

NEWSLETTER

知产快报

● 人工智能浪潮下，很多工作将会被替代。智能语义技术已经在专利检索中大规模应用，未来将成为检索专家的得力助手还是终结者？作者基于其曾在专利审查工作中使用智能语义开展检索的丰富经验，通过梳理智能语义技术产生的原因、基本原理及其在专利检索中的应用方式，给出了一个答案。



智能语义技术在专利检索中的应用

人工智能浪潮下，很多工作将会被替代。智能语义技术已经在专利检索中大规模应用，未来将成为检索专家的得力助手还是终结者？作者基于其曾在专利审查工作中使用智能语义开展检索的丰富经验，通过梳理智能语义技术产生的原因、基本原理及其在专利检索中的应用方式，给出了一个答案。

专利检索经历了百余年的发展历史，早在 19 世纪末美国政府就在咨询工作中应用专利检索技术，当时是针对纸质专利文献的手工检索。从 20 世纪 70 年代起，基于计算机的专利检索逐步普及，专利检索的效率也因此大幅提升，但对于当时的公众而言，获取专利信息仍然是十分困难的事情。直到 1997 年 IBM 开始在互联网上提供专利信息服务，公众获取专利信息才变得相对容易¹。纵观专利检索的发展历史，伴随着信息技术的变革，专利信息获取和利用的方式也在不断发生变化。

进入 21 世纪以来，信息技术创新日益加快，人工智能技术的发展尤为迅速，网络购物时的推荐算法、扫脸支付、语音助理等新技术已经不知不觉融入到我们的日常生活中。在专利检索分析领域，被称作智能语义的新技术也进入到了重要的发展时期。2018 年 5 月，世界知识产权组织召开“知识产权局关于 ICT 战略和人工智能”会议，会议资料显示，多个国家和地区的知识产权局已经开始尝试在专利审查的检索环节应用智能语义技术²。在商用专利数据库领域，多款工具已经具备了智能语义检索功能。新技术的出现为专利检索工作带来了什么样的变化？作者基于其曾在专利审查工作中使用智能语义开展检索的丰富经验，简要

介绍智能语义技术产生的原因、基本原理及其在专利检索中的应用方式。

传统检索面临的困境

困境，不仅存在于专利检索中，而是在于所有的传统搜索领域。信息爆炸是困境的主要来源，随着人类知识以几何级数量增长，想要快速准确的搜索到所需信息并不是一件容易的事情。专利行业同样面临数据增长的问题，仅中国大陆公开的专利文献数量，发明和实用新型专利文献总量从 2008 年底只有不到 250 万件，已经跃升到 2018 年底的 1700 万余件，10 年的时间增长了 7 倍，这为开展专利检索工作带来巨大的挑战。

基于传统的布尔检索方式，检索专家们致力于在检索全面性和检索效率之间寻找平衡：任何一个技术概念都有诸多的表述方式，检索时只有尽可能全面的列举这些表述方式，才能保证不漏掉重要文献；但每一种表述方式又会有多重含义，全面的表述将引入大量不准确的噪声文件，这些噪声文件浪费了检索人员的浏览时间。如今，检索专家们面对着增长了数十倍的专利文献，感到越来越难在检索的全面性与检索效率之间找到平衡。

智能语义检索的兴起

为了解决传统检索面临的困境，出现了一

¹陈燕, 黄迎燕, 方建国. 专利信息采集与分析[M]. 清华大学出版社, 2006: 104-106.

² WIPO/IP/ITAI/GE/18, May 23 to May 25, 2018 (Geneva, Switzerland), https://www.wipo.int/meetings/en/details.jsp?meeting_id=4658

种自动扩展检索内容的语义技术，具体来说构建一个大型的同义词库，将用户输入的内容在同义词库中搜索，扩展相关关键词来提高检索的命中率。但是，这种语义技术虽然提高了检索的全面性，但应用该算法会在每次检索中命中更多的专利，同样引入了噪声文件，增加了检索人员浏览和筛选文献的工作量。

另一种智能语义技术是基于语义相关性的排序。这种技术基于专利文本数据开展机器学习，自动运算词语之间和文档之间的相关性，基于相关性对文献进行排序。这种方法的目标是将最相关技术排在最前，直接提升检索人员浏览文献的效率。应用该技术的专利检索工具，可以允许用户输入任意长度的一段文本或是直接输入一个专利公开号，系统将自动推荐最相关的文献。

智能语义技术在专利检索中的应用

对于基于同义词库的语义搜索技术，由于计算机仅仅是对用户输入的内容进行了扩展，在具体检索时仍然执行了布尔检索，因此这种语义检索工具和传统布尔检索工具的差别不大，在理解和使用上没有太大的困难。

