

●张静<sup>1,2</sup>, 刘细文<sup>2</sup>, 柯贤能<sup>1,2</sup>, 黎江<sup>1,2</sup>

(1. 中国科学院 研究生院, 北京 100049; 2. 中国科学院 国家科学图书馆, 北京 100080)

## 国内外专利分析工具功能比较研究\*

**摘要:** 简单介绍了专利分析的一般方法与流程, 然后从分析前数据准备功能、分析与展现功能两大方面对国内外12种专利分析工具的功能进行了比较研究。其次, 分别针对基本统计分析、引证分析以及聚类分析这3种主要实现方法进行了详细对比。最后对专利分析工具的选择与改进提出了建议。

**关键词:** 专利分析工具; 引证分析; 聚类分析; 统计分析; 比较研究

**Abstract:** The methods and procedures of patent analysis are briefly described, then based on the data preparation function before patent analysis as well as the data analysis and display function, twelve patent information analysis software tools at home and abroad are compared. The latter is focused on the basic statistical analysis, citation analysis and cluster analysis. Finally, the article offers some suggestions about the choice and improvement of patent information analysis software tools.

**Keywords:** patent information analysis software tool; citation analysis; cluster analysis; statistical analysis; comparative study

专利情报分析是指对来自专利说明书和专利公报中大量的、个别的、零碎的专利信息进行加工及组合, 并利用统计手段或技术分析方法使这些信息成为具有总揽全局及预测功能的竞争情报的一项分析工作<sup>[1]</sup>。也就是说, 专利情报分析是将零散的专利信息按照不同的指标或角度来进行整合分析, 以获取整体性的有价值的深层次情报信息。

随着社会的发展, 技术创新的步伐不断加快, 专利数据也随之快速增多。面对海量的专利数据, 要有效率地对其进行多角度分析, 就必须寻求方便快捷的专利分析工具。而专利分析工具的好坏将直接影响到专利分析结果的效率与准确性, 因此, 其在战略情报研究中的重要性也必然会越来越强。

### 1 专利分析方法概述

专利分析方法是以前文献计量学为基础, 借助其他学科的知识及有关工具而进行的。从研究人员的角度来看, 专利分析方法分为定性分析与定量分析两种; 从专利分析工具实现的角度来看, 大致可分为基本统计分析、引证分析和聚类分析。这些方法互相交叉融合, 统计分析与聚类分析对应了定量分析, 而引证分析则既有定性也有定量的成分。

#### 1.1 基本统计分析、引证分析和聚类分析

基本统计分析, 就是简单的定量分析, 主要是通过专

利文献的外表特征, 按照不同的角度来进行统计, 并从技术、经济等不同的角度对统计数据及其变化进行解释, 以取得技术动态发展趋势方面的情报。其分析结果通常以统计报告或图表的形式展现。

引证分析是指对目标专利的引用专利情况和目标专利被引用的情况进行分析, 以揭示相关专利之间的关系, 反映特定技术领域的生命周期, 以及竞争对手之间技术相互依赖关系。引证分析工具的好坏在于引证数据的来源和引证结果的呈现。目前引证数据的来源主要有美国(US)、德国(DE)、欧洲(EP)、英国(GB)和世界知识产权组织(WO)的专利。引证结果的呈现主要有引证表、引证树和引证地图。

聚类分析主要是专利分析系统利用独特的聚类技术将同一专利数据集中的数据按照技术分类聚成不同的子类, 以揭示该特定技术领域内各个子领域的分布情况, 分析各主要竞争对手在各子领域内的专利分布情况。目前专利聚类主要是按主题进行聚类, 呈现结果也可以按竞争对手和时间顺序进行浏览。聚类分析的结果展示方式主要有聚类地图、有结构数据聚类和无结构数据聚类<sup>[2]</sup>。

