

# 基于IPC的专利数据组织初探

□ 宋群 吴广印 / 中国科学技术信息研究所 北京 100038

**摘要:** 专利作为技术信息最有效的载体, 具有很大的实用性和可靠性, 近年来越来越受到人们的重视。文章提出一种有效的专利数据信息组织方式, 旨在为检索用户提供有序的检索结果, 提高专利信息资源的利用率。

**关键词:** 专利检索, 数据组织, IPC

DOI: 10.3772/j.issn.1673—2286.2013.04.010

## 1 引言

专利是一种特殊的文本, 属于知识产权的一部分, 它是一种无形的财产, 是记录和传播科技成果的重要载体, 具有很大的实用性和可靠性。联合国世界知识产权组织曾得出这样的统计结果: 专利文献占据了全世界每年新诞生技术的90~95%, 而其中70%的信息是只能从专利文献中获取的, 因此, 它越来越受到人们尤其是科技工作者的重视, 世界上每年专利申请数量的增长速度已经超过了100万件, 而专利的累计总量也在4000万件左右, 互联网的迅速发展使得信息的传播越来越容易, 但面对这样一堆庞大的数据, 如何方便、快速地获取有价值的信息, 一直是人们关注的话题, 各类专利检索平台和检索软件因此应运而生。

## 2 专利信息检索平台现状及存在的问题

随着互联网的迅速发展, 免费的专利信息资源在网络上日益丰富, 通过互联网来检索专利信息也成了人们获取所需资源的主要手段。目前已有许多机构通过互联网提供专利查询服务, 其中大部分的检索平台是由国家知识产权局及下属部门或者专利数据研究公司主办, 国内较有权威和影响力的检索平台包括中国国家知识产权局网站、中国知识产权网、中国专利中心的

中国专利数据库检索系统、中国专利信息网、万方数据知识服务平台专利检索系统、CNKI中国专利数据库、国家科技文献资源网络服务系统、Soopat专利搜索、Patentics等, 其中国家知识产权局是我国专利行业最权威的政府部门, 其专利数据库收录自1985年以来公布的全部中国的专利信息, 并可浏览到各种说明书全文及外观设计图形, 面向公众提供免费专利检索服务。各个检索平台的基本情况如表1所示。

从表1中我们可以看出, 这些比较权威的网络专利检索平台收录的数据资源均始于1985年, 各个检索平台的收录专利类型、检索界面、检索方式以及检索字段都是大同小异, 而使用相同的分类号进行检索测试, 发现每个网站的检索结果也基本相同, 这表明国内主要的专利信息机构的数据源也基本一致, 虽然在数据库的更新速度、专利相关信息的提供范围等方面可能有些差异, 比如国家知识产权局网、中国专利信息网和Patentics等提供了二次检索功能, 但差异也仅仅是前两个平台的二次检索的关键词需要用户自己作判断, 而Patentics会根据检索的结果自动总结出多个关键词提供给用户; 同时国家知识产权局、知识产权网、Soopat、Patentics提供了“专利统计分析”服务功能, 前三者的统计分析是基于已检索专利, 其中Soopat网站设置了专利对比功能, 且它的专利统计分析功能是免费的, Patentics采用的是检索、分析、提取信息、再检索

