

晨鸟论文剽窃检查系统产品介绍

论文剽窃检查系统（CHENNIAO/PD）是一款全新的网络知识服务系统。该系统利用多年积累的论文诚信判别知识和自主研发的基于互联网文献检索的多维语义知识库和推理技术，依据互联网文献资源以及专业文献库，对科研作品的诚信度实施“精”、“准”、“全”的判别，所形成的详尽报告是甄别科研诚信不佳行为的重要依据。

论文剽窃检查系统（CHENNIAO/PD）是国内唯一支持机房落地安装、单位内网专用的诚信排查系统，并提供“授权帐号”模式下所无法实现的各种定制集成专业服务解决方案，有效保障系统投入运行和效益产出过程的高效平稳。

论文剽窃检查系统（CHENNIAO/PD）凭借其“基于互联网文献检索的多维语义知识库”和“排查服务多重集群与负载均衡技术”的两项专有核心技术，先后获得国家专利、上海市优秀发明奖和科技部中小企业创新基金立项。

一、产品型号

论文剽窃检查一体化设备版

论文剽窃检查系统与高性能服务器的完美一体化组合，经专业性能调优，全面发挥论文剽窃检查引擎的高效性能。实现科研管理平台“从无到有”，“一触即发”。

论文剽窃检查标准/专业版

按需定制论文剽窃检查系统的应用领域，选择与用户专业方向相一致的最佳配置组合。

论文剽窃检查Web服务版

面向服务，将论文剽窃检查的强大功能与第三方信息系统无缝整合，将论文剽窃检查提升轻松融入现行科研管理工作流程。

二、功能描述

论文剽窃检查系统（CHENNIAO/PD）是由基于互联网文献检索的多维语义知识库、论文剽窃检查指标体系、科研作品诚信管理系统和用户附加比对库四个部分组成。与同类产品相比，论文剽窃检查系统（CHENNIAO/PD）不仅是成熟的科研作品诚信判别和监测的专业知识服务系统，也是科研诚信提升和信息化管理的完整解决方案。专业的解决方案中囊括了科研诚信管理系统定制、第三方信息系统无缝整合、用户附加比对库导入、硬件设备一体化构建等一系列技术服务支持。突出的系统扩展性和服务成熟度为用户带来最小的附加投入，最大的增值效益。

核心技术与核心资源

晨鸟科技——网络知识服务专家

服务热线：400-820-3886 服务邮箱：service@chenniao.com

上海总公司：上海市密云路 1018 号复旦大学国家科技园 6 号楼 104（200433）

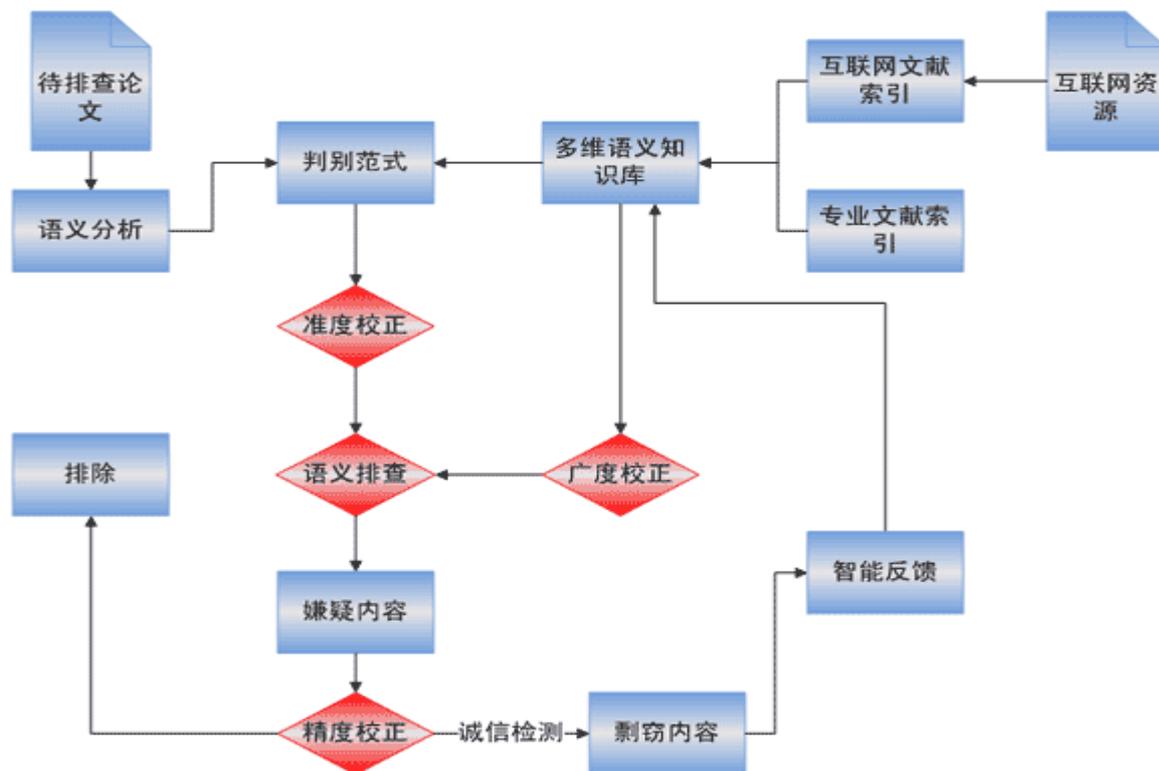
无锡分公司：江苏省无锡市高浪东路 999 号太湖科技中心研发大楼 505（214131）

