

天津大学

本科生毕业设计（论文）开题报告



题目：美国科学研究系统建模及合作模式挖掘

学 院	管理与经济学部
专 业	信息管理与信息系统
年 级	2019 级
姓 名	蒋世华
学 号	3019209018
指导教师	王文俊

一、课题的来源及意义

本课题来源为天津大学新工科毕设项目，将利用知识图谱等相关技术完成预期研究内容。

知识图谱作为一种用图模型来描述知识和对事物之间关联关系进行建模的技术方法，通常由节点和边组成。节点通常是实体，边通常为实体的属性或实体间的关联关系。知识图谱从诞生开始，已经广泛应用于大数据分析、推荐计算、辅助智能问答等领域^[1]。近年来，知识图谱技术在医疗行业、烟草行业、科研数据方面等均已有了相关应用^{[2][3]}。本课题拟根据美国国立卫生研究院（NIH）、美国国家自然科学基金会（NSF）、web of science 论文获取美国科研机构项目发布流程数据。通过构建美国科研体系知识图谱实现针对美国科研系统建模并进一步分析其合作模式，方便相关管理部门进行科研机构并能够提供专业人员的详细参考资料，进而大大降低管理成本，提高相关部门管理效率^{[4][5]}。

针对美国科研机构项目发布流程数据构建美国科研机构知识图谱。为此，课题将基于 Media Wiki 等开发工具构建一种美国科学研究系统建模与合作模式挖掘知识图谱的可视化系统，解决在该领域的迫切需求，具有现实意义和应用推广价值。

二、国内外发展状况

知识图谱的发展可以追溯到上世纪六十年代的语义网络研究，而自 Google 公司于 2012 年正式提出知识图谱概念后，经过近十年的发展，知识图谱已经在各行业、各领域产生了深远的影响。在这其中，对于知识图谱的技术与应用方面的研究也在不断深入。而面对冗杂的科研流程，越来越多的专家学者也将知识图谱技术应用于管理科研机构领域。

在国外，Ganggao (2018) ^[6]等为应对知识图谱中实体在自然语言处理中存在歧义的问题，创建了基于实体的上下文词和信息词之间语义相似度的消歧方法，提高了各个实体之间的独立性，完善了知识图谱的技术体系。Danae Pla (2016) ^[7]等则将目光集中于知识图谱的应用层面，该研究利用知识图谱技术通过从推特文章中检索语义信息，计算用户之间的兴趣相似度，成功降低了推特过度推荐和过度专门化等问题。而对于科研方面的知识图谱应用，Liu W (2018) ^[8]等基于海量科研数据构建知识图网络的方法，以从每个课题设置文档的标题信息中提取主题为关键技术，以及课题的方向，对项目库进行分析，提取项目的基本信息，并对每个课题的文档信息和课题方向进行分析。构建属于同一字段的主题方向的知识图谱。SONG S (2023) ^[9]等利用技术要素和专家知识图操作生成

专家知识图图像。基于科技大数据知识图谱的智能技术诊断专家匹配算法，具有一套为开发技术评估、商业计划、决策咨询、前沿分析和市场预测提供指导的增值服务。

在国内，杨思洛（2012）^[10]等利用可视化方法总结了知识图谱的研究现状，重点对比分析了知识图谱研究的核心作者与所有作者的合作网络，宏观与微观机构合作网络通过高频关键词的共现分析研究知识图谱的发展趋势。李思志（2014）^[11]等则从管理科学与工程宏观的创新轨迹入手，提出利用引文共引分析、文本挖掘、特征抽取的方法，基于图谱分析平台对创新趋势进行分析。雷洁（2020）^[12]等则将知识图谱技术应用于科研档案领域的研究当中。该团队通过构建基于知识图谱的科研档案管理模型，从知识层面将科研档案资源中的科研机构、科研项目、科研成果、人员等要素与项目任务书、合同、研究报告中抽取的知识单元进行关联和融合，丰富科研档案的语义关系，推进科研档案管理系统提档升级。杜悦（2023）^[13]等提出的隐式知识图谱构建方法很好地解决了由于实体信息变动引发的数据一致性问题，适用于大规模科研知识图谱的构建，有助于科技知识的高效管理和传播利用。