而基于潜在语义索引的搜索技术，则完全脱离了传统布尔检索的范畴，与我们长期对检索的理解大不相同。这种搜索技术完全抛弃了从一个文献集中限定出子集的过程，而是按照用户所输入文本内容的相关性，直接对一个文献集进行排序。下面简要介绍应用这种技术开展专利检索的两种方法。

第一种方法是完全独立使用智能语义检索。此时检索人员只需将想要检索的一段文本输入语义检索系统，计算机就能根据用户输入的文本，对数据库中存储的千万乃至上亿条专利做排序，检索人员依次浏览就可能获得所需的专利。由于这种检索方式不需要任何的检索策略，

完全没有检索经验的人也能很快上手。对于检索经验丰富的人，在检索初期也可以优先使用这种方法进行试探性检索，有一定的几率可以很快获得满意的结果，并且可以通过统计分析排序靠前的专利，发现更多的关键词表述方式或相关的专利分类号。

完全独立使用智能语义检索，虽然可以提升检索的效率，但仍然不能替代检索专家和布尔检索。其中一个主要原因是：智能语义技术无法进行技术方案层面的理解，换句话说，智能语义技术还不能像检索专家一样，提炼出技术方案的核心技术特征开展精准检索。例如针对下面的一段文本，检索专家基于其中描述的技术方案和对现有技术的理解，会将“缺口”和“折弯部”这两个特征作为核心特征来检索，而目前的智能语义算法还无法模拟检索专家从文本到技术方案这一抽象的思维过程。

一种一体式自拍装置，包括伸缩杆及用于夹持拍摄设备的夹持装置，所述夹持装置包括载物台及设于载物台上方的可拉伸夹紧机构，其特征在于：所述夹持装置一体式转动连接于所述伸缩杆的顶端，所述载物台上设有一缺口，所述夹紧机构设有一与所述缺口位置相对应的折弯部，所述伸缩杆折叠后可容置于所述缺口及折弯部。

使用智能语义最佳的方式是与专家布尔检索结合使用。具体来说是利用布尔检索式获得检索结果后，再利用一个与布尔检索式无关的排序因子对上述检索结果进行排序。布尔检索式用于精准的表达技术方案的核心特征，智能语义将其他非核心特征用相关性来表述，用排序的方法展示给检索专家。这样检索的好处是，一方面可以充分发挥检索专家定义核心特征的作用，保证了检索的精准度，同时发挥智能语义算法搜索的全面性，避免了因为对非核心技术特征的表述不够全面而导致的漏检。

针对上面的例子，如果独立使用布尔检索，则至少要在检索时限定出自拍装置、夹持装置、

缺口、弯折部等特征，但将语义与布尔检索结合时，可以仅用布尔检索搜索缺口、弯折部等核心特征，再将该全部文本作为排序因子对检索结果排序，就能将与自拍装置、夹持高度相关的专利排序到最前。这种检索方式既提高了检索的准确性，又避免了漏掉重要专利，检索的效率也能大幅提升。

每当听到人工智能将会取代某些工作时，相信不少人会有一丝担忧。但笔者认为，专利

检索本身是一种目标高度不确定、需要大量主观思考判断和综合分析的工作，检索专家们完全不必把智能语义看作是竞争对手，而是当成今后工作中的重要工具为自己赋能。相信语义技术与专家布尔检索的结合会是未来检索的主流，基于海量数据模型的语义排序与基于布尔的精确命中，将会合奏出人与人工智能融合的完美乐章。

本文不等同于法律意见，如需专项法律意见请咨询我公司专业顾问和律师。我公司电子邮箱为：LTBJ@lungtin.com，该电子邮箱也可在我公司网站 www.lungtin.com 找到。

如需更详细的信息或进一步的法律咨询，请与本文作者联系：

周俊：高级项目经理、专利分析师、前专利审查员：LTBJ@lungtin.com



周俊

(高级项目经理、专利分析师、前专利审查员)

周俊，大连理工大学车辆工程硕士，中国国家知识产权局前专利审查员，曾挂职国家知识产权局知识产权发展研究中心(CNIPA-IPDRC)，参与专利分析项目管理及研究工作，现任中专隆天知识产权运营(深圳)股份有限公司高级项目经理。他参与了多项由国家知识产权局组织的专利预警和专利导航项目，协助国家知识产权局及各级政府开展专利与产业决策；主持了多项面向企业的专利竞争性情报分析项目，为企业客户研判研发方向、了解竞争对手技术动态、明晰技术竞争格局以及进行创新资源整合提供帮助。他深耕新能源汽车和智能网联汽车领域，并在专利辅助研发和专利组合战略等方面拥有丰富的知识和经验。