书记关于提高知识产权审查质量和审查效率的重要指示精神，服务于供给侧结构性改革的经济主线，坚持质量第一、效益优先的原则，落实新形势下加快知识产权强国建设的若干意见，落实专利质量提升工程实施方案，加快世界一流专利审查机构的建设，迫切需要着力提升专利审查质量和效率。尤其是在当前专利申请量庞大、文献信息“爆炸式”增长的时代背景下，专利审查工作面临激增的审查负荷和有限的审查力量之间的突出矛盾，提升整体审查质量和效率也就成为当前的重点工作之一。

检索是专利审查程序中的一个关键步骤，是一个非常消耗审查时间的重要

步骤，例如，欧洲专利局平均检索时间为 13 小时。^①《专利审查指南 2010》第二部分第七章对检索作出了一般性的规定，包括检索资料、检索主题、时间界限、检索准备、对发明专利的检索、防止重复授权的检索、中止检索以及特殊情况的检索等。此外，各个审查领域还制定了基本检索规范，对上述一般性规定进行细化，以达到规范检索过程。上述规定或规范主要是起到引导作用，而实际检索工作往往非常复杂。提升检索的质量和效率，归根结底还是要落实到检索的自动化和智能化水平上。专利审查协作北京中心的研究人员开展了一系列的探索性工作并逐渐形成了程式化的检索方法，希冀在计算机的辅助下实现智能化检索。

二、检索工作面临的问题

如图 1 所示，提高检索质量和效率，无外乎是做好检索系统和检索流程两个方面的工作。当前，联机检索系统取得了很大的进展，提升了检索效率；但检索流程的智能化始终徘徊于起始阶段，各个环节均只能依靠检索人员自身的能力和经历“人工地”去完成，加之个案的情况不同，导致检索过程费时费力、检索结果良莠不齐，严重制约检索质量和效率的进一步提升。

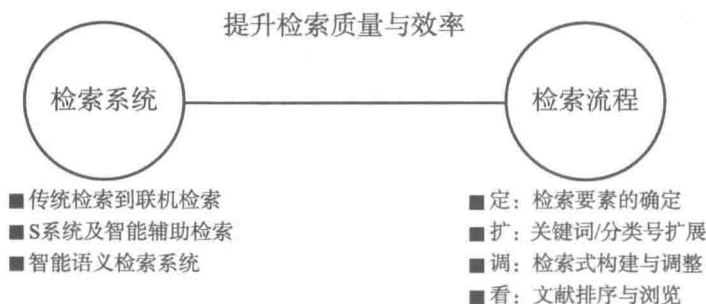


图 1 检索系统与检索流程的关系

对于检索系统来说，电子化、数字化、网络化的联机检索已经取代了传统的纸件资料、胶片、光盘等，使得文献数据的存储和索引方式发生了质的飞跃，大大提高了文献获取的便捷性和检索工作的效率。这些检索系统包括 S 系

^① 葛树，蒋彤．改进实体审查质量的政策研究以及局级检索质量管理模式研究 [R]．国家知识产权局学术委员会（2010 年度一般课题研究项目，ZX201003），2011：156．

统及其智能辅助检索系统^{①②}、基于空间向量理论的语义检索系统等^③。其中，S系统所依赖的布尔逻辑运算是目前最基本、最成熟的检索理论，但布尔逻辑运算理论本身对检索要素的确定及表达要求很高，一旦检索要素的确定或者关键词、分类号的表达不当，会导致漏检、错检的发生。另一个是基于空间向量理论的语义检索系统（有时也称之为语义搜索、语义排序），它能自动提取本申请文本信息并在专利数据库中找到与之语义相关度匹配的专利文献^④，但语义检索系统后台计算机的算法不透明、过程不可控，检索人员无法参与到机器的运算环节中，无法事先预期对比文件在检索结果中出现的位置。因此，单纯的检索系统的改进会遇到发展的“瓶颈”，也就需要我们从检索流程的智能化方面取得突破。

对于检索流程，简单地讲就是《专利审查指南2010》第二部分第七章所涉及的“定”（确定检索要素）、“扩”（扩展关键词与分类号）、“调”（构建与调整检索式）、“看”（文献排序与浏览）四个环节。当前，上述四个环节基本上只依赖于检索人员自身的能力和经验的完成，缺少行之有效的信息化检索工具。这是智能化检索的空白和薄弱环节，是掣肘检索质量和效率进一步提升的关键步骤。检索流程所涉及的“定”“扩”“调”“看”四个环节流程长、检索耗时费力，个案具体情况不同且伴随着诸多人的主观性因素，检索效果取决于受到检索人员专业背景、外语能力、检索能力、检索经验等因素的影响，漏检、错检等情况时有发生。因此，着力推动检索流程的标准化、自动化非常重要。

