基于分布式及增量式爬虫技术与 SSM 框架构建无人机知识库系统

一. 成果完成人及所在单位：

陈秋瑾，王晓荣，马威威，戴加权　　电气工程与自动化学院　　指导教师：何原荣

二. 成果简介：

本系统以互联网分布式及增量式爬虫技术为外部数据获取技术支撑，结合知识管理理念构建知识树体系，收集来自国内外无人机行业领域相关知识文献、新闻讯息存储于本地数据库中，整合实时监测台风路径图便于系统监控预测台风。此外，优化系统框架为功能丰富强大的 SSM 框架集，在对爬取数据提取、分析、去重后，进一步实现自动提取文章摘要信息及语音播放摘要功能；在利用 Echarts 组件及知识图谱等大数据相关算法分析有效数据后，以图表形式展示行业研究趋势，以及通过关键字云图展示行业研究热点及机构合作情况等信息。本系统通过强大的云计算、大数据和分布式存储技术实现研究所的内容的采集和归档，通过严谨的细颗粒度权限管控保护文件的安全应用，通过灵活的元数据及编目模板技术让用户更加方便、高效的进行知识共享、协同工作以及互动交流，已投入我校计算机信息工程学院的数字福建自然灾害监测大数据研究所使用，具有一定的研究价值和实用价值。

三. 主要研究内容及创新点：

本系统为专属定制化的无人机知识库专家系统，将隐形知识转为机构知识，提高知识的利用效率，方便科研人员进行学术研究，实现大数据技术应用在无人机行业领域上的研究和知识积累。本系统的两种平台模式（前端、后台），分别针对用户（教授、老师、学生等科研学者）和管理员使用，前端可以看作一个关于无人机行业专业知识的搜索引擎和知识库，后台可以看作为机构内（高校、研究所、企业）的一个知识库管理及资源监控系统。**主要研究内容有 5 点**：（1）知识获取及存储模块：采用分布式及增量式相结合爬虫技术获取互联网知识，使用 Redis 数据库结合 MySQL 数据库高效且稳定存储数据；（2）知识智能管理模块：通过知识采集、知识树知识图谱模型和提取技术实现知识的挖掘，达到知识智能利用；建立了包含知识库、搜索引擎、知识激励、知识社区、专家系统、个人知识管理等系统管理模块；（3）知识树形式展示知识模块：知识树形式展示数据有利于知识传递，关键字是知识获取的钥匙，通过知识节点和关键字关联，方便知识的传播和理解；（4）台风实况模块：台风信息来源于全球国际台风监控网站的详细数据和通过无人机采集台风过后的灾损情况的正射影像图；（5）数据可视化分析及知识推理模块：对爬取到的数据利用 Echarts 组件进行分析以图表形式可视化展示，通过知识图谱等算法分析出爬取数据的关键字云图、机构同现及作者同现图，以此研究该行业领域的研究趋势、机构/个人产出等情况。**本研究的创新点有 4 点**：（1）在原来分布式爬取的基础上结合增量式结合爬取的技术，优化系统原有 MVC 框架为功能更丰富强大的 SSM 框架，（2）使用结构化语义分析处理机制，进行数据的预处理以提取有用信息，使用成熟的多来源的自动化翻译接口翻译整理来自不同国家的相关的研究资料，最终将去重、提取、分析、结构化处理后的数据以知识树形式展示在系统中；（3）使用协同工作机制及知识管理理念，将爬取到的新闻及文献数据通过大数据、知识图谱等相关数据分析算法可视化分析专业领域内的研究现状；（4）使用自动文摘生成技术，将爬取下的新闻及文献等文章中关键信息提取出来形成摘要信息，并且实现语音播放摘要的功能。如图 1、2 所示分别为系统前端的新闻资讯模块、专业文献模块，图 3、4 所示分别为系统前端数据可视化模块、后台的知识图谱模块。



图 1 前端新闻资讯模块



图 2 前端专业文献模块

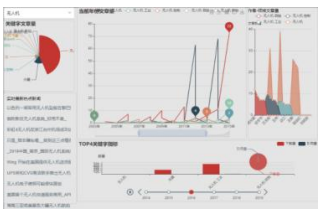


图 3 数据可视化分析模块

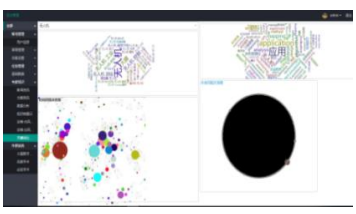


图 4 后台知识图谱模块

四. 主要研究成果及应用情况

1.投稿参加第九届赣桂闽遥感科技论坛（收录论文集）、第十届桂闽赣遥感科技论坛（评获一等奖）。

2.投稿参加 2019 年 ICGBD 大数据测绘地理信息科学国际研讨会（EI 会议论文已录用），投稿 CEESD 2019 第四届环境工程和可持续发展国际会议（EI 会议论文审稿中）。

联系人：陈秋瑾　　电　话：18506927528　　Email:892725709@qq.com