北京分公司：北京市西直门北大街 32 号枫蓝国际写字楼 D 座 1702（100082）

论文剽窃检查系统（CHENNIAO/PD）采用“基于互联网文献检索的多维语义知识库”与“排查服务多重集群与负载均衡技术”两项核心技术以保障诚信排查“精”、“准”、“全”和软硬件最优化状态稳定运行。论文剽窃检查系统（CHENNIAO/PD）应用“基于互联网文献检索的多维语义知识库”的核心技术，将互联网海量文献资源进行检索、缓存和索引，并有机融合一般专业文献资源，形成系统核心资源。经深入研究发现，国内多数科研作品创作的不诚信行为多表现为从互联网抄袭、剽窃有关文献资源，类似情况占总体不诚信行为的60%以上。海量互联网文献资源的应用价值，是一般文献出版类资源库无法比拟的。论文剽窃检查系统（CHENNIAO/PD）在互联网文献资源检索技术上所取得的成绩，无疑使我国文献诚信判别查全度领先于国际水平。



论文剽窃检查业务流程



论文剽窃检查指标体系

- ✦ 重合字数 重合字数分为总重合字数和子重合字数，总重合字数是指被检查论文抄袭的字数之和，可以反映被检查论文的剽窃程度，子重合字数是指被检查论文抄袭某篇文献的字数，可以反映被检查论文抄袭该文献的程度。
- ✦ 总抄袭字数 总抄袭字数是指被系统判断为抄袭内容的字数，抄袭内容包括重合内容，也包括不重合但相似的内容。
- ✦ 总字数 总文字数是指该检测论文所有包含的字数。
- ✦ 总文字复制比 总文字复制比是总重合字数和论文总字数的比例，如果重合字数较少，但总文字复制比较大，说明该论文剽窃情况也比较严重。
- ✦ 总抄袭比例 总抄袭比例是指总抄袭字数和论文总字数的比例。
- ✦ 相似度 相似度是指某段剽窃内容和被剽窃的原文之间的接近程度。
- ✦ 首部重合文字数 首部重合文字数指学位论文前 20%文字中重合的文字数量。对于论文，一般开头部分均是综述性的报告介绍，其重要性远低于论文尾部。
- ✦ 尾部重合文字数 尾部重合文字数是指剩下部分中重合的文字数量。

- ❖ 抄袭文献数 抄袭文献数是指被检查论文共抄袭多少篇文献。
- ❖ 抄袭处数 抄袭处数分为总抄袭处数和子抄袭处数，总抄袭处数是指被检查论文共有多少处剽窃，子抄袭处数是指被检查论文抄袭某篇文献的处数。
- ❖ 最大抄袭段长 连续抄袭内容作为一个抄袭段，最大抄袭短长是指连续抄袭字数最多的抄袭内容长度
- ❖ 平均抄袭段长 连续抄袭内容作为一个抄袭段，平均抄袭段长是指所有抄袭段的平均抄袭字数。

三、核心技术

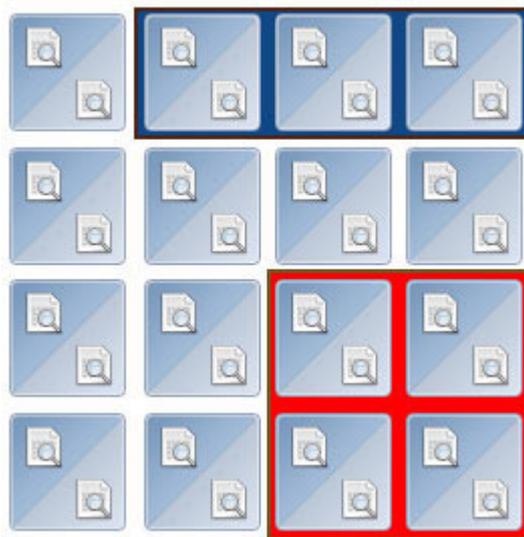
基于互联网文献检索的多维语义知识库

论文剽窃检查系统（CHENNIAO/PD）采用了具有自主知识产权的“基于互联网文献检索的多维语义知识库”来保障文献诚信判别的“精”、“准”、“全”。该多维语义知识库历时 4 年、调研访问数百位科技查新专业人士和评审专家，融合专家学者多年来积累的判别知识，提炼出采用人工智能技术模拟的判别范式。由此形成的判别结果，与同类产品相比，在文献比对时不局限于关键词和段落而专注于文献核心贡献，在诚信判别时超越抄袭嫌疑而是给出精确的剽窃判断，并不受算法干扰行为的影响。

