

# 专利检索中的关键词扩展研究

覃冬梅 丁君军

国家知识产权局专利局专利审查协作北京中心

## 摘要

关键词检索是专利检索时的主要方式,关键词属于人类自然语言,其表达方式多种多样。本文从关键词扩展的角度出发,论述了关键词扩展的必要性及常见关键词扩展方式,通过实际检索案例论证关键词扩展的有效性,在一定程度上减少漏检,从而能更快、更准地检索出最相关的对比文件。

## 关键词

专利检索;关键词;扩展;现有技术

## 引言

在专利审查过程中,检索是发明专利申请实质审查程序中的一个关键步骤,其目的在于找出与申请的主题相关的现有技术,以确定申请的主题是否具备专利法所规定的新颖性和创造性<sup>[1]</sup>。

在检索过程中,通常采用关键词和分类号组合方式进行检索,而以关键词作为检索手段更是首选检索方式。关键词属于人类自然语言,表示其相近含义的方式多种多样,例如计算机术语“存储器”,在中文专利文献中还可表示为“储器”、“存储设备”、“存储单元”、“存储装置”等,表示具体的存储器形式还可分为“内存”、“外存”、“缓存”、“贮存器”、“寄存器”等,或者用英文缩写“ROM”、“RAM”、“DRAM”等表示,此外,在台湾申请人的专利申请文献中,“存储器”还可能被称为“记忆体”、“记忆棒”等<sup>[2]</sup>。因此,不同背景、地区的人对同一特征的技术术语表达方式可能截然不同,如果不进行必要的关键词扩展,势必会导致检索结果不全面,造成现有技术漏检等问题。

## 1 关键词扩展概述

对关键词进行扩展首先得选取基本关键词,对检索主题进行分析,以权利要求为基础,结合说明书背景技术和具体实施例的内容,重点关注发明申请中的发明点,准确提炼出表示发明申请主题和发明点的词语作为基本关键词。提炼出基本关键词后,对关键词进行筛选、补充、扩展。

专利检索中,对关键词进行适当扩展是检索时的惯用手段。关键词的扩展可以借用现有工具,如搜索引擎的关键词工具、相关搜索、搜索提示或者专利局专利检索与服务系统(S系统)的关联词查询工具,也可以利用本领域词典、标准和技术术语库等工具书进行扩展,从关键词的同近义词表达形式、上下位概念、音形相近词、外来词及其相关概念等角度进行扩展<sup>[3]</sup>。例如触摸屏领域中的关键词“触摸”,对其进行同近义词扩展,可对“触摸”扩展关键词为:触控、触屏、触碰、接触、触感、触点、触点、触控、触控等。又如对软件进行补丁中的“补丁”一词,扩展关键词“升级”、“更新”能表达与“补丁”相近的含义。

关键词的扩展还有很多角度,可以从说明书中解决的技术问题、达到的技术效果入手,例如一种改变角度进行焊接的发明,

在针对具体角度检索未果时,可以尝试从改变焊接角度达到防止碎片飞溅的技术效果入手,利用扩展关键词“飞溅”在全文库中进行检索;还可以对关键词的同在词进行扩展,例如根据触摸屏传感特征,关键词“触摸”还可扩展为传感、感应、电感、多点、多手势、电极、扫描等,上述词语一般是触摸屏文献中的和“触摸”同在的词语。

关键词扩展没有统一的策略和标准,具体案例需要具体分析,扩展是建立在本领域技术人员对发明方案理解的基础上来实现的。检索时,关键词扩展也不是一步到位,检索过程中需根据检索结果不断地调整检索策略,以提高检索查全率。此外,关键词扩展也应准确把握尺度,盲目的扩展会带来检索噪声,影响检索的查准率。合理、有效的关键词扩展,能更快、更准地检索到现有技术。如以下案例,就是通过对关键词的恰当扩展,较好地检索到一篇能够影响权利要求创造性的对比文件。

## 2 发明申请检索案例

### 2.1 案例概况

申请号:200910085036.3,申请日:2009年5月27日,发明名称:一种数据过滤与合并的方法。该申请修改后的权利要求1为:

一种数据过滤与合并的方法,包括:按照数据的类型将数据的结构信息、过滤规则和合并规则存储在对应的配置文件中;