\* 本文得到国家高技术研究发展计划(863计划)重大专项支持(编号: 2011AA01A206)。

表1 各检索平台基本情况

	收录范围	专利类型	检索界面	检索方式	检索字段	二次检索	专利分析
国家知识产权局	1985年至今	发明、实用新型、外观设计	中文	字段检索、IPC分类号检索、法律状态检索	公开号、申请号、申请日、名称、主分类号、分类号、申请人、发明人、代理人	有	有
知识产权网	1985年至今	发明、实用新型、外观设计	中文	字段检索、IPC分类号检索、法律状态检索、行业导航检索	申请号、申请日、公开号、名称、主分类号、分类号、申请人、发明人、代理人等	有	有
专利信息检索系统	1985年至今	发明、实用新型、外观设计	中文和英文	字段检索	专利号、名称、申请日、公开日、分类号、申请人、发明人等	无	无
专利信息网	1985年至今	发明、实用新型、外观设计	中文	字段检索	申请号、公开号、公告号、分类号、公开日、名称、申请人、发明人、代理人等	有	无
万方数据专利检索	1985年至今	发明、实用新型、外观设计	中文	字段检索、IPC分类号检索、法律状态检索	名称、申请号、申请日、公开号、发明人、申请人、代理人、主分类号、分类号等	有	无
CNKI专利数据库	1985年至今	发明、实用新型、外观设计	中文	字段检索	名称、摘要、申请号、公开号、分类号、主分类号、申请人、发明人、代理人等	有	无
Soopat	1985年至今	发明、实用新型、外观设计	中文	字段检索、IPC分类号检索	申请号、申请日、名称、公开日、摘要、公开号、分类号、主分类号、申请人、发明人、代理人等	无	有
Patentics	1985年至今	发明、实用新型、外观设计	中文和英文	字段检索	专利号、发明人、名称、公开日、申请日、摘要、申请人、代理人、IPC分类号等	有	有

的方式,将检索和分析结合到了一起。

目前网络上的大部分专利数据库都具备了这些检索的基本功能,方便了人们检索专利文献,但随着专利信息数量的迅速增长、检索技术的日趋成熟,人们对于检索结果的要求也越来越高,这就需要检索机构能够对专利数据信息进行有效的组织并呈现给用户,但现在大多数网站都只是简单地将检索结果罗列出来,未能将数据信息有序地反馈给用户,很难让使用者将精力集中在所需要的专利信息上,用户也就不能充分利用得到的信息资源。以在中国知识产权网的专利信息服务网平台检索为例,用“计算机安全”作为关键词进行检索,共得到22页212条检索结果,浏览默认排序的前5条检索结果,我们发现,对应标题为计算机安全装置、计算机安全网卡、计算机安全控制器、计算机安全防护系统及方法、计算机安全防护方法及计算机安全防护工具的IPC分类号分别为G06F1/00、H04L12/24、

G06F15/00、G06F21/00、G06F21/00,最后两条专利具有相同的分类号,属于同一个IPC大组,又与第一条和第三条专利属于同一个IPC小类,第二条专利与其他四条不属于同一个部,虽然检索的结果都包含检索词,但所属的分类并不同,甚至有很大差异,但是国内的大部分专利检索平台都只是将这样不规律的检索结果直接反馈给用户,这就为用户有效利用专利信息带来了不便,也在很大程度上降低了用户利用信息的效率,因此,检索结果的有效组织具有重要意义。

而谈到专利信息的组织,熟悉专利的人想到的都会是国际专利分类(IPC)。IPC是最早的专利文献分类组织工具,作为一种语言独立的专利文件检索工具,对于海量文献的组织与检索作出了很大的贡献,虽然随着专利文献的不断增长,出现了一些其他的专利分类体系(如美国专利分类法、英国专利分类法等),但IPC一直是世界范围内唯一通用的国际专利分类,是世

界各国必须执行的分类体系,IPC使得各国专利文献获得了统一分类,它的权威性是不容置疑的,每件专利都有一个或多个分类号,通过它可以避免因关键词选择不当造成检索结果不全面或不准确,因此,如果能够充分利用国际专利分类体系,那么实现专利信息的有效组织将会变得容易。然而在国内的众多专利信息检索平台中,很多网站甚至都没有基于IPC的检索方式,即使有一些网站可以进行IPC分类导航检索,如国家知识产权局网站,但其提供的就是查询者可以将IPC分类表依照分类次序由“部一大类一小类”逐级展开,通过点击IPC分类号进行检索,然而由于专利信息平台的服务对象大部分并不熟悉IPC的分类内容,也就很难确定想要检索的专利属于哪一个IPC部,因此这一部分使用者在实际查询时很少直接使用导航检索,IPC分类导航功能的意义对于这些使用者并不显著。