需要注意的是, 引证分析与聚类分析也包含了部分统计分析的思想在内。

#### 1.2 专利分析一般流程

从情报工作者的角度来说, 专利信息分析流程通常分为3个阶段: 准备期、分析期和应用期。准备期的主要工

\* 本文受中国科学院国家科学图书馆研究生科研项目资助。

作包括建立专利信息分析队伍、确定分析目标、研究背景资料、选定分析工具以及选择专利信息源等。准备期是保证专利信息分析达到目标的基础,而分析期是专利信息分析工作的主体,通常我们把分析期分为数据采集和数据分析两个阶段。应用期的主要工作包括对分析报告进行评估、制定相应的专利战略以及专利战略的实施等<sup>[3]</sup>。

专利分析工具的主要作用就在于要在分析期提供准确的数据并进行科学的分析,同时要为应用期报告的撰写提供可视化的分析结果展示。

因此,专利分析工具需要支持分析前的数据准备工作,支持多种统计分析维度和对专利分析的指标,并能将分析结果以直观的形式展现出来,当然也要能方便地导出用户所需的详细专利信息。

## 2 国内外专利分析工具数据准备功能的比较

笔者对国内外的 12 种专利分析工具进行了调研,包括 Aureka<sup>[4]</sup>, Delphion Citation Link<sup>[5]</sup>, Delphion Text Clustering<sup>[6]</sup>, Delphion PatentLab-II<sup>[7]</sup>, Matheo Patent<sup>[8]</sup>, Thomson Data Analyzer<sup>[9]</sup>, 北京东方灵盾科技有限公司与 M-CAM 合作的 East Linden Doors 工具<sup>[10]</sup>, 保定大为<sup>[11]</sup>, 知识产权出版社<sup>[12]</sup>, 北京彼速<sup>[13]</sup>, 中国台湾地区连颖科技的 Patentguider 2.0 试用版<sup>[14]</sup> 和恒和顿的 HIT\_恒库<sup>[15]</sup>。其中,前 6 种工具为国外专利分析工具,East Linden Doors 为中外合作开发的,后 5 种为国内专利分析工具。

在进行专利分析前,有必要对专利数据进行一定的处理,以保证分析结果的准确性。目前主要的数据处理包括同一专利权人的不同名称合并、同族专利的合并;另有部分分析工具支持自有数据的导入。

### 2.1 同一专利权人不同名称的合并

Delphion 的 Corporate Tree 可查找同一专利权人的不同表述并生成统一表达式,并在检索时直接使用该统一表达式进行检索。Thomson Data Analyzer 支持专利权人合并,有自动合并和手动合并两种方式。PatentLab-II、知识产权出版社和北京彼速支持专利权人的手动合并。

### 2.2 同族专利合并

Aureka 支持同族专利合并和展开,并可选择合并后显示 US, EP 或 PCT 的专利号。Matheo Patent 支持同族专利合并和展开。

### 2.3 自有数据的导入

Delphion 支持导入自己的一组专利号,进行统计分析、聚类分析。Thomson Data Analyzer 可导入各种内部和外部商业数据库的数据。PatentLab-II 可导入 Delphion 数据格式。北京彼速可导入自有数据,导入格式是 Excel。

## 3 各专利分析工具分析与展现功能的比较

专利分析工具提供分析的实现方法主要可以分为基本统计分析、引证分析与聚类分析 3 大类。对每种分析方法,其分析维度、结果展现方式均各有不同。

### 3.1 基本统计分析

基本统计分析是最基础的专利分析方法,因此基本上所有专利分析系统都具备此项功能,但不同的专利分析系统在分析结果的呈现和专利分析的维度这两个方面往往不一样。

据调查,统计分析结果呈现主要有 4 类工具:报表类(Reports)、图表类(Charts)、矩阵类(Matrices)、网络类(Networks)。统计分析的维度可分为 7 大类,分别是一般统计、专利权人、发明人、IPC 专利分类号、存活期分析、区域分析和引证统计,其中引证统计将放在后面的引证分析和聚类分析中讨论。这里仅讨论前 6 种基本统计分析维度,而这 6 类分析维度下又各有更细的划分。