综上所述，当前国内外知识图谱研究大体分为技术研究与实际应用两大方向。技术研究主要集中于数据挖掘、实体识别、关系抽取等方向；实际应用则主要集中于智能问答、辅助决策等方面。针对美国科研机构项目的知识图谱研究^{[14][15][16]}，由于时间积累不足，主要集中在其科研档案或文献数据方面^{[17][18][19]}，对于科研机构项目的研究较少，且并未将其作为研究主体。因此本研究将美国科研机构项目作为研究主体符合当前科研机构项目管理的迫切需要。

三、研究目标、研究内容与研究方法

1.研究目标

本文的工作目标是借助知识图谱技术，基于美国科研机构项目的开源大数据，通过对美国科研机构项目信息的语义网络分析，结合对美国科研机构项目的实体与关系抽取，构建美国科研机构项目知识图谱，并以 Media Wiki、SMV 等开源工具为基础搭建一个美国科研机构项目流程图谱的可视化管理系统。在对该知识图谱进行管理和可视化分析的基础上，本文拟采用图神经网络实现对于知识图谱数据的构建^[20]，并通过美国科研机构项目流程数据挖掘之间的关联性，进而满足对科研机构项目的管理需求。

2.研究内容

(1) 知识图谱相关概念调研：包括知识图谱概述、知识抽取与知识挖掘、知识表示与建模、知识图谱融合与推理。语义网络的结构方式、特点、自然语言处理的基本用法。

(2) 国内外研究现状分析：包括国内外知识图谱的技术研究、应用研究以及针对美国科研机构的知识图谱研究现状；

(3) 美国科研机构项目知识图谱的数据获取：基于(1)(2)当中的理论知识，完成前期科研项目数据搜集和网络爬取，形成数据库；

(4) 美国科研机构项目知识图谱的设计与实现：借助语义网络分析、知识抽取与挖掘、知识表示与建模，将美国科研机构项目相关数据处理为结构化的 RDF 文档，构建出美国科研机构项目知识图谱框架。

(5) 美国科研机构项目知识图谱的可视化：通过 Media Wiki、Semantic-Media Wiki 等知识图谱开源工具搭建美国科研机构项目知识图谱信息系统，编程实现知识图谱的管理和可视化搜索。

(6) 结论与展望：论文最后对全文的研究工作进行总结，对未来的研究内容和方向进行展望。

3.研究方法

(1) 文献搜集法

通过搜集、整理知识图谱领域相关文献资料，归纳总结知识图谱研究现状，并总结整理知识图谱构建方法。

(2) 系统分析与设计

借助知识图谱技术，基于美国科研机构项目的开源大数据，通过对美国科研机构项目的语义网络分析、结合针对美国科研项目的实体与关系抽取，利用 SMW 构建美国科研机构项目知识图谱的可视化系统。

四、进度安排

2022.12.21—2022.11.28 论文选题，查阅资料

2022.11.28—2022.12.26 参照任务书撰写完成开题报告，并完成文献翻译

2022.12.26—2023.01.14 参照开题报告撰写并确定论文提纲

2023.01.15—2023.03.06 文献阅读、理论学习，掌握 SMW 等工具使用方法。

2023.03.07—2023.04.17 理论模型研究、图谱设计与实现

2023.04.18—2023.05.15 论文初稿撰写与修改

2023.05.16—2023.05.31 论文终稿撰写与修改

2023.06.01—2023.06.10 论文定稿打印，准备答辩

五、研究或实验方案的可行性分析

1、经济可行性

本知识图谱的设计与开发基于 Media Wiki 开源环境，将先在个人 PC 机调试，之后上传网络，具有经济可行性。

2.技术可行性

学生具备一定的图数据库编写能力，文本语义网络分析能力，python 等编程语言的编写能力，具有技术可行性。

3.社会可行性

美国科研机构项目知识图谱的可视化系统能够为相关科研机构及管理部门提供清晰、准确的参考，降低科研机构运行及管理成本，具有较高的应用推广价值，具备社会可行性。

六、主要参考文献

[1]Shen, Y., et al. Knowledge graph constructing, searching and visualization method, involves finishing query in knowledge graph presenting user search content in visual pattern form in responding to personalized retrieval requirement of user, and improving user search experience, Univ Shanghai (Ushn-C).

[2]Wang, W., et al. (2021). "Key technologies of the service platform of scientific research knowledge graph in tobacco field." Acta Tabacaria Sinica 27(4): 83-91.