### 3 基于IPC的专利数据组织

目前网络上使用的IPC的检索平台基本上都是为方便初学者而设计的辅助检索方式,这并没有发挥IPC真正的优势,如果想要从专利检索中获取更多有价值的信息,有效的数据组织方式是关键,而由于每一项专利都有其对应的分类号,因此使用IPC组织专利数据是一个很好的办法,重新组织的专利数据在数据库中的存储形式及示例数据如图1所示。

基于IPC的专利数据组织的基本思想是在存储专利数据时,将具有相同IPC分类号的专利信息划分为一类,存储在同一张表中,并根据专利的分类号在IPC分类体系中所处的不同深度赋予其相应的层级标识,再将专利信息的申请人、申请数量、日期等属性添加到表

的每条记录中,这样,对于检索的每一条结果,都很容易获得所有具有相同IPC的其他专利数据,如果用户需要,就能够快速地将所有具有相同分类号的专利信息反馈给检索者,同时可以根据层级关系,将相应数据的分类号的上级与下级专利信息一并返回给用户,这样可以在很大程度上帮助检索者发现自己未留意或被埋在很多检索结果深处的有用的专利信息。

在使用IPC分类号作为专利数据的标识对数据重新进行组织以后,我们将很容易得到具有相同分类号和相应上下层级的专利数据。仍以上面进行的检索为例,如果用户需要,那么计算机安全防护系统及方法和计算机安全防护方法及计算机安全防护工具以及所有的IPC分类号为G06F21/00的专利能够被快速方便地组织在一起返回给用户,同时该分类号的上级与下级数据信息也可以一并返回,如分类号为G06F21/02的专利电脑电源的无线加密锁电路和分类号为G06F21/20的计算机安全取证装置等等。我们知道,专利分类中最具有权威性的是IPC分类方法,它把专利信息中互相关联最紧密的数据组织到一起,分在相同或具有层级关系的分类号下,因此这样获得的专利数据与用户想要的信息是最相关的,也就能为用户提供最可能需要的信息。

将专利数据基于IPC组织起来的另一个好处是方便了对某一领域的统计分析。对于任何机构而言,如果制造出来的产品是已经存在的,或者是使用了相似的技术,那将会造成很大的研发浪费,还有可能因不小心侵权而产生法律纠纷,而对于企业组织,专利信息具有更多的经济价值,因为专利是竞争者之间唯一的不得不公开的关键信息,通过深度发掘专利数据,可以方便地进行各类统计分析并生成图表,从中能够得到重要的围绕专利的信息与知识,再进一步通过细致、综合、严密、

	IPC	层级	专利总数量	发明专利数量	实用新型数量	外观设计数量	分析维度类型	分析维度名称	fid	申请专利号
4	A01B1/00	NULL	1	0	1	0	申请人	贺庆宾	4	CN92226113.X
5	A01B1/00	NULL	1	0	1	0	申请人	吉文	5	CN93204803.X
6	A01B1/00	NULL	1	0	1	0	申请人	徐继先	6	CN94215275.1
7	A01B1/00	NULL	1	0	1	0	申请人	董昭夫	7	CN96207780.1
8	A01B1/00	NULL	1	0	1	0	申请人	魏明泰	8	CN96212567.9
9	A01B1/00	NULL	1	0	1	0	申请人	刘志华	9	CN97218957.7
10	A01B1/00	NULL	1	1	0	0	申请人	四川省万县市农业科学研究所	10	CN97107633.2
11	A01B1/00	NULL	1	0	1	0	申请人	焦路波	11	CN98220618.6
12	A01B1/00	NULL	1	0	1	0	申请人	吴昌生	12	CN98218983.4
13	A01B1/00	NULL	1	0	1	0	申请人	许秋林	13	CN98251878.1
14	A01B1/00	NULL	1	0	1	0	申请人	周怀翰	14	CN00205157.5
15	A01B1/00	NULL	1	0	1	0	申请人	黄以兰	15	CN01245467.2