笔者调研的工具中,提供基本统计分析功能的有:Aureka, Delphion PatentLab-II, Matheo Patent, Thomson

表 1 基本统计分析工具功能比较

分析工具	统计维度	结果呈现方式
Aureka	专利权人、发明人、专利分类号(主要、全部)、一般统计	以报表为主,每一类又分为基本报表和标准报表,共 25 个报表
Delphion PatentLab-II	专利权人、专利分类号和国家统计、一般统计	每一类分别提供报表;支持 10 种图表类型
Matheo Patent	发明人、申请人、IP 分类号(4 位、全部)、欧洲分类号、公布年份、自定义组和同族专利	对称矩阵分析与非对称矩阵;对称网络和非对称网络
Thomson Data Analyzer	发明人、组织、国家、申请时间、技术内容	直方图等统计图表;对称矩阵
保定大为	概况、申请国家、申请人、发明人、存活期、IPC;支持自定义分类分析自定义分类申请量	各种统计图表和报表;矩阵图
知识产权出版社	趋势分析、区域分析、申请人和国内分析;申请人综合比较	统计图表和报表
北京彼速	趋势分析、地域/国别分析、公司分析、专利权人、发明人和 IPC;专利权人综合比较	自定义统计图表;以蛛网图形式展示专利权人综合比较结果
Patentguider 2.0	发明人、国别、IPC、UPC、公司分析;专利件数分析	折线图的分析图表
HIT_恒库	分为授权信息、技术信息、时间信息、同族和组 4 个部分;用户自定义的组统计	统计图表;自动生成各种统计报告

Data Analyzer, 保定大为, 知识产权出版社, 北京彼速, Patentguider 2.0 试用版以及 HIT\_ 恒库这 9 种。不同工具的统计分析维度及结果呈现方式有很大的不同 (见表 1)。

其中, 保定大为主要提供自定义各分类趋势分析、各分类区域分析、各分类申请人分析、各分类发明人分析等。

知识产权出版社的专利分析工具提供申请人综合比较, 以表格的形式揭示申请人在 (专利数 + 申请活动年期 + 公开活动年期 + 所属国数 + 发明人数) 5 个方面的综合情况。

北京彼速的经纬线工具提供专利权人综合比较, 与知识产权出版社类似, 但以蛛网图形式揭示专利权人 (专利数 + 发明人数 + 国家 + 申请活动年期 + 公开活动年期) 的综合情况, 其中, 国家是专利权人的申请国家数; 申请活动年期是指在一个给定的时间区间内, 专利权人有申请专利的年数。

Patentguider 2.0 的专利件数分析包括技术生命周期统计和历年专利件数比较两个方面。

HIT\_ 恒库的 4 个部分的统计角度包括: 同族、优先权年、IPC4、IPC、欧洲分类、发明人、申请人、公告年、专利审查员、助理审查员、代理机构 (人)、美国分类、基本发明信息、扩展发明信息基本附加信息等统计。

### 3.2 引证分析

上述工具中, 提供引证分析的有 7 种: Aureka, Delphion Citation Lin, East Linden Doors, 保定大为, 知识产权出版社的引证分析工具, Patentguider 2.0 试用版以及 HIT\_ 恒库。其中, Aureka 主要通过 Hyperbolic Trees 工具描述特定专利的引用和被引的情况; East Linden Doors 通过麦哲伦模块进行引证分析, 其包含 3 个分析系统: 望远镜 (Telescope)、沙漏 (Hourglass) 和人造卫星 (Satellite)。

