

当前位置: 科技频道首页 >> 军民两用 >> 计算机与网络 >> 高性能东方文字文档智能全信息数字化系统

请输入查询关键词

科技频道

搜索

高性能东方文字文档智能全信息数字化系统

关键词: [文字识别](#) [文档重构](#) [东方文字文档数字化系](#) [数字化技术](#)

所属年份: 2003

成果类型: 应用技术

所处阶段:

成果体现形式:

知识产权形式:

项目合作方式:

成果完成单位: 清华大学

成果摘要:

在全球信息化时代, 文档数字化是急需解决的战略性难题。清华大学最新研究成功的“高性能东方文字文档智能全信息数字化系统”是集”高性能中日韩文档识别理解重构系统”和”基于识别的原文重现自动电子出版物制作系统”为一体的、解决多文种(中、日、韩、英、藏)原文重现文档全信息数字化规模化生产, 及低质、手写文档和表格识别数据录入的智能信息化工具, 整体性能居于国际领先水平。1. 国际领先的高性能中日韩文档识别理解重构系统; 2. 鉴定意见: “系统首次在一个统一的框架下实现了印刷体简繁汉英、日英、韩英文字的识别和文档重构, 其整体性能达到了国际领先水平。”在文字识别核心技术方面有诸多创新, 极大提高了识别的鉴别能力和鲁棒性: 识别字符集空前扩大(简/繁体字6763/13053, 日文6524、韩文7238), 覆盖了中日韩基本字符集; 各种字体的平均识别率在99.5%以上; 脱机手写汉字识别率达98.7%、自由书写达92%。经Scansoft公司评测, 即使不擅长的日文、韩文的识别性能都超过目前国际上相应的系统。在国际竞标中夺标, 授权给微软公司使用。首创Gabor变换文字识别算法, 有效提高了输入图像质量低情况下的文字识别。首创藏文识别系统, 实际藏文识别率95%, 已在藏民区试用。3. 提出文档全信息数字化, 研发成功自动全信息数字化系统。鉴定意见“是国内外首次推出的, 将复杂报纸文档经版面分析、识别、理解、最后自动精确重构为原式原样的标准格式电子出版物等功能, 为中国信息资源建设给出了一个高质、高效、规模化生产的工具, 是一个有意义的创举。作为一个整体居于国际领先水平。”在中国电子出版、数字图书馆工程中发挥重要作用。4. 金税工程增值税发票识别防伪系统在全国推广, 极大打击了偷漏说, 保证金税工程顺利实施; 银行单据识别等行业应用解决方案, 为这些行业的信息化建设创造了条件。成果已产品化, 广泛应用于办公自动化、银行、税务、公安各行业。THOCR2000、名片和身份证识别、邮政分拣等软件, 销往海内外。已创收100万美元外汇和1亿2千万元效益。

成果完成人: 丁晓青;刘长松;吴佑寿;陈明;彭良瑞;方驰;张嘉勇;文迪;郭繁夏;郑冶枫

[完整信息](#)

行业资讯

- [新疆综合信息服务平台](#)
- [准噶尔盆地天然气勘探目标评价](#)
- [维哈柯俄多文种操作系统FOR ...](#)
- [社会保险信息管理系统](#)
- [塔里木石油勘探开发指挥部广...](#)
- [四合一多功能信息管理卡MISA...](#)
- [数字键盘中文输入技术的研究](#)
- [软开关高效无声计算机电源](#)
- [邮政报刊发行订销业务计算机...](#)
- [新疆主要农作物与牧草生长发...](#)

成果交流

推荐成果

- [液压负载模拟器](#) 04-23
- [新一代空中交通服务平台、关...](#) 04-23
- [Adhoc网络中的QoS保证\(Wirel...](#) 04-23
- [电信增值网业务创意的构思与开发](#) 04-23
- [飞腾V基本图形库的研究与开发...](#) 04-23
- [ChinaNet国际\(国内\)互联的策...](#) 04-23
- [电信企业客户关系管理\(CRM\)系...](#) 04-23
- [“易点通”餐饮管理系统YDT2003](#) 04-23

Google提供的广告

>> 信息发布

[版权声明](#) | [关于我们](#) | [客户服务](#) | [联系我们](#) | [加盟合作](#) | [友情链接](#) | [站内导航](#) | [常见问题](#)
国家科技成果网

京ICP备07013945号