接收到数据后,根据所接收的数据的类型读取对应的配置文件,根据所读取的配置文件中的结构信息和过滤规则滤除或保留该数据;若保留该数据,则根据所读取的配置文件中的结构信息和合并规则将该数据与相应数据进行合并;所述数据的结构信息包括:数据的各字段的偏移量。

在现有技术中,实现数据过滤与合并时,由于不同数据类型具有不同的结构,在每增加一种数据类型时,都需要在代码中重新编码写入该数据类型的结构信息,不利于系统的扩展。

该申请采用配置文件存储各种不同类型数据的结构信息、过滤规则和合并规则,其发明点为:除了将数据的规律和合并规则存储在配置文件中,还将数据的结构信息存储在配置文件中。该申请的方案不需要修改系统代码,只需要将新增的数据类型的结构信息增加到配置文件中即可进行数据的过滤与合并,增强了系统的可扩展性。

### 2.2 检索策略

#### 1)、提炼基本关键词

发明专利权的保护范围以其权利要求内容为准,通过对权利要求1进行分析,并根据说明书记载和申请人强调的发明点“数据的结构信息存储在配置文件中”,将检索的重点选定到该发明点上,提炼出基本关键词“过滤”、“合并”、“配置文件”、“结构信息”等。

#### 2)、扩展关键词

对关键词进行扩展,根据本领域的公

知常识,配置文件一般是采用XML或INI的格式进行存储,而XML是本领域通用的层次性数据交换配置文件,具有跨平台、可移植性好等优势,因此,可将关键词“配置文件”扩展为关键词“XML”、“INI”。在提炼出本申请的检索要素后,构造体现技术方案发明构思的检索要素表,如表1所示:

OR	基本关键词	过滤、合并	合并、过滤、规则、数据	AND	配置文件	AND	数据结构
分类号	IPC	H04L29/06、H04L9/36	Filter*, script*, rule*, data	AND	Configuration file*, XML-INI	AND	Data structure-, describe-, script-, type-, field-, offset

在中国专利检索系统文摘数据库(CPRSABS)和中国全文文本库(CNXTXT)中进行检索。

CNXTXT? ..HI  
编号 所属数据库 命中记录数 检索式

75 CNXTXT 23 (((合并 or 过滤) 8d (数据 or 规则)) S (配置文件 or XML or INI) S (数据结构 or 数据描述 or 脚本 or 类型 or 字段 or 偏移量)) AND PD<20090527

通过上面的检索式,检索到一篇能评价本申请创造性的X类对比文件1,公开号:CN101355774A,公开日2009年1月28日。

### 2.3 检索结果分析

对比文件1公开了如下内容:采用XML配置文件对源数据进行描述,该源数据描述指定如何过滤数据;将各业务数据结构和过滤条件存储在XML配置文件中;定义根据不同参数值以对数据进行处理的内容。因此对比文件1公开了本申请的发明点:将数据的结构信息和过滤条件一起存储到XML配置文件中,针对不同类型的数据进行分别处理。

而两者的区别仅在于,配置文件中还包含合并规则,并利用该合并规则合并数据。在对比文件1中已公开了将过滤规则存储于配置文件中的基础上,本领域技术人员很容易实现将合并规则也存储于配置文件中,因此权利要求1不具备创造性。

## 3 总结

在对上述案例进行检索时,通过将“配置文件”扩展为本领域常用的配置文件格式“XML”、“INI”,很快的检索到一篇已公开该申请发明点的对比文件。因此,对关键词进行有效扩展,可以使检索事半功倍,从而准确、高效地完成检索任务,提高检索到XY类对比文件的成功率。

随着语义技术的发展,现有的语义系统已能自动对关键词进行扩展,只是目前机器扩展关键词的效果还远不如人工扩展,有待今后进一步的探讨和研究<sup>[2]</sup>。

## 参考文献

[1]中华人民共和国国家知识产权局.专利审查指南2010[M].北京:知识产权出版社,2010

[2]周胜生.专利文献的特点与关键词检索[J].PATENT EXAMINATION REVIEW,2010,2:59~63

[3]陈予琳.关键词检索方法在科技查新中的应用研究[J].河南师范大学学报(自然科学版),2011,3:171~173