这几种分析工具都能将引证分析结果可视化显示出来, 但在分析数据来源、引证结果显示方式、专利信息呈现与统计等方面存在较大的不同 (见表 2)。

值得注意的是, East Linden Doors 的望远镜系统除显示专利的相关专利号、名称等基本信息外, 还会显示引证率: Pr/Su, Pr 表示此专利引证过的先有专利数量, Su 表示此专利被其他专利引证的次数。这个引证率可以成为诠释此专利对先有专利的依赖程度和对后有专利的重要性的影响程度的重要指示器。如果一项专利拥有很高的 Pr 数量, 这将暗示着此项专利相对于其所引证的先有专利可能仅仅是进行技术累加而呈现的不同, 而自身技术创新的成分可能比较小。这种情况经常显现在专利技术趋向于饱和的领域。

表 2 引证分析工具功能比较

分析工具	引证数据来源	引证结果可视化显示	专利信息呈现与统计	
Aureka Hyperbolic Trees	US, DE, EP, GB 和 WO	可选择多级引证, 最多可显示 5 级; 引证树中目标专利可同时显示多个字段; 可自动或手动给不同专利标注不同颜色	可对检索结果集的后继引证专利进行统计	
Delphion Citation Link	USPTO	引证关系可显示 1 级、2 级、3 级以及所有引证关系; 目标专利只能显示手动选择的一个字段; 能手工对不同专利标注颜色	可按年份查看引证关系; 点击节点查看专利详细信息	
East Linden Doors 的麦哲伦模块	望远镜系统	US	目标专利的引证及被引证的专利数量通过代表各专利的图标的大小变化和图标上的数字直观展现	鼠标置于专利图标可显示该专利相应信息
	沙漏系统	US	通过横向时间条来表述目标专利及与其有引证关系专利从申请到公告的历史; 每年公告的所有与目标专利相关的专利都会显示在纵向的彩色条形图表中	先有技术与现有技术以树状结构列表呈现
	人造卫星系统	US	通过峰状投射图形来展现目标专利创新范围内的所有相关专利; 不同色峰代表不同统计角度的相关专利	点击山峰, 会出现该范围内的专利信息列表
保定大为	US	按专利的引证数量和相互引证关系生成引证图; 只有一级引证关系	可统计申请人专利被引证次数; 可统计申请人自我引证专利数; 可统计申请人引证的非专利文献	
知识产权出版社	US	引证树可分引证方向显示、被引证方向显示和整体显示 3 种情形; 可自动按不同颜色进行标引	引证分析图可导出	
Patentguider 2.0 试用版	US, EP, WIPO, Delphion, TWPAT, WebPat	雷达分析可形成专利引证族谱图; 引证相关数据表; 专利引证次数图	可通过 PatentGrabber 工具下载 USPTO 和 EPO 专利全文	
HIT_恒库	US 和 EP	单条专利的引证分析, 可快速建立该专利的前向及后向引证树, 以箭头方向示意引证关系	引证树可保存为位图	

3.3 聚类分析

提供聚类分析的工具较少, 只有 Aureka, Delphion Text Clusterin 和 Thomson Data Analyzer。其中, Aureka 的聚类分析功能是通过 Vivisimo 和 ThemeScape 这两个工具实现的。这三者在数据来源、聚类角度与结果呈现方式上有较大区别 (见表 3)。

表 3 聚类分析工具功能比较

分析工具	数据来源	聚类角度	结果呈现方式
Aureka	Vivisimo	自身包含的专利数据库	专利标题、文摘 将所选专利自动聚成若干个 Cluster 和子 Cluster, 呈多级显示; 一次最多聚类分析 1000 个专利文档
	ThemeScape	自身包含的专利数据库	按主题聚类; 按标题和文摘、权利要求书、全文的内容进行聚类, 每类还可进一步选择更多字段 以主题地图的形式可视化展示; 可在主题地图上将不同的公司标注成不同颜色; 可用不同颜色标注不同的年份; 可进一步浏览地图上特定区域的单个或分组专利文档, 或导出地图上特定区域的文档
Delphion Text Clusterin	自身包含的专利数据库; 可导入 Derwent 专利数据	专利标题、文摘	可选择要聚类的 Cluster 数量; 每个 Cluster 都会用一些 Descriptive Words 来描述
Thomson Data Analyzer	可导入多个其他数据库的数据, 也可导入用户自己的数据	专利标题、分类	可导出为 Excel 或其他工具中; 导出结果能以 ThemeScape 地图的形式显示

