

非专利文献技术内容分析及其深加工策略研究

专利检索咨询中心 葛亮 马照英 孙夏 李安玲 关楠楠 郭兰



摘要:发明专利申请在被授予专利权前应当进行检索。检索的目的在于找出与申请的主题密切相关或者相关的现有技术中的对比文件。引用对比文件判断发明或者实用新型的新颖性和创造性等时,应当以对比文件公开的技术内容为准。本文研究了非专利文献技术内容在摘要和全文中的分布情况,提出在非专利文献检索系统中设置技术内容这样一个全新字段,以满足专利审查员的检索需求,提高审查效率。最后,研究了非专利文献技术内容字段的数据深加工策略。

关键词:非专利文献 技术内容 数据加工 专利审查
摘要检索 全文检索

引言

非专利文献中包含了大量的实用技术信息,在专利审查中所起的作用越来越重要^[1]。近年来,我国的电学、

通信领域的专利申请量非常大,相关领域内的非专利文献在相应的审查领域内的引用频次也逐年增加,专利审查工作中对非专利文献的需求量不断增长。但因非专利文献内容庞杂,在专利审查检索过程中还存在许多亟待解决的问题。

我们知道,每一件发明专利申请在被授予专利权前都应当进行检索。检索的目的在于找出与申请的主题密切相关或者相关的现有技术中的对比文件,以确定申请的主题是否具备专利法第二十二条第二款和第三款规定的新颖性和创造性。对比文件是客观存在的技术资料。引用对比文件判断发明或者实用新型的新颖性和创造性等时,应当以对比文件公开的技术内容为准。该技术内容不仅包括明确记载在对比文件中的内容,而且包括对于所属技术领域的技术人员来说,隐含的且可直接地、毫无疑义地确定的技术内容^[2]。在CNKI非专利数据库中,审查员可选择的与技术内容相关的检索项包括:题名、主题、摘要、关键词和全文。其中,摘要检索字段和全文检索字段反映的技术内容最多。因此,非专利文献的摘要和全文中的技术内容对于专利检索的查全率和查准率影响很大。基于此,我们对非专利文献的摘要和全文中的技术内容进行了统计和分析。

一、非专利文献技术内容统计方法

为了解我国非专利文献摘要是否能够全面体现技术内容,满足专利审查检索需求。我们对电学、通信领域的电子学报、半导体学报、光学学报、通信学报等50种期刊进行了统计。统计步骤如下:

(1)每一种期刊选择2000-2010年的其中一期进行统计,例如选择《电子元器件应用》2008年第1期;

(2)从所选期刊筛选出有技术内容的文献;

(3)筛选文献的统计项目包括摘要字数、摘要技术内容字数、全文字数、全文技术内容字数;

(4)对筛选的所有文献进行一个总的统计,得到一个非专利文献技术内容的整体分布状况。

以《通信技术》2000年第1期为例,列出非专利文献技术内容调查表如表1所示。

按上述统计方法,得到非专利文献技术内容统计结果如表2所示,涉及所选的50本期刊1512篇非专利文献中896篇有技术内容的文献。

表 1 非专利文献技术内容调查表

序号	标题	摘要 (字)	摘要技 术内容 (字)	全文 (字)	全文技 术内容 (字)	摘要技 术内容 / 全文技 术内容	全文技 术内容 / 全文	摘要技 术内容 / 全文	摘要是否反映 全文技术内容
1	DS/CDMA 移动通信系统的同步技术	111	86	3466	508	16.93%	14.66%	2.48%	否
2	平流层信息平台有效载荷技术研究	114	0	3797	550	0.00%	14.49%	0.00%	否
3	有限精度混沌映射跳频多址序列	125	0	1440	229	0.00%	15.90%	0.00%	否
4	块交错方案在移动图像通信系统中的应用	150	113	2927	570	19.82%	19.47%	3.86%	否
5	高清晰度电视信号串行传输系统的研究	63	0	1631	565	0.00%	34.64%	0.00%	否
6	多媒体控制协议 H.245 的实现	32	0	2517	232	0.00%	9.22%	0.00%	否
7	ISDN 网络终端 NT2 的设计与实现	114	114	3634	330	34.55%	9.08%	3.14%	否
8	GPS/DGPS 仿真中并发延时算法的设计	136	0	2578	188	0.00%	7.29%	0.00%	否
9	合计	845	313	21990	3172	9.87%	14.42%	1.42%	否

