附件4：

北京市工程技术研究中心三年绩效考评报告

工程中心名称：北京市海量语言信息处理与云计算

应用工程技术研究中心

依托单位：北京理工大学

联系人：祝烈煌

联系电话：010-68913536

手机：13522998905

电子邮箱：liehuangz@bit.edu.cn

依托单位科技主管部门联系人：杜宁

联系电话： 68918655

手机：13810982505

电子邮箱：ninger1979@bit.edu.cn

北京市科学技术委员会

二○一四年制

报告说明

1.本报告是为北京市工程技术研究中心（以下简称“工程中心”）绩效考评而设计。各工程中心确保所写内容真实、客观、准确。

2.本报告中的相关统计数据时间为自2011年1月1日起至2013年12月31日。各年份相关数据必须和当年提交的年度报告保持一致，与年度报告相关数据不符均视为无效数据。

3.在确认本报告编写准确无误后，应在依托单位内部进行公（不少于5个工作日）。依托单位应在承诺函的相应位置签字盖章，否则本报告无效。

4.本报告附件各栏可根据实际需求自行加页，所加附件一律使用A4纸张。

5.本报告中不得出现《国家科学技术保密规定》中列举的属于国家科学技术涉密范围的内容。

北京市工程技术研究中心绩效考评承诺函

根据北京市工程技术研究中心绩效考评有关文件要求，依托XXXX单位组建的XXXX工程技术研究中心参加本次绩效考评，并承诺如下：

1. 所提供的报表数据、文字资料及有关附件材料真实、

准确、完整；

1. 对所提供的资料真实性负责；
2. 不干预绩效考评工作。

工程技术研究中心主任（签字）：

年 月 日

工程技术研究中心依托单位盖章：

年 月 日

**一、工程中心基本情况统计表**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 基本信息 | 中心名称 | 北京市海量语言信息处理与云计算应用工程技术研究中心 | | | 依托单位 | | | | 北京理工大学 | | | | | | 共建单位 | 华建宇通科技（北京）有限责任公司 | | | | |
| 目前中心主任 | 黄河燕 | 职称 | | 教授 | | | | 手机 | | 13910846166 | | | | 电子邮箱 | hhy63@bit.edu.cn | | | | |
| 认定时中心主任 | 黄河燕 | | | 认定时技术委员会主任 | | | | 李德毅 | | | | | | 目前技术委员会主任 | 李德毅 | | | | |
| 主要运行地址 | 北京理工大学中心教学楼十层 | | | | | | | | | | | | | | | | | | |
| 认定时研究方向 | 机器翻译、语义计算与语言资源建设、信息检索与知识管理、信息抽取与信息挖掘、海量数字资源管理、云计算技术研究 | | | | | | | | | | | | | | | | | | |
| 技术水平与成果转化 | 目前研究方向 | 机器翻译、语义计算与知识工程、机器学习与WEB挖掘、信息检索与社会计算、海量数字资源管理、云计算应用与安全 | | | | | | | | | | | | | | | | | | |
| 技术成果水平 | 承接项目 | 年份 | | | 国家科技计划项目（科技部项目）、国家自然基金委员会项目 | | | | | 其他省部级项目 | | | | | 横向合作项目 | | | | |
| 数量 | | 年度实到经费（万元） | | | 数量 | | 年度实到经费（万元） | | | 数量 | 年度实到经费（万元） | | | |
| 2011 | | | 3 | | 35.3 | | | 3 | | 362.6 | | | 12 | 199.8 | | | |
| 2012 | | | 7 | | 196.8 | | | 5 | | 502.0 | | | 24 | 461.6 | | | |
| 2013 | | | 19 | | 597.4 | | | 9 | | 205.7 | | | 18 | 255.7 | | | |
| 总计 | | | 29 | | 829.5 | | | 18 | | 1070.3 | | | 67 | 917.1 | | | |
| 新技术/新产品/新工艺（项） | 0 | | | | | | | | 累计经济效益（万元） | | | | | 0 | | | | |
| 发明专利申请（项） | 国内 | | 国外 | | | | 发明专利授权（项） | | 国内 | | | | | 国外 | | | | |
| 47 | | 0 | | | | 16 | | | | | 0 | | | | |
| 研究论文（篇） | 国内 | | | | | | 国外 | | | | | SCI/EI/收录 | | 著作（部） | | | | |
| 2 | | | | | | 9 | | | | | 11 | | 1 | | | | |
| 技术标准（项） | 国际标准 | | | | | | 国家标准 | | | | | 行业标准 | | | | | | |
| 0 | | | | | | 1 | | | | | 0 | | | | | | |
| 其他 | （主要填写等同于发明专利的成果数量，如新药证书、动/植物新品种、临床批件等） | | | | | | | | | | | | | | | | | |
| 获奖情况（项） | 国家级奖项 | | | | | | | 其他省部级奖项 | | | | | | 行业协会等其他奖项 | | | | |
| 特级 | 一等 | | | 二等 | | 三等 | 特等 | | 一等 | | 二等 | 三等 |  | | | |
| 0 | 0 | | | 0 | | 0 | 0 | | 0 | | 0 | 3 |  | | | |
| 技术创新的贡献度 | 成果转化（项） | 19 | 产生直接经济效益（万元） | | | | | | 2273.3 | | 其中在京转化（项） | | | 12 | 产生直接经济效益（万元） | | | 1590.4 | |
| 队伍建设与  人才培养 | 队伍结构情况 | 认定时专职人员数量 | 54 | 现有专职人员数量 | | | 60 | | 副高级（含）以上职称数量及所占比例 | 33.3% | | 博士数量及所占比例 | | | 58.3% | 副高级（含）以上职称中40岁(含 )以下数量及所占比例 | | | 61.1% | |
| 青年骨干人才培养情况 | 人才引进数量 | 0 | 千人计划 | | | | | 0 | 海聚工程 | | 0 | | | 其他 | 0 | | | | |
| 人才培养数量 | 6 | 科技北京领军人才 | | | | | 0 | 科技新星 | | 0 | | | 其他 | 6 | | | | |
| 工程人员培训情况 | 培训次数 | 2 | | | | | | 培训数量 | 6 | | 职称晋升（人/次） | | | | 5 | | | | |
| 开放交流与  运行管理 | 开放交流 | 开放课题（项） | 0 | 总金额（万元） | | | | | 0 | 访问学者或兼职人员（人/次） | | 6 | | | 对外技术服务（次） | 2 | | | | |
| 技术委员会召开次数 | 2 | 主办国际会议（次） | | | | | 2 | 主办全国性会议（次） | | 0 | | | | 在国际会议做特邀报告（人/次） | | 2 | | |
| 仪器设备纳入首都科技条件平台数量（台/套） | 0 | 总金额（万元） | | | | | 0 | 国际科技合作基地（国家级/市级/否） | | 否 | | | | 科普基地（是/否） | | 是 | | |
| 依托单位支持 | 实验室现有  科研面积（㎡） | 考评期内新增科研面积（㎡） | 实验室现有仪器设备数量（台/套） | | | 现有仪器设备原值（万元） | | 考评期内新增仪器设备数量（台/套） | 新增仪器设备原值（万元 | | 经费投入（万元） | | | 2011年 | 15 | | | | |
| 2012年 | 75 | | | | |
| 4100 | 0 | 1320 | | | 2190 | | 123 | 420 | | 2013年 | 15 | | | | |