从表 3 可以看出, Vivisimo 侧重检索结果管理的聚类分析, 而 ThemeScape 更侧重于检索结果的深度分析。使用 ThemeScape 将不同公司以颜色标注, 可以查看自己的专利组合或比较自己与竞争对手的专利布局侧重点; 而年份的标注, 则可以看出技术的生命周期。Delphion Text Clusterin 可以导入 Derwent 专利数据分析, 但是它的实际聚类效果并不太理想。Thomson Data Analyzer 的分析数据来源种类和格式很多, 可以导入 Derwent, INSPEC, Excel 等的数据库, 而其聚类结果则能以 ThemeScape 地图的形式来显示。

4 国内外专利分析工具总体比较

根据前文的分析比较, 12 种专利分析工具的比较结

果见表 4。

表 4 12 种专利分析工具的比较结果

专利分析工具	分析前数据准备				引证分析	聚类分析	统计分析	特点
	专利权人不同名称的合并	同族专利合并	自有专利数据导入					
Aureka		支持			支持	支持	支持	功能与数据很全的优秀的专利分析工具
Delphion Citation Link					支持			使用方便的引证分析工具
Delphion Text Clustering						支持		有较大自主性的聚类分析工具
Delphion PatentLab- II	支持		支持				支持	统计结果显示多样
Matheo Patent		支持					支持	统计分析全面且深入
Thomson Data Analyzer	支持		支持		支持	支持		支持多个数据库的数据整合
East Linden Doors					支持			多角度全面的引证分析
保定大为					支持		支持	中文界面, 统计功能灵活
知识产权出版社	支持				支持		支持	功能比较全面, 统计分析具有独特性
北京彼速	支持		支持				支持	中文界面, 统计功能全面
Patentguider 2.0 试用版					支持		支持	包含中国台湾地区的专利数据库, 雷达分析便于侵权分析
HIT_恒库					支持		支持	专利数据动态更新, 具有翻译功能

从表 4 可以看出:

1) 由于 Aureka 集成了包括 Hyperbolic Trees, Vivisimo 和 ThemeScape 这 3 种工具, 因此它是这 12 种工具中功能最为全面的分析软件。

2) 而 Delphion 在专利分析中也具有很强的优势, 它有多种不同功能侧重的专利分析工具, Delphion Citation Link 是专门的引证分析工具, Text Clustering 是专门的聚类分析工具, 而 PatentLab-II 则侧重统计分析, 如果能将这些工具集成并整合起来, 必定是一个优秀而全面的分析工具。

3) Thomson Data Analyzer 具有很强大的整合功能, 它支持多种格式数据的导入导出, 且具备较好的数据清洗功能。

4) 北京东方灵盾科技有限公司与 M-CAM 合作的 East

Linden Doors 具有很强的引证分析功能,但 M-CAM 本身还没有将六分仪等工具整合进来。

5) 保定大为、HIT\_ 恒库虽然支持引证分析,但由于引证级数过低,并不太具有实用性。

总体而言,国外专利分析工具比国内专利分析工具更加灵活,使用细节上要更加方便,这主要体现在数据的导入导出以及数据的显示与标示上。在统计分析功能上,国内外的各个工具都进行了很大程度的开发,统计维度与显示方式种类很多。但在引证分析与聚类分析上,国内工具与国外工具就具有了较大的差距。

从 3.2 节与表 4 可以明显看出,国外专利分析工具支持的引证级数普遍高于国内具有引证分析功能的专利分析工具;而调研的这 12 种工具中,尚没有能支持聚类分析的国内工具。

