基于Spark的人工智能知识的知识图谱构建

摘要：

关键词：知识图谱；Spark；可视化

Abstract:

Keywords:

0 引言

人工智能是当前最热门研究领域之一，甚至被誉作世界三大尖端技术之一[1]，近年来我国甚至将其上升到国家战略的高度：2017、2018与2019年的政府工作报告中均被提及[2]。可见，人工智能在现代科学技术与经济社会中有着不可替代的地位，随着5G时代的到来，人工智能必将展现更广阔的应用前景。与此同时，人工智能相关方向的人才匮乏也正越来越成为（市场）关注的议题[3]，而在培养人才时，如何准确把握所授相关领域知识的准确性、全面性与前沿性成了一个难题，知识图谱是解决这一难题的有效工具。知识图谱是人工智能领域重要的一个技术分支，其目的是将现有的人类知识构建为一个结构化的知识库。目前，已经有许多大型知识图谱被构建出来，如DBpedia 、Freebase等，然而，当前的知识图谱工具普遍存在以下问题：1）通用知识图谱工具涉面较广，但知识冗余混乱、组织零散、系统性差，不利于用户的专业学习；2）垂直知识图谱工具种类少，成熟的应用仅限于某些领域，在一些具有较大应用需求的领域未获重视，前景广阔[4]。

综上所述，本文的目的是构建一个面向学习者尤其是本科生的人工智能领域的垂直知识图谱。人工智能领域繁多，我们选取机器学习、自然语言处理与机器视觉等三个领域作为代表。

1 相关工作

知识图谱的构建技术仍在持续发展中，目前存在多种流派，每一种技术手段途径各异、效果良莠不一。随着相关技术的不断演变与发展，新的知识图谱构建方法被不断推出，有些研究也在尝试使用经典的方法在新的应用领域构建相应的垂直知识图谱。金婧等[6]侧重于知识图谱表示学习方法，在TransE[5]模型的基础上提出了一种融合实体类别信息的知识表示学习模型（TEKRL）；杨玉基等[7]在对领域知识图谱的系统研究上，提出了一种构建领域知识图谱的“四步法”；孙昊天等[8]实现一种基于带权三元组构建时政类知识图谱的方法；董永强等[9]提出了一种基于YANG[10]模型由数据模型驱动的网络领域知识图谱构建方法。而在通过经典方法构建垂直知识图谱上，熊晶等[11]

2 数据来源

3 Spark与Jiagu模型

4 结果与分析

5 结束语