备注：《通信技术》2000 年第 1 期的文献总量为 22 篇，其中有技术内容文献的为 8 篇。

表 2 非专利文献技术内容统计结果

统计项目	摘要字数	摘要技术内容字数	全文字数	全文技术内容字数	摘要技术内容 / 全文技术内容	全文技术内容 / 全文	摘要技术内容 / 全文
合计	129608	79428	2751653	484412	16.40%	17.60%	4.71%

二、非专利文献技术内容统计结果讨论

1. 非专利文献包含技术内容的文献量

针对电学、通信领域非专利 50 本期刊的技术内容调查发现：涉及的 50 本期刊共有 1512 篇文献，除去无技术方案的非科技类文献如会议、产品介绍、广告、政策等，机理或理论研究类，技术方案不具体或技术方案提供不全的，筛选其中 896 篇文献包含技术内容，具备成为专利对

比文件的可能性，占调查全部文献的 59.26%。对于不同调查期刊包含技术内容的文献所占该期文献总数的比例也不同，例如无线电工程、电路与系统学报、通信学报、微计算机信息、电子器件包含技术内容的文献占 90% 以上，而电视技术、光子学报、中国集成电路、发电设备包含技术内容的文献占 40% 以下。

2. 非专利文献技术内容分布特点

根据表 2 非专利文献技术内容统计结果，得到非专利文献期刊技术内容分布情况如图 1 所示。

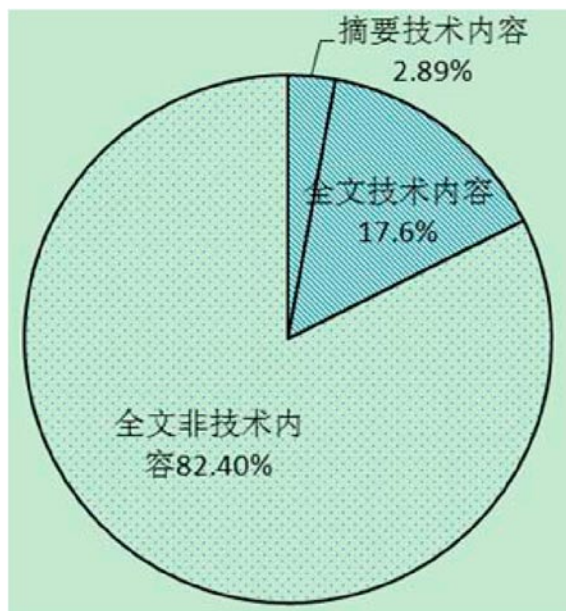


图 1 非专利文献期刊技术内容分布情况

经过数据统计, 17%的非专利文献是不提供摘要的, 在有摘要且包含技术内容的非专利文献中, 摘要包含的技术内容仅占全文内容的 2.89%, 全文的技术内容占全文内容的 17.6%, 摘要反映的技术内容是其应该反映技术内容的 16.4%。在统计的 896 篇非专利文献中, 875 篇无法充分反映文献的技术内容, 占 97.66%。上述数据表明, 非专利文献技术内容分布特点是摘要反映的技术内容很少, 而全文中多数文字与技术内容无关。非专利文献技术内容分布特点导致了以下几个方面的专利检索问题:

首先, 在非专利文献数据库中, 如果以全文作为检索入口, 速度慢, 查准率低。在非专利检索中, 利用全文检索, 虽然能够保证文献的查全率, 但系统需要处理大量数据, 速度慢是毋庸置疑的。另外, 数据统计发现, 非专利文献的技术内容在整篇文章中占的比重又很少, 审查员在全文检索时, 要浏览大量的“冗余”信息(全文非技术内容, 约占 82.4%), 费时费力, 降低了审查效率, 提高了审查成本。非专利文献引用起来效率不高是目前我局非专利文献引用率相比欧洲专利局较低的一个重要原因。全文检索时, 技术性文章与非技术性理论阐述文章等同时检出, 难以区别, 检索效率低, 全文检索将会检出

大量噪声数据, 虽然保证了文献的查全率, 却降低了检索效率。从审查角度而言, 完成一件检索的时间并不是无限期的, 需要满足效率优先的检索原则。因此, 全文检索无法实现检索质量和效率的平衡, 这对专利审查检索不是最优选择。