经深入研究发现，国内多数科研作品创作的不诚信行为多表现为从互联网抄袭、剽窃有关文献资源，类似情况占总体不诚信行为的 60%以上。论文剽窃检查系统（CHENNIAO/PD）的多维语义知识库领先融合互联网文献元搜索技术，对互联网海量文献资源进行检索、缓存和索引，形成基于互联网文献检索的多维语义知识库。海量互联网文献资源的应用价值，是一般文献出版类资源库无法比拟的。基于互联网文献检索的多维语义知识库无疑在文献诚信判别查全度上填补了一项技术空白。

排查服务多重集群与负载均衡技术

论文剽窃检查系统（CHENNIAO/PD）是国内唯一支持机房落地安装、单位内网专用的诚信排查系统，规避了“授权帐号”模式下无法提供定制化技术服务的弊端，有效保障设备资产的投入和利用度。由于诚信排查系统在运行过程中会对所承载的服务器内存、CPU 性能消耗较大，论文剽窃检查系统（CHENNIAO/PD）为适应普通服务器的硬件性能，发挥其最佳运行效能，而自主研发了排查服务多重集群与负载均衡技术。该技术针对排查服务进行智能集群和调度，在一台服务器上实现多台服务器的集群与负载均衡效果，有



效分配服务器内存和 CPU 资源，以最优化的性能来运行多线程并发的排查服务。每一个排查服务在并发运行过程中会占用 3-5 个服务器集群来进行负载均衡，确保普通服务器上能够顺利运行多个并发排查服务，不会因为资源抢占而意外中止或死锁。

四、产品特色

基于互联网文献检索的核心技术与核心资源

论文剽窃检查系统（CHENNIAO/PD）应用“基于互联网文献检索的多维语义知识库”的核心技术，将互联网海量文献资源进行检索、缓存和索引，并有机融合一般专业文献资源，形成其核心资源。基于互联网文献检索是论文剽窃检查系统（CHENNIAO/PD）保障文献诚信判别“精”、“准”、“全”的核心。基于复旦大学的一项深入研究发现，国内多数科研作品创作的不诚信行为多表现为从互联网抄袭、剽窃有关文献资源，类似情况占总体不诚信行为的 60% 以上。因此海量互联网文献资源的应用价值，是一般文献出版类资源库所无法比拟的。论文剽窃检查系统（CHENNIAO/PD）基于互联网文献检索技术的创新，无疑使我国文献诚信判别查全度领先于国际水平。

基于大量专家判别知识积累的人工智能模拟推理

论文剽窃检查系统（CHENNIAO/PD）所采用的诚信判别技术是“晨鸟科技”与复旦大学历时 4 年、调研访问数百位科技查新专业人士和评审专家，从而融合专家学者多年来积累的判别知识，提炼出采用人工智能技术模拟的判别范式。由此形成的判别结果，与同类产品相比，在文献比对时不局限于关键词和段落而专注于文献核心贡献，在诚信判别时超越抄袭嫌疑而是给出精确的剽窃判断，并不受算法干扰行为的影响。

基于集成接口与集群调度优化的成熟化产品解决方案

论文剽窃检查系统（CHENNIAO/PD）创新地应用 Web 服务标准接口和“排查服务多重集群与负载均衡技术”，全面保障了产品的灵活性、稳定性、扩展性、安全性和高效性，并将其从成熟的科研作品诚信判别和监测的专业知识服务系统提升为科研诚信提升和信息化管理的完整解决方案。摒弃“帐号授权”模式的呆板与距离感，以真挚、专业的服务为用户带来最小的附加投入，最大的增值效益。

五、所获荣誉

 2006 年 11 月，晨鸟专利——“一种利用网络资源实现剽窃和格式检查的在线论文管理方法”（申请号：200610118150.8）获国家知识产权局受理。

 2007 年 12 月，“支持剽窃和格式检查的晨鸟在线协同论文管理系统”荣获上海市优秀科技发明三等奖。

-  2008年6月,“以智能论文剽窃检查技术为特色的全流程教学协同平台”获国家创新基金支持。
-  2008年10月,晨鸟专利——“一种支持格式比对和剽窃检查的文档编辑器原理与方法”(申请号:200810207552.4)获国家知识产权局受理。