

公示材料

一、基本信息			
项目名称	中文	新一代人工智能专利机器翻译研究与应用	
	英文	Research on New Generation of AI Patent Machine Translation System and Application thereof	
成果申报等级		<input checked="" type="checkbox"/> 一等奖 <input type="checkbox"/> 二等奖 <input type="checkbox"/> 三等奖	<input checked="" type="checkbox"/> 是否同意调级
主要完成人		张孝飞、吕荣波、马伟、张迁、周聪、葛昱晖	
主要完成单位		知识产权出版社有限责任公司、北京中献电子技术开发有限公司	
推荐单位(盖章)		国家知识产权局	
奖项的主要项目来源		<input checked="" type="checkbox"/> 国家级 <input type="checkbox"/> 省部级 <input type="checkbox"/> 其他	
具体计划、基金的名称和编号: 1、国家科技支撑计划重点项目“协同式多语言云翻译服务平台与应用”课题六“知识产权领域多语言翻译应用示范”(2012BAH14F00)。 2、知识产权出版社有限责任公司 2019 年十六项重点工作项目“多语种人工智能机器翻译系统”。			
成果的主要项目起止时间		起始: 2012-5	完成: 2021-6
组织验收/鉴定单位		中华人民共和国科学技术部; 知识产权出版社有限责任公司	
成果登记号	5602020Y0004	成果登记时间	2021 年 9 月 20 日

二、奖项简介

（主要技术内容、技术指标、创新点、授权知识产权情况、应用推广及取得的经济、社会效益等；限 1 页）

本项目致力于新一代人工智能专利机器翻译的研究与应用，形成了成熟、完整的知识产权数据服务生态，提供多款面向不同用户需求的企业级产品，突出展现了在知识产权数据应用领域的技术领先优势和产品布局前瞻性。

本项目能够提供专利数据多语言机器翻译，突破了全球知识产权交流中的语言障碍，一方面，有利于国家知识产权局利用信息化手段提升专利审查能力，发挥人工智能核心技术赋能知识产权强国战略；另一方面，可以增强企业对全球知识产权资源的深度利用，推动知识产权基础信息互联互通和传播利用，推动有自主知识产权的服务和产品‘走出去’，持续提升国际影响力和竞争力。

本项目以国际主流神经网络机器翻译模型为核心，同时，融合模板、统计、规则和实例等多种机器翻译方法，结合专利大数据针对专利文献特定域信息的用词习惯、句法结构以及语义选择等特点，对翻译知识库进行优化配置和构建；系统在漏译处理、术语一致性处理、长句复杂句和专利特定句式处理、权利要求项特殊格式处理及化学分子式、上下标等特殊字符处理上取得了突破性进展。

本项目开发的机器翻译引擎实现了专利数据在中、英、日、德、法、俄、韩、西、葡、阿 10 大 PCT 语种、共计 90 个语种方向的互译，机器翻译质量可媲美中高级人类译员翻译水平，中日双向翻译的人工评测准确率接近 98%，中英、中德双向翻译的人工评测准确率接近 95%，处于国际领先地位。同时，本项目的机器翻译引擎作为一款成熟的机器翻译产品，具有极高的稳定性和商业应用价值，在响应速度方面不低于 1000 词/秒，并发数不低于 1000，能够保证多用户调用机器翻译服务进行批量翻译时的整体效率，并且拥有高性能的服务器集群，平均每天翻译专利超 30 万件。

本项目采用的核心创新技术有：多策略神经网络机器翻译模型、面向神经网络的多语种词法分析和多源信息融合技术、海量翻译知识库的自动获取及多层次语料双语对齐技术、大规模预训练迁移学习、用词一致性处理和高效的基于句子级细粒度负载均衡调度算法等。

本项目创新性突出，其核心算法及相应的支撑软件共获得了 8 件发明专利及 15 件软件著作权的授权，形成了比较完善的核心技术知识产权布局体系。

本项目发布以来，获得新华网、光明日报、经济日报、China Daily 等多家媒体的关注和报道，指出新一代人工智能专利机器翻译将为缩短专利审查周期、提高专利领域从业人员的工作效率作出突出贡献。目前，对局服务和商业化服务体系日臻完善，现有审查员用户 1 万多名，每日翻译超过 5 万人次，在线翻译系统用户注册量 20 多万，并相继与 IBM、法国 Questel、中国贸促会、香港永新和中咨等大型涉外专利代理机构、律师事务所和科技公司签订合作协议，经济效益和社会效益显著。