其次, 在非专利文献数据库中, 摘要反映的技术内容很少, 通过摘要这一重检索入口对技术内容进行检索不能满足目前的专利检索需求。摘要反映的技术内容较少, 直接会导致漏检, 影响专利检索的查全率, 导致专利审查质量下降, 授予专利权处于相对不稳定状态。审查员在非专利文献检索系统中, 输入检索要素进行检索时, 如果摘要字段未能提供或者提供不完整的技术内容, 检索效果会大打折扣。由于摘要反映的技术内容较少, 不利于审查效率提高, 主要体现在摘要的浏览功能上, 如一些技术细节包括参数、材料等信息不能直接通过摘要直接获得。审查员在浏览摘要后无法确定非专利文献是否可以作为对比文件被引证时, 需进一步下载全文, 在全文中查找所需的技术内容, 全文内容庞杂, 不利于技术内容的快速浏览。非专利文献摘要往往注重描述研究机理或研究结果的分析, 而不侧重描述文章中的技术方案。而且, 摘要中通常包含了许多与技术方案无关的

内容，这也会给专利检索带来噪声。

总之，非专利文献的全文检索保证了查全率，但其速度慢，噪声大，查准率低。摘要反映的技术内容少，且含有许多与技术内容无关的文字，既影响了查准率，也影响了查全率。检索过程既要满足查准率又要满足查全率的要求。速度和质量本身就是一对矛盾，因此，在满足上述要求的前提下，建设满足双方平衡点的非专利文献数据库检索字段，是亟待解决的问题。为了平衡非专利数据库的查全率和查准率，我们需要整合非专利文献包含的全部技术内容，构建一个服务于专利审查的全新字段——“技术内容”，使其具有非专利文献技术内容的检索功能和浏览功能。该字段的技术内容信息量远大于非专利文献摘要的信息量，保证查全率，且去除了全文中与技术内容无关的信息，减少噪声，方便浏览，提高审查效率。

三、技术内容字段的数据深加工策略研究

1. 技术内容字段与摘要的区别

技术内容字段的标引以提供技术信息为目的，主要是将非专利文献的技术领域、要解决的技术问题、技术方案以及获得的技术效果等信息整合在一起。与摘要的不同之处在于技术内容字段涵盖的技术主题及主要技术

特征，该字段去除一些与技术内容无关的文字，全部文字均为技术信息，其不受字数限制，提供更加完整的技术内容，包括一些技术细节，如参数、材料等。

2. 技术内容数据深加工策略研究

技术内容字段数据深加工的主要方式则是采取从全文中拷贝技术信息，使其具有检索功能，然后对拷贝的信息进行编辑修订，使其具有阅读浏览功能。下面对技术内容的深加工方法进行了研究。

首先，确定全文中的技术内容，拷贝技术信息，实现技术内容检索功能。技术内容包括：要解决的技术问题和有益效果、技术方案、用途或技术领域。要解决的技术问题是指所要解决的现有技术中存在的技术问题。有益效果是指由技术特征直接带来的，或者是由所述的技术特征必然产生的技术效果。有益效果可以由产率、质量、精度和效率的提高，能耗、原材料、工序的节省，加工、操作、控制、使用的简便，环境污染的治理或根治，以及有用性能的出现等方面反映出来，也可以通过列出实验数据的方式予以说明。技术方案是对要解决的技术问题所采取的技术改进，该内容能够使读者在最短的时间了解非专利文献技术内容的创新之处。用途是公开的技术方案在不同领域的实际应

用，技术领域是技术方案所属的具体技术领域。

拷贝技术信息，在内容上对应文献中的主要技术方案，体现文献主要技术创新之处。如文献主要技术方案涉及对设备、产品、方法或组合物的改进，技术内容中应体现相应的改进之处。如文献主要技术方案整体是全新的，则技术内容应从整体上体现该方案。如文献涉及产品（化合物、组合物及其他等），同时其制备方法或用途信息也应包含在技术内容字段中。如果文献中有多种可供选择的方案，在阐明优选方案的同时，还应拷贝其他方案。技术内容还应包含文献中提到的对检索有用的其他信息。

对拷贝的技术内容进行编辑修订，使其具有可读性，实现技术内容字段的阅读浏览功能。由于拷贝技术信息庞杂、分散，有些内容还会出现大量重复，要想实现阅读浏览功能需要对这些技术信息进行编辑修订。编辑修订要遵从客观原则，不得加入自己的主观见解。机械、电学、通信领域的非专利文献有许多涉及技术方案的附图，为了方便用户对附图的阅读和浏览，在技术内容字段还可加入相应的附图说明。