填表说明：1、表中各项数据均需和附件明细表保持一致，请各机构先整理附件明细表，后填表。数据前后不一致均视为无效数据；

2、对外技术服务以是否有发票为次数凭证。

**二、工程中心在考评期内的运行绩效**

**（一）发展规划与目标完成**

1.认定时规划目标完成情况

工程中心瞄准国家战略需求和学术前沿，结合工业化、信息化、国防现代化建设需要，以促进北京市经济建设和社会发展需求为目标，凝聚学科发展特色，围绕多知识多策略融合的机器翻译模型、基于多特征知识标记的语言分析算法、基于篇章语义的上下文相关处理研究、有效的文本译前预处理和复杂长句分析处理机制、实用型人机互助机器翻译、少数民族语言机器翻译、特定领域的对话翻译系统等认定时工程中心规划的机器翻译领域重点研究内容，以及语义计算与资源建设、信息检索与知识管理、信息抽取与信息挖掘、海量数字资源管理、云计算技术研究等研究方向，开展了协同式多语言云翻译服务平台与应用、面向言语的多策略融合机器翻译方法、基于本体的多策略民汉机器翻译研究、基于自然语音人机交互的信息搜索系统研发和产业化、社交网络分析与网络信息传播等重点项目的研发，在学术论文、发明专利、科技获奖等方面都取得了新的突破，达到国际领先、国内一流，实现预期目标与建设水平。

工程中心自认定以来，承担了国家973计划课题、国家863计划课题、国家科技支撑计划项目、国家自然科学基金重点项目等国家重大/重点项目10余项，科研经费达到近3000万元，超额完成认定时工程中心规划的研发经费增收指标。通过研发投入、成果转化以及依托单位专项经费支持，工程中心的科研条件和配套设施进一步改善，有利支撑了工程中心关键技术研究和重要成果转化。工程中心在北京落地的成果项目达到12项，直接经济效益达1500余万元，转化项目不仅取得经济效益还取得巨大的社会效益，影响范围覆盖全国，基本完成预定目标。

工程中心现有员工58名，其中教授8名，副教授/副研究员11名。自认定以来，工程中心新增教授2名、副教授3名，博士生导师3名；新获批教育部新世纪优秀人才2名，北京市高等学校青年英才4名。黄河燕主任遴选为国家863计划主题专家，并担任中国人工智能学会副理事长、中文信息学会副理事长等职务。

2、未来三年发展规划

工程中心将基于现有工作基础，瞄准大数据时代下的智能信息处理与信息安全等战略机遇，以服务首都经济建设与文化发展为目标，以国家重大战略需求、重大行业应用需求为导向，开展关键技术攻关、重大应用系统开发、特色技术产业化应用、高水平人才队伍建设、拔尖人才培养，全力打造该领域国家级创新团队。

具体规划分为如下方面：

1. 面向社会媒体文本大数据处理与舆情服务

基于互联网的社会媒体大数据正在成为人类社会中社会关系维系和信息传播的重要渠道和载体，对国家安全和社会发展都会产生深远的影响：社会个体通过各种连接关系在社交网络上构成“关系结构”，包括以各种复杂关系关联而成的虚拟社区；基于社交网络的关系结构，大量网络个体围绕着某个事件而聚合，并相互影响、作用、依赖，从而形成具有共同行为特征的“网络群体”；基于社交网络关系结构和网络群体，各类“网络信息”得以快速发布并传播扩散形成社会化媒体，并反馈到现实社会，从而使得社交网络与现实社会间形成互动，并对现实世界产生影响。

工程中心将面向国家重大战略需求，重点围绕社交网络的结构大数据、群体大数据、信息大数据这三个维度开展前瞻性的理论研究。并在网络结构分析、特定社区发现、个人建模、群体情感分析、社会网络热点和敏感问题发现等关键技术上形成一批知识产权，并开展舆情分析服务产业化示范验证，直接为国家安全与首都政治、经济建设服务。

2.语言深层分析理解与智能信息服务

基于工程中心现有研究基础，继续通过对本体和描述逻辑、计算语义学、机器学习和自然语言处理等多研究方向交叉融合的探索，并在智能知识理解与推理的关键科学技术展开研究及智能信息服务产业应用。