图1 专利数据存储格式示意图

相关的分析,就会很容易地完成对某技术领域现状的分析,识别关键技术,也能够检测竞争对手在技术方面的竞争力,让企业做到知己知彼。近年来,专利信息分析凭借能够提供发明创造的轨迹、揭示技术发展趋势、展现竞争态势以及为发展战略制定提供参考等作用越来越受到人们的重视,无论用户的目的是了解同一领域的研发动态、主要申请人和发明人,还是为了借鉴其他发明进行科技创新和避免知识产权的侵权行为,分类存储的专利信息都为更方便快速的反馈结果提供了保障。

将专利数据基于IPC组织在一起也能够为国家的各个专利机构审批专利提供一定的参考价值。将同一类的专利数据放在一起,并为处于不同层级的信息赋予标识,在检索新申请的专利时,同类的信息以及上级

与下级的信息都能够快速获取,这样就能够很方便地知道是否已经存在相同或类似的技术。

## 4 结语

专利信息作为技术信息最有效的载体,囊括了全球90%以上的最新技术情报,如此巨大的信息资源如果能够充分加以利用,无论对于科研机构进行科学研究、企业进行发展决策还是国家进行战略决策,都将是取之不尽用之不竭的宝库,而有效的数据组织方式是充分利用信息的前提,国际专利分类法凭借其在专利分类方面的专业性和权威性,必将受到越来越多的检索平台的推广。

## 参考文献

- [1] 李鹏.国际专利分类的困境与出路——IPC的发展与展望[J].中国发明与专利,2009(8):76-79.
- [2] 曹平.中外专利文献信息检索平台比较研究[J].情报探索,2011(7):67-69.
- [3] 张帆.Internet免费专利数据库检索与利用[J].昆明理工大学学报(社会科学版),2009,9(8):104-108.
- [4] 李健康,许四洋,张政宝.九个常用中国专利检索网站比较研究[J].图书馆论坛,2010,30(6):191-200.
- [5] 张春华,王磊,王向红,等.中外专利数据库服务平台简介及检索应用[J].中国发明与专利,2012(1):61-63.
- [6] 胡晓,魏雪梅.我国网络专利检索平台分析和评价[J].科技管理研究,2010,30(14):75-77.
- [7] 马嘉,宋建玮.IPC分类法在科技查新工作中的应用[J].图书馆学报,2012(3):32-34.
- [8] 马海群.网络环境下的国际专利分类法IPC变革与发展[J].现代图书情报技术,2002(6):41-43.
- [9] 丁海燕.国际专利分类法与分类号[J].今日科技,1985(8):34-35.
- [10] 蒋昌生,陈志.国际专利分类法简介[J].赣图通讯,1985(4):17-20.
- [11] 王昌亚.《国际分类法》的概况及其发展[J].情报学刊,1988(3):23-26.
- [12] 黄迎燕.第8版国际专利分类表的特点[J].中国发明与专利,2007(5):26-27.
- [13] 贺杰,孙福新.技术创新与专利技术(第十一讲)专利的国际分类[J].工程机械,2003,34(11):56-58.

## 作者简介

宋群(1989-),男,硕士研究生,就读于中国科学技术信息研究所。E-mail: songqun101@sina.com  
吴广印(1965-),男,硕士生导师,中国科学技术信息研究所研究员。E-mail: wuguangyin@hotmail.com

## Research on Data Organization of Patent Based on IPC

Song Qun, Wu Guangyin / Institute of Scientific and Technological Information of China, Beijing, 100038

Abstract: As the most effective carrier of technological information with much practicality and reliability, patent is paid more and more attention in recent years. The article involves an effective way on data organization of patent. The purpose is to provide ordered retrieval results for users and to improve availability of patent information.

Keywords: Patent retrieval, Data organization, IPC

(收稿日期: 2012-08-06)