3. 技术内容字段数据深加工案例

下面通过两个典型案例进行分析，阐述电学、通信领域非专利文献

技术内容字段的数据深加工方法。

例 1：水平铅酸蓄电池

摘要：介绍了水平铅酸蓄电池的研究现状、组成和结构特点。初步研究表明：水平铅酸蓄电池具有高比能、高比功率、快速充电的特点，但寿命较短。寿命已成为水平铅酸蓄电池实用化的主要制约因素。报导了 40 Ah 水平铅酸蓄电池的试验结果，讨论了影响水平电池寿命的 3 个关键因素——铅布的耐腐蚀性、各单格蓄电池的一致性、蓄电池的失水问题，指出只有突破寿命难关，水平铅酸蓄电池才有可能从研制阶段转为实用阶段。

加工的技术内容字段：本文涉及一种水平铅酸蓄电池。水平铅酸蓄电池的铅布是由玻璃纤维外包覆金属铅（或铅合金）经过挤压拉丝、编织而成。组成铅布内芯的玻璃纤维不但增加了铅布的强度，同时，由于玻璃纤维的密度小，也减轻了铅布的质量。此外，由于铅布很薄、可制成薄极板，这一特点使水平铅酸蓄电池具有优良的大电流放电能力。正极、负极可制作于同一块铅布上（称作双极板）。为减少水平铅酸蓄电池的自放电、增强铅布抗腐蚀能力，需在电极的正极、负极双连体之间填涂耐酸、绝缘的隔离材料。水平铅酸蓄电池采用的是准双极性的电池结构设计。由于极群是由双极板之间的铅布连接起来的，省

略了普通阀控铅酸蓄电池的焊接工艺，减轻了半极柱、汇流排的质量，因此使其具有电阻小、比能高的特点，同时，由于水平放置，减轻了电解液的分层现象。

分析：摘要中只是笼统的介绍了其研究工作情况，没有提及实质的技术内容，通过摘要提供的技术信息只能得知文献介绍的是关于水平铅酸电池的，根本无法获取该文献解决技术问题的技术方案创新之处。这样的摘要不能实现审查员对该水平铅酸蓄电池的特征检索，很可能导致漏检。另外，阅读摘要后对文献技术方案不得而知，需要进一步浏览全文，全文中技术内容分散，不方便阅读浏览。加工后的技术内容字段将文献核心技术方案及其达到的效果进行充分说明，具备了检索功能和阅读浏览功能，方便审查员将专利申请文件与该文献公开的现有技术内容进行对比。

例 2：线性光电耦合隔离放大器

摘要：介绍一种利用运算放大器对非线性光耦进行补偿使之具有线性传输特性的隔离放大器，对其工作原理进行了分析。并给出了实测结果。

加工的技术内容字段：采用运放对光电耦合器补偿而构成的隔离放大器如图 2(参考文献 4 的附图 1)所示，两只同型号光电耦合器的发送端串联接至运算放大器 A_1 的输出端，电阻

R_4, R_5 组成的分压器为 A_1 提供一个偏置电压，光电耦合器 G_1, G_2 的发送端在无信号输入时也有一定的静态电流流过，以使光电耦合器脱离其工作点上的“死区”。光电耦合器 G_2 输出的电流 I_{g2} 在 R_7 上产生的电压降 V_{R7} 与分压器 R_9, R_{10} 产生的偏置电压 V_{R9} 之间形成的差动电压经 A_2 放大后输出 V_0 。这就是隔离放大器的输出电压。为了保证隔离放大器能传输从零值起始的模拟信号，可在其输出端引入电压偏置电路如图 3(参考文献 4 的附图 2)所示，可使在单电源电压下工作的隔离放大器仍具有传输从零电压起始的模拟信号的能力。采用运放对普通光电耦合器进行非线性补偿，使之构成具有良好的传输线性度的隔离放大器是一种简便实用的方法。

分析：摘要中没有对技术方案要点进行充分描述，没有给出该隔离放大器的电路结构和工作原理，而且对技术方案的改进点描述也不完整，没有涉及单电源工作时隔离放大器的输出偏置电路，以及对于产品改进带来的有益效果也没有提及，通过摘要不能了解产品的关键技术特征和产生的效果，加工后的数据将电路工作原理以及改进点进行了充分的说明，并简要说明了技术效果，结合摘要附图更便于审查员的浏览阅读。