工程中心将围绕大规模数据深层特性利用，采用大规模异构源的深层语义关联计算模型、本体映射、基于多任务学习与迁移学习的跨领域本体自动构建等方法，研究大规模本体知识库并构建高质量Word Embedding库；基于其良好的知识结构和丰富的知识来源，研究词汇语义、篇章结构、领域概念扩展的多层次理解方法，实现基于本体的篇章结构理解和基于海量知识库的概念扩展化理解；还将针对多种类型的推理任务以及大规模本体知识库上查询推理系统的可扩展性（Scalability）问题，进行算法层优化和系统层优化，研究并实现面向海量知识的可扩展高效推理机制。最终将面向社会重大行业和特色领域（如：自动问答）进行技术产业化应用，提供智能信息服务。

3.优化协同式多语言云翻译服务

据前瞻产业研究院发布的一项报告《2013-2017年中国语言服务行业发展前景预测与投资战略规划分析报告》，2011年世界翻译市场的规模为314亿美元，中国大陆翻译市场规模为300亿元，年复合增长率为9%。至2011年，预计的我国的翻译市场规模达到503亿元。近年来，翻译市场的迅猛发展也带动了机器翻译技术的发展。据市场调查公司Research and Market的一份报告支持，全球机器翻译市场在2012年达到16亿美元，2012-2016年期间的额年复合增长率将达到18.05%。北京作为我国首都，是中国与世界在政治、经济、文化等各个方面的沟通与交流的重要窗口。越来越多的国际活动选在在京举办，如：刚刚结束的2014年APEC，以及2016年将要召开的G20，迫切需要实时、高效、高质量的多语翻译服务支撑。

工程中心将继续深入开展海量结构化和非结构化数据管理、高性能海量语言信息计算等典型多语言机器翻译研究问题，综合采用云存储、MapReduce云计算模式等技术，优化系统平台层，扩大语种支持及示范应用范围，提供面向北京、辐射全国的协同式多语言云翻译服务。

4.以开源共享与协同创新为主导的研究成果推广，构建良性的产学研生态系统

产业推广是工程中心的特色和优势所在，与国内其他同类工程中心一样，产业转移成功率较低是当前技术转移的客观现实。工程中心在当前三年的建设期也存在科研产业辐射程度相对较低的问题。，未来三年内，工程中心将探索更加科学有效的技术转移策略。将以开源共享与协同创新为主导，将中心的研究成果向社会公开共享，通过当前已有的共建单位华建宇通、协同创新平台、技术开放日、大数据工作委员会、中国人工智能学会、中国中文信息学会，及工程中心现已开放的自然语言处理与信息检索共享平台等多种形式开展有效的技术服务工作，全力构建良性的产学研生态系统。

5. 以服务北京为基点，面向国家重大信息安全需求，整合产学研资源，全力打造国家级创新团队。

工程中心拥有一批有影响力的科研团队和产业精英，目前，国家将信息安全列为国家的重要战略，结合我们已有的研究基础，工程中心将全力整合各类资源，服务首都社会经济发展，积极开展工作，全力打造国家级创新团队，申请国家级创新平台，辐射全国，分布有序地建设工程中心，成为本领域创新的品牌。

**（二）技术水平与成果转化**

1、定位与研究方向情况

着眼产学研用相结合，形成独具特色的研究和科技转化成果，将工程中心建设为海量语言信息处理及云计算应用领域内具有国际影响力的研究开发平台。

工程中心认定时，设定的研究方向是机器翻译、语义计算与语言资源建设、信息检索与知识管理、信息抽取与信息挖掘、海量数字资源管理、云计算应用技术研究。结合国家重大战略需求、社会重大应用需求，以及相关领域的技术发展趋势和工程中心技术优势，同时综合国家经费资助及科研人才培育与引进情况、技术委员会历次专家建议工程中心对原有研究方向进行了凝练和调整，调整后的研究方向为：机器翻译、海量数字资源管理、机器学习与Web挖掘、信息检索与社会计算、语义计算与知识工程、云计算应用与安全。

1. 机器翻译

主要研究多知识多策略融合的机器翻译模型、基于多特征知识标记的语言分析算法、基于篇章语义的上下文相关处理研究、有效的文本译前预处理和复杂长句分析处理机制、实用型人机互助机器翻译、少数民族语言机器翻译、特定领域的对话翻译系统研制。

1. 海量数字资源管理

主要研究海量数字资源管理系统体系框架研究与设计，抽象数据类型描述理论，分布、并行和协作环境下的海量信息管理的事务模型及处理方法，海量资源加工及多级存储体系构建，海量信息的科学组织和互操作，数字资源的索引、查询和个性化服务等。

1. 机器学习与Web挖掘

主要研究机器学习的理论与算法、社会网络分析与信息传播的基础理论与关键技术、深度学习（Deep Learning）的理论与应用、大规模Web数据的处理与利用为核心的数据挖掘与推荐算法。

1. 信息检索与社会计算

主要研究面向社会媒体的自然语言处理技术、社会网络分析与复杂系统，采用信息检索、信息抽取等方法进行社会媒体挖掘、预测与推荐，有效支撑社会媒体舆情分析与精准营销等新型互联网应用。

1. 语义计算与知识工程

以语义深层计算为目标，主要研究面向多领域应用的基于本体的专业分类机制、多语专业术语库的建设与标准体系建立、基于多特征知识标记的多语言知识库/资源库建设。

1. 云计算应用与安全

主要研究海量视频云数据管理、海量地震数据成像云计算技术、面向仿真的云计算支撑平台技术研究、数字化制造协同管理云计算技术等。研究面向云存储的加密数据检索、基于属性加密的云存储密级管控、面向云存储的数据持有性审计技术等技术。

2、技术成果水平

工程中心瞄准国家重大战略发展需求和世界科技发展前沿，坚持“强化基础、着力前沿、协同创新、持续发展”的科研发展思路，不断加强产学研合作，成功获批包括国家973计划课题、国家863计划课题、国家科技支撑计划项目、国家自然科学基金重点项目等国家重大/重点项目10余项。

工程中心非常重视成果产业化和首都重大需求。“协同式多语言云翻译服务平台”的研发与实际应用结合，增强了知识产权领域专利文献在云翻译平台的翻译效率和应用覆盖面；“多策略智能辅助翻译系统及工程管理平台”解决了翻译知识分散化，工作流程的无序化，促进了翻译过程的协同化和标准化。