三、程式化检索方法的提出

（一）程式化检索的含义

经过前期的研究和摸索，计算机辅助实现检索流程的智能化是可行的。程式化检索是专利审查协作北京中心提出的，利用人、计算机与数据三要素相互关系，通过设置运算规则实现检索流程的智能化，包括检索要素的确定、关键词和分类号的扩展、检索式的构建和调整，以及文献排序和浏览。值得注意的是，本文所指的程式化检索（Stylized Retrieval）强调在人、计算机和数据三要

① 国家知识产权局智能辅助检索项目组. 智能辅助检索系统用户手册 [Z], 2017: 44-50.

② 马昊, 宫宝珉. 分领域检索手册 (整体组) [R]. 国家知识产权局学术委员会 (2015年度专项课题, ZX201509_08), 2015: 39-41.

③ 李惠. 语义检索排序方法研究进展 [J]. 现代情报, 2015, 35 (2): 159-164.

④ 赵良, 刘以成. 关键词程式化扩展提升专利文献检索效能的应用模式研究 [J]. 审查业务通讯, 2017, 23 (5): 63-72.

素共同作用下按照步骤进行检索，并非计算机实现的单纯化的程序化检索（Programmable Retrieval）。

（二）程式化检索的三要素

如图2所示，人、计算机和关联数据构成了程式化检索的三个基本要素。以数据为基础建立起检索规则，计算机发挥运算的优势，而检索人员发挥判断、调整和决策的作用。简言之，让人和机器去做各自擅长的事情。

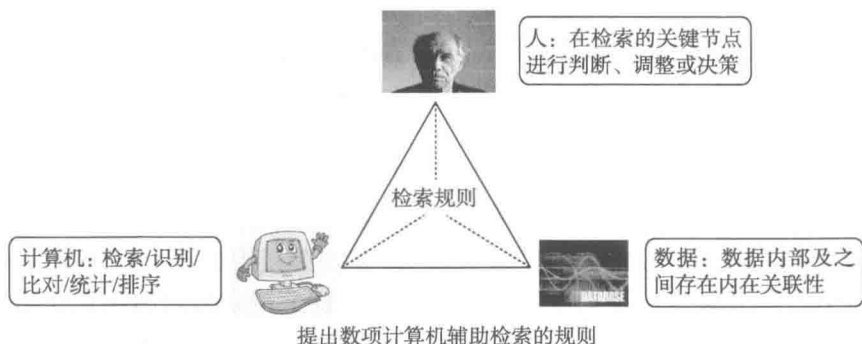


图2 程式化检索理论的三要素

第一，检索人员的工作应该是专注于判断、调整和决策。个案的具体情况不同，需要检索人员根据检索的预期以及计算机运算的结果在关键节点上进行必要的判断。此外，允许检索人员的参与，有利于将人的智慧融入到程式化检索当中。

第二，计算机的优势在于按照既定的程序代替人的简单劳动，而且运算速度快。程式化检索以计算机运算为主，充分发挥计算机在数据识别、比对、统计分析等方面的优势，并将分析结果呈现给检索人员进行判断、调整和决策。

第三，数据关联性很重要，是提取检索规则并进行计算机运算的基础。程式化检索的本质是数据挖掘（Data Mining），是从大量的数据信息中挖掘关联规则并用于检索工作。关联性是文献的特征之一，一些文献存在“天然的”关联性，例如同族文献；另一些文献则可以通过检索式来关联，例如，因享有相同的检索式，检索结果之间的文献出现关联性。

其中，基于数据关联性提取检索规则是实现程式化检索最难的部分。规则源于数据的关联性，并成为人、计算机和数据三要素之间的中间链接环节。只有建立起检索规则之后，计算机才能对关联数据进行处理。不仅如此，针对检索流程所涉及的“定”“扩”“调”“看”环节，需要提炼出不同的计算机可运算的检索规则，从而实现程式化的检索。

四、程式化检索方法的实现

（一）检索要素的确定

检索要素的确定是检索的起点。检索人员在阅读申请文件、充分理解了发明内容并初步确定了分类号和检索的技术领域后,应进一步分析权利要求,确定检索要素。其中,基本检索要素是体现技术方案的基本构思的可检索的要素。一般地,确定基本检索要素时需要考虑很多因素,包括技术领域、技术问题、技术手段、技术效果等方面。不仅要考虑技术方案明确的技术特征,必要时还应当考虑技术方案中的某些技术特征的等同特征。检索要素的确定存在诸多难点问题,例如,权利要求技术特征很多或者技术特征表达比较晦涩、不明确时,基本检索要素就不容易提取或概括出来。