## 5 现有专利分析工具的局限与展望

从以上分析可以看出,目前的专利分析工具在以下 4 个方面还存在着较大的不足,在其发展过程中应予以考虑改进。

1) 分析数据来源单一。虽然多数工具的检索数据来源很全面,但很多分析工具仅能针对美国专利进行分析,这在引证分析功能中表现得尤为明显,这对专利分析结果的全面性具有很大的影响。分析数据来源单一同样也体现在对自有数据导入数据的功能上,多数分析工具不提供自有数据导入功能。

2) 分析前数据处理功能有待加强。在进行专利分析前,将检索结果按照相应的标准进行去重、整合是非常必要的,但目前提供这项功能的分析工具却很少。因此,专利分析工具应增加这项功能,同时也要增强此项功能的灵活性,使分析人员能够从不同角度进行数据处理。

3) 聚类分析功能还有待挖掘。提供聚类分析功能的工具很少,且聚类结果并不十分理想。可以肯定,聚类分析对技术竞争分析、专利布局等方面会起到越来越重要的作用,因此,文本挖掘、自然语言处理等有利于加强聚类深度与准确性的方法,以及 XML 等有助于增强聚类灵活性的方法都应更多的应用于专利分析工具中。

4) 专利分析工具与分析指标的不匹配度较高。专利分析的准确程度在很大程度上依赖于专利指标的选取与计算。有些技术指标的实现可以通过基本统计方法来实现,如专利数量、同族专利指数、专利成长率等;有些是通过引证分析来实现的,如引证指数、即时影响指数、科学关联性、技术生命周期等;有些要通过聚类分析来实现,如相对专利产出率、技术重心指数等。

现有的专利分析工具一般都支持专利数量、同族专利

指数等,能够通过基本统计方法得出结果的分析指标;具有引证分析功能的工具都能支持技术生命周期指标的直接统计。但其他需要经过再次计算的分析指标,如相对专利产出率、科学关联性则需要用户迂回进行计算,使用起来并不方便。因此,专利分析工具的功能设计应该尽可能地向着既定分析指标靠拢,以方便用户使用。

可以预见,专利分析工具的功能会越来越完善,使用会越来越灵活,分析范围会越来越广。这就要求专利分析工具提供商要在扩展工具功能的同时,增强工具的可扩展性与易用性。

需要注意的是,这里的易用性不仅仅是指工具的易操作性,更重要的是分析功能同分析指标的高匹配性带来的易用性。□

### 参考文献

- [1] 唐炜,刘细文. 专利分析法及其在企业竞争对手分析中的应用 [J]. 现代情报, 2005 (9): 179
- [2] 刘佳佳,董曼,方曙. 国外专利分析工具的比较研究 [J]. 现代图书情报技术, 2007 (2): 67-74
- [3] 陈燕,方建国. 专利信息分析方法与流程 [EB/OL]. [2007-05-10]. <http://www.zydg.net/magazine/article/1672-6081/2005/12/298578.html>
- [4] <http://scientific.thomson.com/products/aureka/>
- [5] <http://www.delphion.com/products/research/products-citelink>
- [6] <http://www.delphion.com/products/research/products-cluster>
- [7] <http://www.delphion.com/products/research/products-patlab>
- [8] <http://www.matheo-patent.com/>
- [9] <http://scientific.thomson.com/products/tda/>
- [10] [http://www.eastlinden.com.cn/Product/Product\\_Lindendoor.asp](http://www.eastlinden.com.cn/Product/Product_Lindendoor.asp)
- [11] <http://www.daweisoft.com/chproduct/chuser/view.asp?proid=118>
- [12] [http://www.cnipr.com/cp/w/zlxcp/t20050822\\_52639.htm](http://www.cnipr.com/cp/w/zlxcp/t20050822_52639.htm)
- [13] <http://www.bizsolution.com.cn/SoftIntroduce/Graticle.aspx>
- [14] <http://www.learningtech.com.tw/products/pg/pg.aspx>
- [15] <http://www.all-patent.com/>

作者简介:张静,女,硕士生。

刘细文,男,研究馆员,硕士生导师。

柯贤能,男,硕士生。

黎江,男,硕士生。

收稿日期:2007-08-03