工程中心在科研奖励、高水平学术论文、标准和专利等科研成果方面均取得了新的突破，形成一批标志性研究成果。

（1）协同式多语言云翻译服务平台

据前瞻产业研究院发布的一项报告《2013-2017年中国语言服务行业发展前景预测与投资战略规划分析报告》，2011年世界翻译市场的规模为314亿美元，中国大陆翻译市场规模为300亿元，年复合增长率为9%。至2011年，预计的我国的翻译市场规模达到503亿元。近年来，翻译市场的迅猛发展也带动了机器翻译技术的发展。据市场调查公司Research and Market的一份报告支持，全球机器翻译市场在2012年达到16亿美元，2012-2016年期间的额年复合增长率将达到18.05%。北京作为我国首都，是中国与世界在政治、经济、文化等各个方面的沟通与交流的重要窗口。越来越多的国际活动选在在京举办，如：2008年奥运会、2014年APEC，以及2016年将要召开的G20，迫切需要实时、高效、高质量的多语翻译服务支撑。

工程中心及时把握技术发展趋势，聚焦专利信息应用服务领域，研发了协同式多语言云翻译服务平台。该平台综合采用基于规则、实例和统计知识的多种自然语言处理技术，实现多策略机器翻译；在译后加工及工程管理的基础上，通过为翻译人员提供良性人机交互接口使其能够灵活控制翻译结果，通过规则分析实现多种翻译知识的积累与动态转换，实现面向对象的工程、任务和用户管理；在系统平台层，综合采用云存储、MapReduce云计算模式等技术，处理海量结构化和非结构化数据管理、高性能海量语言信息计算等典型多语言机器翻译研究问题。

同时，构建了大型英\日\汉知识产权专业知识库，研究针对知识产权领域的机器辅助翻译系统和云翻译服务平台优化策略，开展英\日\汉云翻译服务应用示范，不但有效提高了翻译工作的效率，还在保证翻译质量的同时，提高速度、降低成本，对翻译服务的未来产业发展有深刻意义。该平台目前已在国家知识产权局部署实施，示范应用。

（2）多策略智能辅助翻译系统及工程管理平台

语言分析理解是极其复杂的思维活动,是图灵测试机器是否具备智能的标准之一。辅助翻译需求领域相对集中，具有显著领域特点，亟待高效流畅、安全可控的流程化工程管理，以满足不同需求（特定行业、海量、涉密等）的翻译任务。工程中心充分发挥共建单位各自的技术优势，共同研发了多策略智能辅助翻译系统及工程管理平台。该系统平台集交互翻译、双语语料对齐、翻译知识及工程任务流程化管理于一体，覆盖8个语种、14个汉外语种对。创新性地提出了多特征知识一体化的实例模式表示方法及基于多层次特征相容匹配的启发式约束实例模式检索方法，不仅能表示复杂的模式成分及其相互关系，而且可提高模式检索速度，保证检索到的模式具有深层次的相似性，提高类比翻译的准确性和效率；设计实现了人性化、智能化的交互式辅助翻译平台，为专业翻译人员和企业情报信息处理中心提供了集成信息翻译处理平台，支持人机交互过程的自动跟踪、积累和学习，提高了系统的适应能力；实现了多元信息流特征识别与变换子系统，支持对HTML、DOC、TXT、RTF、XLS、PPT等多种格式文本信息的特征识别与变换恢复处理，消除了翻译人员繁琐的文字排版工作；提出了基于多知识分析的双语语料对齐算法，结合多层次分段对齐及长度和词汇信息结合的句子对齐方法，增加了计算句子相似度的可靠性，提高了对齐效率，大大缩小了发生错误的影响范围。语料对齐的召回率高达97%以上，精确率超过95%；提出并实现了面向对象实例模式知识库的分布共享和管理机制及面向工程的多用户协同式翻译及基于安全职责权限分类的安全管理机制，人员互相制约、所辖数据相互独立，避免了对象间的并发冲突；对翻译资源、任务、人员与目标实现统一管理和翻译处理流程的电子化、标准化，从而降低了翻译人员在翻译流程中的劳动强度，大幅提高了总体翻译效率。专家鉴定该平台“在人机互动的智能辅助翻译平台、双语对齐平台和翻译工程信息管理平台等方面具有创新性，在汉外、外汉多语机器翻译领域达到国际领先水平” 。该成果先后在国家知识产权出版社、中国科学技术信息研究所信息技术支持中心、航空工业信息中心、北京利云技术开发公司、中航工业洪都、北京三信时代信息公司、中国空空导弹研究院、中航工业综合技术研究所等企事业单位应用，并已获得北京市科委科技奖（已公示 2014）。

（3）云计算应用及云存储数据安全处理平台

将具有超级计算能力的图形处理单元-GPU添加到云计算平台的虚拟化管理中，并提供具有一定通用性的高性能开发套件，简化高性能计算应用开发过程。

围绕云存储安全问题重点研究了面向云存储的加密数据检索、基于属性加密的云存储密级管控、面向云存储的数据持有性审计技术等关键技术实现了面向云存储的加密数据处理平台，用户使用自己的密钥加密数据后再存储到云端服务器有效防止云存储服务提供商和非授权用户访问私有数据的同时，通过多种面向云存储的加密数据处理技术实现云服务提供商对加密数据的高效处理。

云计算环境中，存储设备是信息处理过程的起点及终点，具有十分重要的地位，而国外存储装备暴露出的信息安全问题十分严重。为适应国家信息化建设中“自主可控”的战略要求，研究了具有自主安全功能的存储设备，切实保障系统存储安全。所研制的固态盘、U盘、存储卡等存储设备具备自主可控安全功能，在安全控制传输通道、功能隐蔽性、闪存介质数据映射、数据自毁及逻辑存储空间保护等方面进行了原始创新，实现了存储设备安全功能的自主、可控、可信。产品已投入实际应用，通过了部级鉴定， 获得国家发明专利12项，获得部级科技进步奖2项。