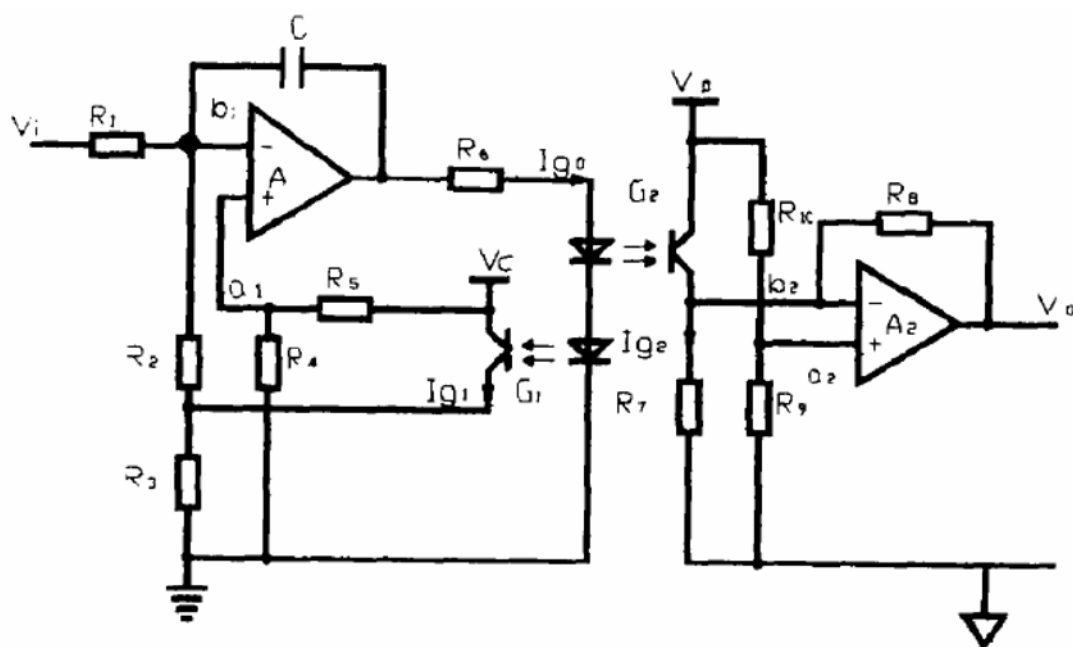


图2 线性光电耦合隔离放大器电原理图

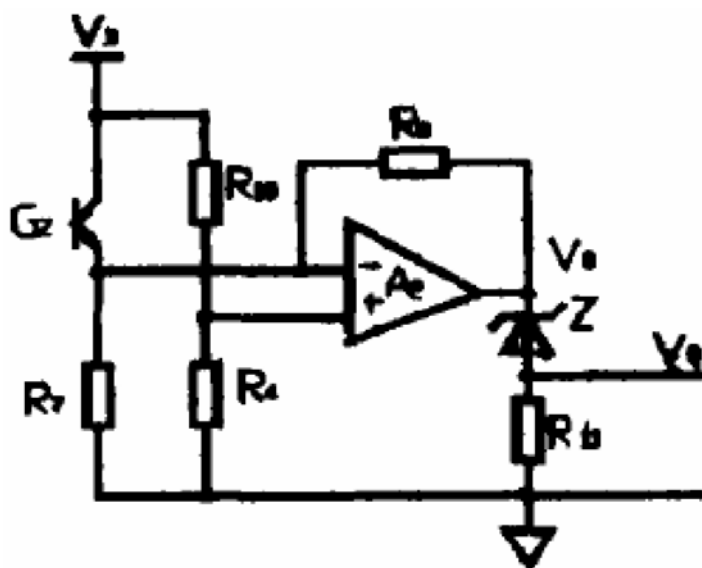


图3 单电源工作时隔离放大器的输出偏置

结语

在专利审查过程中，对比文件公开的技术内容是评判发明或实用新型专利申请的新颖性和创造性的重要证据。技术内容的加工可以平衡专利审

查对非专利文献检索查全率和查准率之间的矛盾，提高非专利文献在专利审查中的利用率，从而提高我国专利审查质量，加强我局授予发明专利权的稳定性。技术内容字段数据深加工

以拷贝加编辑的方式进行,方便快捷,加工成本低,产出效益高,可成为今后我局非专利文献数据深加工的一个新的方向。

(专利检索咨询中心 杜军 审校)

参考文献

1. 董林水、张秉斋、李敬辉、孙亮. “中文非专利对比文献在专利审查中的应用情况分析”. 《数据加工通讯》2008年第1期第20-27页。
2. 专利审查指南(2010版),第203-210页。
3. 许艳芳、司凤荣、钱志刚、彭元亭、郑克文. “水平铅酸蓄电池”《电池》2003年第1卷第33期第33-35页。
4. 刘大健. “线性光电耦合隔离放大器”. 《仪表技术》1997年第6期第18-20页。

