“基于Spark平台的人工智能知识的知识图谱构建”

**项目摘要**

**队员名单：**文华、周余、刘宏鑫、陈叶红

**院校：**合肥工业大学宣城校区 计算机与信息系

**摘要：**随着计算机大数据的快速发展,可以借助于互联网平台的各种工具找到有价值内容,但海量数据給筛选、组织与评价带来极大困难。项目拟利用分布式爬虫实现PathFinder 算法获得学科信息并借助 Spark 平台过滤无效信息,再借助 Spark 优化基于CNN的文本特征选择算法完成学科专有名词分类,同时采用 Bayes 统计推断完成关键词与相关信息的联想,完成内容的可视化。项目意义在于通过 Spark 完成人工智能知识的重整,实现了一个学习者尤其是本科生适用的知识图谱工具。

**关键词：**知识图谱、人工智能、大数据、Spark平台

**作品优点**

1. **Spark结合CNN构建人工智能知识图谱；**

**技术创新：**Spark平台广泛运用于大数据处理，而鲜有人将其运用于知识图谱的构建，这是我们的独立思考出的创新路径。

1. **实现PathFinder算法限制阈值爬取数据；**

**技术创新：**我们创新性地将PathFinder算法用于数据的爬取，过滤大量无用的信息，只爬取阈值高的数据。。

1. **实现了非传统思维导图的可视化效果。**

**模式创新：**我们运用D3，amcharts等一系列JS库及前端文档实现不同种类数据的多样可视化。传统的知识只具有简单的表格，清单等，有着天然的枯燥，关联性不强等缺点，本项目多样化的可视化图谱显然，其直观形象、易于理解。