（4）基于领域本体的特定领域知识管理平台

该平台充分利用本体良好的语义信息，实现领域知识的存储、检索及复用；利用自然语言处理技术，获取并存储用户工作过程中产生的领域知识，实现领域知识的积累。针对各领域内项目流程日趋复杂的现状，研究开发流程设计器，使管理人员可以将一个项目分为多个流程，分配给各部门的相应人员去完成，充分发挥知识管理平台的强大作用。该成果申请技术发明专利12项，已获得工信部国防科工局科技成果鉴定。

(5) 网络数字资源组织与个性化检索服务平台

该平台针对网络数字资源特点，支持可扩展的元数据与数字资源互操作规范，支持多种数字资源知识抽取方法，为用户提供基于浏览和点击的个性化检索服务，该平台可用于网络新闻及各类垂直领域数据检索，可为舆情分析等提供基础平台。该平台的相关技术及应用获国家发明专利1项，软件著作权1项，获国防科技进步3等奖一项。

3、成果转化与市场结合能力

工程中心是由北京理工大学与北京市海淀区高新技术企业华建宇通科技（北京）有限公司校企联合共建，是集双方特色与技术优势的产学研一体化联合体，具备天然的科研深化、成果转化和市场推广能力。目前已共同研制基于云计算的海量语言信息处理系列平台，提升了机器翻译、知识管理等产品的服务模式，提高了中心产品的核心竞争力。

目前，工程中心的相关成果通过专利转让、技术转化等多种形式，直接助力在京企业取得了显著的经济效益。同时，程中心结合北理工计算机学院的技术和人才优势，在科学研究、联合立项、人才培养和科研成果转化等方面与合作单位开展全方位多领域合作，先后与中国电子信息安全研究院、中国电子六所、微软（中国）有限公司、同有飞骥等公司展开密切合作，签订战略合作协议，促进了工程中心技术成果应用水平的提高。

工程中心在科研立项之前就与市场需求紧密结合，进行有目的的开发研究，提升成果的时效性，防止成果与市场脱节，避免成果成为“无源之水，无本之木”，使创新资源得到充分利用。

4、技术创新贡献度

近三年来工程中心在技术创新方面进行了极大的人力、物力投入，已经在经济效益、社会效益和人才培养与储备方面取得明显收益。

工程中心所属技术直接在京转化成果12项，产生直接经济效益1500余万元。仅“航空协同辅助翻译系统”一项，在北京已在航空领域（航空发展研究中心、中航工业综合技术研究所）取得直接经济效益近30万元，目前已推广到中国空空导弹研究院、中航工业洪都等厂所。工程中心成果不仅带来可观的经济效益，促进了北京的经济发展。工程中心现阶段的技术贡献在服务社会方面的体现尤为显著。多项成果直接服务于国家重要部门、国家重要行业，及社会公益性组织机构。

“专利领域专业文献机器翻译系统”是工程中心与国家知识产权出版社合作项目，主要实现专利文献的翻译。随着世界经济一体化发展，国际交往会越来越频繁，语言理解的障碍也将越来越明显。随着经济领域的交流的逐步增多和各国对专利的重视程度的加强，专利的查询及翻译需求日益增多。目前，我国已经成为仅次于美国的第二大专利申请国，专利申请数量的快速增长，带动了我国专利数据资源服务市场规模的快速发展。经常使用专利数据的个人用户已经达到数百万的规模。专利行业的翻译需求市场巨大，前景广阔。然而伴随翻译需求的快速增加，我国整体的翻译能力却无法快速提升，翻译能力无法满足翻译需求的不断增加。目前翻译主要是依靠人工翻译，然而人工翻译周期长，成本高，高质量翻译人员短缺，远远解决不了实际问题。“专利领域专业文献机器翻译系统”的应用可以有效的解决专利领域翻译遇到的瓶颈，为北京乃至全国的专利服务行业提供全面的翻译服务，配合另一成果“协同辅助翻译系统”也可为其他行业的翻译工作提供有力的帮助。系统的投入与应用成功突破专利翻译语言间障碍，增强了我国专利多语信息翻译处理能力，提高了我国在政治、经济、文化等各方面的国际地位和竞争能力。

中心成员作为项目负责人承担新闻出版重大科技工程项目-“数字版权保护技术研发工程-数字内容分段控制技术研发”,该系统是“数字版权保护技术研发工程”技术核心包，主要研发基于分段的数字内容分段技术，综合考虑数字内容的动态性、授权的安全可靠性，以及版权保护在资源、流通模式等不同层面的需求和版权所有者、用户等各参与方的要求，通过数字内容分段加密、防复制动态分段授权和动态使用控制的方式，为数字内容的版权保护提供更细粒度的授权控制方法，以适应数字化市场发展和版权保护发展的需要，具有较大的应用和社会服务价值。

“盲文书籍发布系统”是工程中心成员与中国盲文出版社合作的项目,实现了盲文图书出版的数字化。系统符合自动化盲文操作特点的字处理和排版软件，全程支持语音方式提示和点字输出显示，确保盲人工作人员能够顺利使用。中国现有盲人1,200多万，是世界上盲人最多的国家。而盲文——又称点字是盲人接受教育、获取知识的最重要的工具。中国盲文出版社是全国惟一为我国1691万盲人提供服务的出版机构，是全国盲人文化科技资源中心。但是当前，有限的盲文图书出版能力与不断增加的盲人读者需求，盲文图书制作的高成本与盲人群体普遍的低收入，这两大矛盾日益突显。由于盲文编码的特殊性，致使音乐类、外语类等盲文图书成为我国盲文图书出版的薄弱点，与国际相比存在巨大差距。每年出书的种类和数量都相当有限。据统计每年出版的盲文出版物只有140种左右。工程中心研发的盲文书籍编辑发布系统的研发解决了上述问题，改善盲人这一特殊的社会群体的生存状态，使他们能获得平等的教育、就业机会，从而真正融入主流社会，平等参与社会生活的各个方面，极大地促进了盲文出版领域信息化水平提升，降低工作人员的劳动强度和出版成本，提高盲文出版物的出版效率和速度，加快盲文出版事业发展步伐，扩大盲人读物出版规模，满足适龄盲人少年儿童的教材供应和日益增长的阅读需求。让盲人朋友平等享受信息技术发展进步的成果，是实现盲文朋友中国梦的重要体现，我国盲文出版事业的发展也具有十分重要的意义。结合工程中心在机器学习、数据挖掘、以及信息检索技术的核心数据保护平台，在市场中具有很强竞争力的优势。目前已经先后为军队、公安、卫生、金融等行业提供核心加固产品数百套。现阶段在智慧城市的云端核心数据保护方面进展顺利，与多家国有上市公司形成合作意向，预计在未来几年将会有较大的突破。并与青云（QINGCLOUD）形成战略合作，未来将在公有云和金融行业核心数据安全保护等方面形成突破。