[3]Liu, Y., et al. Medical use-based knowledge graph system, has medical application layer for providing service to user based on knowledge graph, where medical application layer comprises decision assistance module that provides decision assistance scheme for user, Univ Northeastern (Unen-C).

[4]李志民.美国科研机构概览[J].世界教育信息,2018,31(05):6-10.

[5]育东.美国科研机构的运作与管理[J].全球科技经济瞭望,1995(04):5-8.

[6]Zhu G , Iglesias C A . Exploiting Semantic Similarity for Named Entity Disambiguation in Knowledge Graphs[J]. Expert Systems with Applications, 2018, 101(JUL.):8-24.

[7]Danae Pla Karidi,Yannis Stavrakas,Yannis Vassiliou. Tweet and followee personalized recommendations based on knowledge graphs[J]. Journal of Ambient Intelligence and Humanized Computing,2018,9(6):2035-2049.

[8]Liu, W., et al. Method of constructing knowledge graph network based on massive scientific research data, involves extracting subject word from title

<p>information of setting document of each topic as key technology, and clustering direction of topics, State Computer Network & Information Saf (Stcn-C).</p> <p>[9]Song, S., et al. Science and technology big data knowledge graph based intelligent technical diagnosis expert matching algorithm, has set of instructions for developing technology evaluation, business plan, decision consultation, front analysis and market prediction value-added service, BEIJING WANFANG SOFTWARE CO LTD (BEIJ-Non-standard).</p> <p>[10]杨思洛,韩瑞珍.知识图谱研究现状及趋势的可视化分析[J].情报资料工作,2012(04):22-28.</p> <p>[11]李思志,李佳骏,李艳红.管理科学与工程领域的创新轨迹研究——基于 TOP 期刊的文献计量和文本挖掘视角[J].中国管理科学,2014,22(S1):56-62</p> <p>[12]雷洁,赵瑞雪,李思经等.知识图谱驱动的科研档案大数据管理系统构建研究[J].数字图书馆论坛,2020,No.189(02):19-27.</p> <p>[13]杜悦,常志军,董美,钱力,王颖.一种面向海量科技文献数据的大规模知识图谱构建方法[J/OL].数据分析与知识发现:1-14[2023-01-19].</p> <p>[14]廖帅.高校 SRM 科研管理系统建设与协同平台研究——基于 CiteSpace 知识图谱软件的计量分析[J].现代信息科技,2019,3(10):172-175+178.</p> <p>[15]杜悦,常志军,董美等.一种面向海量科技文献数据的大规模知识图谱构建方法[J].数据分析与知识发现,2023,7(02):141-150.</p> <p>[16]王洪旭,温晓会,刘万明.基于知识图谱的高校科研经费管理研究[J].行政事业资产与财务,2022(05):108-110.</p> <p>[17]H. Cai, Z. Liu and C. Wang, "Intelligent recommendation system based on knowledge graph for scientific research teams," 2021 13th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), Hangzhou, China, 2021.</p> <p>[18]Huang, S. S. and X. J. Wan (2013). AKMiner: Domain-Specific Knowledge Graph Mining from Academic Literatures. 14th International Conference on Web Information Systems Engineering (WISE), Nanjing, PEOPLES R CHINA.</p> <p>[19]雷洁,赵瑞雪,李思经,鲜国建,寇远涛.知识图谱驱动的科研档案大数据管理系统构建研究[J].数字图书馆论坛,2020,No.189(02):19-27.</p> <p>[20]Z. Ye, Y. J. Kumar, G. O. Sing, F. Song and J. Wang, "A Comprehensive Survey of Graph Neural Networks for Knowledge Graphs," in IEEE Access, vol. 10.</p>	
<p>选题是否合适: 是<input type="checkbox"/> 否<input type="checkbox"/></p> <p>课题能否实现: 能<input type="checkbox"/> 不能<input type="checkbox"/></p>	<p>指导教师 (签字)</p> <p>年 月 日</p>
<p>选题是否合适: 是<input type="checkbox"/> 否<input type="checkbox"/></p> <p>课题能否实现: 能<input type="checkbox"/> 不能<input type="checkbox"/></p>	<p>审题小组组长 (签字)</p> <p>年 月 日</p>