以往,确定检索要素的信息通常来源于申请文件,但申请文件的信息往往非常有限,常常不足以让检索人员快速抓住可检索的要素。笔者认为,确定检索要素应该以权利要求及申请文件的信息为基础,同时也应充分地从现有技术整体去考量检索要素,尤其是查看与本申请近似文献对相同或相近检索要素的表述方式,避免主观猜测或臆想检索要素,从而提高检索要素确定的准确性和有效性。基于这一原理。程式化检索使用计算机代替人去浏览关联文献,获取必要的检索信息,而人所做的是根据检索要素选取的预期,从机器推送的结果挑选出所需的检索要素即可。尝试性地,笔者提出了基于语义原理,利用相似的文本(例如语义相似文本)具有相似的索引词及分类号的规则^①,用于基本检索要素的确定。

作为示例,图3为基于语义相似原理的确定检索要素。首先,根据本申请权利要求的技术方案或全文信息,利用语义搜索系统获取一组与本申请文本相似的关联文献。其次,计算机统计分析关联文献的索引词(关键词)和分类号信息,呈现关联文献中关键词与分类号的分布情况。最后,检索人员根据本申请权利要求确定检索要素的预期,从统计分析结果中挑选适合的基本检索要素。如果统计分析结果的信息未达到预期,可以重新调整关联文献获取的范围,重新由计算机推送候选的检索要素信息。

（二）关键词和分类号的扩展

关键词和分类号是检索要素最重要的表达形式,理论上所有的文献都可以采用合适的关键词或分类号检索得到。然而,关键词会因撰写人员的国别、习

^① 赵良,王秋丽,刘以成.检索要素程式化扩展方法的理论探讨[J].审查业务通讯,2017,23(11):26-33.

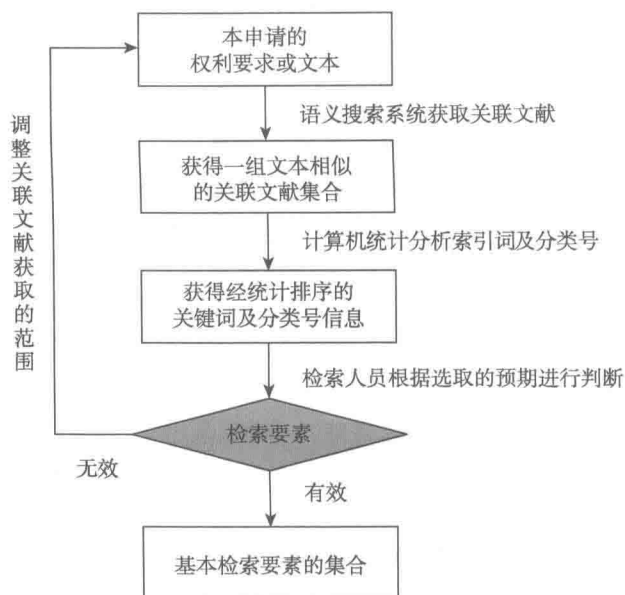


图3 基于语义相似原理的确定检索要素

惯、知识水平以及文献本身的内容等而不同，分类号会因分类体系、应用/功能规则及分类员的理解而不同，这导致检索人员在有限的时间内很难穷尽文献的关键词或分类号的表达形式。为了全面检索，通常需要尽可能地扩展关键词、分类号等的多种形式，并将不同表达形式检索到的结果合并作为针对该检索要素的检索结果。因此，关键词和分类号的扩展一直是检索的难点，也是导致文献漏检的最主要的因素。^① 为了解决该问题，我们建立了三种程式化扩展的方法。

1. 检索历史原理

基于检索历史原理的程式化扩展关键词或分类号是一条重要的扩展思路。如图4所示，在使用检索系统（例如S系统）构建检索式的过程中，检索式会记录每个检索人员使用的各种关键词信息，关键词之间因布尔逻辑运算构成关联数据或关联信息。其中，“OR”的检索式通常涉及的是同义词或近义词，有时还会涉及上下位或其他相关概念。以检索式“OR”涉及的关联数据作为检索要素库，对用户群体所使用的关键词进行收集、整理和统计分析，就能汇集

^① 赵良，万新，杨秦．合理选取关键词提高医药主题检索的准确率[J]．审查业务通讯，2014，20（2）：40-47.