在核心数据保护方面的技术创新和突破为本细分行业形成一个技术龙头作用，在一定程度上成为带领友商发展的方向。而云端核心数据安全保护技术又将为北京和全国云计算产品的发展保驾护航。此外，工程中心特别重视科学研究的开发共享，并对产业采用开源共享的形式输出关键技术，推出了自然语言处理与信息检索共享平台（www.nlpir.org），共享平台发布了大量的共享软件、一手的研究论文和资料，早在2009年其已经开始向社会免费发布了微博语料库、Twitter语料库500万条，是国内外最早对外共享微博语料库的机构。工程中心整合了多年的研究成果，完成了NLPIR网络搜索与挖掘共享开发平台，针对互联网内容处理的需要，融合了自然语言理解、网络搜索和文本挖掘的技术，提供了用于技术二次开发的基础工具集。开发平台由汉语分词、词性标注、大数据分类、大数据聚类、大数据搜索、新词发现、关键词提取、文本摘要等十一个个中间件组成，各个中间件API可以无缝地融合到客户的各类复杂应用系统之中，可兼容Windows，Linux， Android，Maemo5, FreeBSD等不同操作系统平台，可以供Java，C，C#等各类开发语言使用。该技术平台使用的用户遍布全球30余万机构，具有较大的社会效益。

人才培养也是工程中心认定时的一项关键目标，通过技术专研与创新，工程中心在此前三年的建设期内，锻炼了现有的科研梯队，培养了具备产学研实战经验的青年骨干人才，并吸纳了“新鲜”的科研力量。同时工程中心积极响应了北京市科协的号召，遴选优秀专家指导“北京青少年科技后备人才早期培养计划”中来自北京四中等六所中学的高一学生，赢得了良好的社会声誉与口碑。

**（三）队伍建设与人才培养**

1、工程中心主任与工程技术带头人作用

（1）工程中心主任

工程中心主任黄河燕教授是中心的创立者和领导核心，在工程中心的建设和发展中起主导作用。2009年，北京理工大学通过公开招聘从中国科学院引入黄河燕教授（时任共建企业华建集团董事长、总工程师）并聘任为计算机学院院长。黄河燕教授凝聚北京理工大学计算机学院从事海量语言信息处理、网络与信息安全方面研究的师资队伍，与共建企业华建宇通科技（北京）有限公司核心成员联合成立了海量语言信息处理与云计算应用工程技术研究中心。2011年5月，工程中心在黄河燕主任领导、技术委员会专家指导，及全体成员共同努力，申请并成功被认定为北京市工程技术研究中心。黄河燕主任具备深厚的科研背景，拥有丰富的工程和产业化经验，是难得的集产学研于一身的复合型领军人才，学术思想敏捷、思路开阔，善于遵循“产学研用”的战略发展思路，具有极强的组织管理和战略布局能力、能够前瞻地把握领域前沿趋势，制定了工程中心的中长期的战略规划，围绕海量信息处理与云计算应用两条主线，调整优化了工程中心的主要研究方向，亲力亲为参与部署了重要研发工作。通过两年左右时间，工程中心发展平稳，逐步建设成为国内在该领域具有很高影响力的科研平台。

工程中心主任黄河燕教授长期从事自然语言处理和机器翻译、智能信息系统、社交网络与信息检索等领域的科学研究与产业化应用推广，组织管理能力强，先后承担20多项国家级科研攻关项目、大型工程应用和国际合作项目，包括：国家自然科学基金重点项目“基于语义本体的多策略民汉机器翻译研究”、国家“973计划”课题“面向言语的多策略融合机器翻译方法”、国防工业基础科研项目“\*\*\*知识库管理系统”、总参预研项目“网络信息\*\*\*处理技术”、国家发改委软件产业化专项项目“多语信息处理及其应用技术产业化”、北京市科委科技奥运专项项目“面向奥运的多语言智能信息服务系统”、国家“863计划”重大项目课题“面向海量信息处理的多策略机器翻译研究”等；获得了国家科技进步一等奖、中国科学院科技进步一等奖、北京市科学技术二等奖、中央国家机关十大杰出青年等国家级和省部级奖励，1995年享受政府特殊津贴。发表学术论文100多篇，获得授权技术发明专利9项，软件著作权16项，多次应邀在国内外学术会议及大学作特邀报告。黄河燕主任同时具有很强的学术和社会影响力，现任北京理工大学计算机学院院长， BIT-DFKI（中-德）语言信息处理联合实验室主任，兼任国家高技术研究发展计划（863）主题专家组成员、中国人工智能学会副理事长、中国中文信息学会副理事长、国家自然科学基金委专家评审组成员,北京市学位委员会委员、教育部计算机专业教指委委员、International Journal of Advanced Intelligence副主编、《计算机研究与发展》编委等重要学术兼职。

（2）工程中心学术骨干

牛振东，工程中心副主任，计算机学院副院长/教授/博士生导师，中国索引学会副理事长，香港大学语言学系荣誉教授，匹兹堡大学信息学院兼职教授，中央编译局信息化咨询专家委员会主任委员，北京市信息化专家顾问委员会专家。研究领域包括复杂软件体系结构、海量数字资源组织管理、数字图书馆、数字化教育等，近年完成和承担包括973课题在内的近10项国家级课题及国际合作项目，担任多个国际专业期刊编委,在美国科学院院刊PNAS等顶级期刊和会议发表论文多篇。于2004年获霍英东基金会青年教师奖励，2005年获得IBM全球高校教师创新奖励，2006年入选教育部新世纪优秀人才计划。

廖乐健，工程中心副主任，计算机学院教授/博士生导师，。主要从事人工智能及其互联网应用方面的研究，包括语义Web及本体，Web挖掘与问答，自然语言理解，多智能体系统。北京理工大学网络与信息安全研究方向学科责任教授；培养博硕士研究生三十余名。先后主持国家自然科学基金项目“开发语义Web的设计方法学”、“基于约束规划的语义Web 动态服务组合” 等。在国内外学术刊物及会议上发表SCI/EI 检索论文六十余篇。先后担任多个国际会议的程序委员。

洪琳，工程中心副主任，华健宇通公司总裁。长期从事机器翻译系统的应用推广工作，参加过国家863计划重点项目、发改委产业化专项等多个国家级项目的研发工作获得了国家科技进步一等奖等省部级以上奖励5项。

谭毓安，计算机学院教授/博士生导师。现任北京理工大学计算机实验教学中心（北京市示范教学中心）主任，中国计算机学会高性能计算专委会委员, 信息存储技术专委会委员。主要研究领域为网络存储、信息安全、云计算等，近年来承担863计划子课题“智能云服务与管理平台核心软件及系统”、国家自然科学基金“连续数据存储系统节能关键技术研究”等项目10余项，发表学术论文50余篇，获得国家发明专利15项，国防科学技术进步奖等部级奖2项。

祝烈煌，计算机学院教授/博士生导师。于2012年入选教育部新世纪优秀人才，担任中国人工智能学会常务理事、智能信息网络专业委员会主任委员，中国计算机学会YOCSEF学术委员会副主席、互联网专委委员、传感器网络专委委员，中国电子学会信息论分会委员。主要研究领域为密码算法及安全协议、物联网安全、云计算安全、移动互联网安全和可信计算等，发表学术论文50余篇，获批国家发明专利8项，国防科学技术进步奖1项。

李侃，计算机学院教授/博士生导师。国家863评审专家，博士点基金/博士后基金评审专家，CCF互联网专委会委员，计算机应用专委会委员，YOCSEF委员，北京市青少年后备人才导师。目前主要从事机器学习、数据挖掘及分布式系统方面的研究。承担了国家973课题、863课题和国家自然科学基金等多项国家级和部级项目，获得了国防科技进步奖等科研奖励。在国内外期刊、学术会议上发表SCI/EI检索论文50余篇，发明专利近20项。

戴林 1977年生，博士，副教授，博士生导师。研究方向为信息检索、自然语言处理。主持和参与国家级项目超过20项。共发表学术论文超过40篇，其中SCI检索5篇，EI检索超过30篇。担任国家自然科学基金项目评阅人， IEEE计算机学会会员，IEICE会员，中国计算机学会会员，国际WordNet协会（GWA）会员。担任International Journal of Computer Mathematics（SCI）、IPL（SCI）、AICIT(EI)、Optics & Laser Technology（SCI）等国际期刊评阅人。

李凡 博士，副教授，博士生导师。主要从事物联网、无线网络、移动计算、传感网、车联网、容迟网络方面的科研工作。近年来在国际期刊和国际会议发表学术论文40余篇，其中SCI索引10余篇，EI索引30余篇，授权国家发明专利4项，申请专利10余项。作为项目负责人主持多项国家自然科学基金项目，北京市自然科学基金项目和教育部教育部留学回国人员科研启动基金项目，并作为主要研究者参与了国家863智能空间重点项目等。

2、队伍结构与创新团队建设

工程中心具有一支由北京理工大学专职教师、华健宇通公司管理人员和技术人员组成的学历层次高、年龄结构合理、学缘结构好、发展稳定的科研队伍。现有成员60名，其中教授/正高9名，副教授/副研究员11名，博士生导师9名，获得博士学位的比例达到58%。教育部新世纪优秀人才3名，北京市高等学校青年英才4名。

围绕机器翻译、语义计算与知识工程、机器学习与WEB挖掘、信息检索与社会计算、海量数据处理与云计算技术等工程中心重点研究方向，加强创新团队建设。每个方向团队由教授牵头，并由管理人员、学校骨干教师、公司骨干技术人员和市场人员组成。

3、青年骨干人才培养

工程中心通过培养与引进相结合，构建一支思想活跃有创新精神的青年骨干人才队伍。围绕工程中心的主要研究方向，引进具有国际影响力的优秀学科带头人和青年学术骨干。加快青年骨干人才培养，有计划选派优秀青年教师出国进修和深造。破除阻碍青年人才成长的诸多门槛和藩篱，为青年人才成长创造宽松的社会环境，鼓励青年教师承担国家科研项目。

通过近三年的建设，工程中心从香港中文大学、清华大学、北京大学、北京航空航天大学等高校引进6名优秀青年骨干教师。工程中心各种青年骨干教师培养政策促进了一批科研人才的快速成长，新获批教育部新世纪优秀人才2名，北京市高等学校青年英才4名。聘请的微软雷德蒙特研究院高级研究员华先胜，获批为北京理工大学徐特立讲座教授。

**（四）开放交流与运行管理**

1.技术委员会作用

技术委员会是工程中心的技术咨询机构，负责审议工程中心的发展战略、研究开发计划，评价工程设计与试验方案，提供技术经济咨询和市场信息，审议工程中心年度工作等。技术委员会是工程中心有序健康发展的重要保证。

在考评期内，工程中心每年定期召开技术委员会工作会议。工程中心主任向技术委员会汇报年度建设任务完成情况，介绍取得的建设成果和面临的问题，制定后续工作规划。技术委员会对年度工作进行评估，对面临的问题给出建议，对后续规划进行指导。

2.开放交流

工程中心非常重视与领域内的科研机构和高技术企业开展合作交流，有效提升了工程中心技术水平和科研成果转化能力。

工程中心与清华大学、北京大学、中国科技大学、中科院自动化所、中科院声学所、北京语言文化大学等科研机构开展项目联合申报、关键技术联合公关、青年骨干联合培养等多层次、全方位的交流合作。与中国电子信息安全研究院、中国电子六所、微软（中国）有限公司、同有飞骥等公司展开密切合作，签订战略合作协议。  
 工程中心非常重视国际交流，加强美、德、日、英、韩等国的机器翻译研究机构的学术和产业化交流关系。与德国人工智能研究中心联合成立了“BIT-DFKI语言信息处理联合实验室”，与悉尼科技大学联合成立“BIT-UTS数据挖掘与服务技术联合研究中心”，开展深入的交流合作。

每年邀请来自美国、澳大利亚、欧洲等地国际知名专家进行学术交流，接收国内高校青年教师作为访问学者来工程中心进行合作研究。

3.协同创新

工程中心非常重视与相关领域的其他实验室、工程中心加强合作交流，共同申报和承担国家973项目、国家863项目等项目。

工程中心积极加入中关村可信计算等产业联盟，加强与联盟单位的科技合作。工程中心早在2010年就创办了国际上第一家大数据论坛（www.bigdataBBS.com），成为了大数据学术和产业最早也是最有影响力的科技普及和交流平台之一，并被中国大数据专家委员战略白皮书列为大数据战略阵地之一。工程中心的张华平博士发起成立了中国互联网协会大数据工作委员会，并担任执行主任。中国互联网协会大数据工作委员会（筹）的宗旨定位为：中国大数据的政策制定、学术研究、产业应用与教育培训的公益组织，构建大数据政、产、学、研、用多位一体的良性生态体系，以众筹方式搭建大数据交流与交易的开放平台。中国互联网协会大数据工作委员会（筹）每周三晚8点30分开始，在“中国大数据千人会”微信群进行交流与探讨。在具体运营方式上区隔于传统的专家与精英为核心，采用众筹方式，自组织互动。目前已经运营半年多，极大地体现了工程中心与产业融合发展协同创新的定位。

4.运行管理与机制创新

“工程技术研究中心”设主任1名，副主任3名，在北京市科委与北京理工大学的共同领导下，实行主任负责制。工程技术研究中心下设技术委员会、研究开发部、项目开发部、产品开发部、市场开发部、管理办公室等部门。

按照《北京高等学校工程研究中心建设项目管理办法》的规定，制定适合本工程中心的管理机制，具体包括：

1）按工作性质设立若干岗位，对每个岗位规定相应的职责，实行行政、学术、项目三维管理体制。

2）实行周工作报告、月工作汇报、周工作例会、月工作例会、不定期技术报告会的课题日常管理制度。

3）鼓励研究人员积极申请课题，课题的申请、实施和管理服从统一规划，优先保证重大课题的组织实施。按课题的性质和来源不同，实施不同的质量管理体系。对军工课题实施国军标GJB9001A-2001质量管理体系，重大课题实施ISO9001质量管理体系。

4）所有资源均指定专人管理，包括设备、工位、图书、软件、办公用品、文档、研制用元器件、IP地址和Email地址等，研究人员须遵照相应的规定。使用内部信息管理系统辅助日常管理。

5）工程技术研究中心人员须认真遵守工作纪律，以保证科研工作的开展和协同。

6）实行年初规划和年终逐级考核制度，考核结果作为奖金发放、评选先进、薪酬调整、职务升降、岗位调整、不合格淘汰以及上报或存档用的各种总结表的重要依据。

工程中心严格执行各项管理制度，保障教师和员工有序开展科学研究和科技转化，学术论文、专利、软件著作权的数量和质量得到大幅提升。

5.依托单位支持

依托单位针对工程中心发展的实际需求，制定了专门的政策支持措施。在人才引进方面，给予工程中心大力支持，优先改善引进人才的工作和生活条件；依托单位对工程中心给予专项建设经费，用于支持工程中心在国际交流合作、学术论文发表与专利申请、专家咨询等方面的费用支出；在后勤保障方面，依托单位给予工程中心提供工作场所等配套支持。**三、工程技术研究中心自评表**

|  |  |  |
| --- | --- | --- |
| 评价内容 | | 自评分 |
| 发展规划与目标完成 (10分) | 认定时规划目标完成情况 | **9** |
| 未来三年发展规划 |
| 技术水平与成果转化  (45分) | 定位与研究方向情况 | **43** |
| 技术成果水平 |
| 成果转化与市场结合能力 |
| 技术创新的贡献度 |
| 队伍建设与人才培养  (25分) | 工程中心主任与工程技术带头人作用 | **23** |
| 队伍结构与创新团队建设 |
| 青年骨干人才培养 |
| 开放交流与运行管理  (20分) | 技术委员会作用 | **17** |
| 开放交流 |
| 协同创新 |
| 运行管理与机制创新 |
| 依托单位支持 |
| 总分 | | **92** |

**四、依托单位内部公示情况**

|  |
| --- |
| （说明相关材料在依托单位公示的时间、地点，依托单位内部对于公示的反馈情况。）  依托单位（盖章）：  年月日 |

**五、技术委员会意见**

|  |
| --- |
| 技术委员会主任（签字）：  年月日 |

**六、依托单位意见**

|  |
| --- |
| 依托单位（盖章）：  年月日 |

七、附件目录

| **序号** | **附件名称** |
| --- | --- |
| 1 | 技术成果情况明细表 |
| 2 | 队伍建设情况明细表 |
| 3 | 技术委员会召开情况表 |
| 4 | 开放交流情况明细表 |
| 5 | 依托单位支持情况表 |
| 6 | 绩效报告公示照片 |