

天津大学

本科生毕业设计（论文）任务书



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、原始依据

1、工作基础

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具有相关应用^{[2][3][4][5]}。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数据。实现一种针对美国科学研究系统建模及合作模式挖掘建立知识图谱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^{[6][7][8]}。

针对美国科研机构项目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。

2、研究条件

目前国内外关于知识图谱的构建理论和方法已经趋于成熟，美国相关网站数据库可获取大量相关文献以供参考。与此同时，对于当前美国科研机构的项目数据，有美国国立卫生研究院（NIH）、美国国家自然科学基金会（NSF）的科研机构开源大数据，同时还可参考 web of science 中相关论文数据进行参考。近两年来，随着社会各界对科研机构项目的持续关注，使得该领域的相关信息比以往更容易获取，其中包括美国国立卫生研究院

NIH(<https://grants.nih.gov/funding/searchguide/index.html#/>)、美国国家自然科学基金会 NSF(<https://dellweb.bfa.nsf.gov/AwdLst2/default.asp>)、web of science (<https://hfbicd85ae6022a1f4d78s6ob6v60q66u66xkffiac.eds.tju.edu.cn/wos/alldb/basic-search>)、万方(<https://trend.wanfangdata.com.cn/>)、知网 (<https://www.cnki.net>) 等开源数据资源网站。在研究能力上，课题指导教师对相关理论模型和方法具有研究基础，指导经验丰富。学生已具备一定的图数据库编写能力、文本语义网络分析能力、编写爬虫能力与 Media Wiki 的基本操作。

3、应用环境

美国的科学技术一直位于世界前列，人类今天使用的很多先进技术，其发明都源于美国。经过多年的积累和多项计划、改革的实施，美国的科技水平得到了突飞猛进的发展，在火箭技术、武器研究、材料科学、医学、生物工程、计算机等许多领域都处于世界领先地位。在科技水平发达的美国，各学科方向的研究机构数量繁多，主要可分为四种类型：政府科研机构、高等院校、工业界和其他非营利机构^{[9][10]}。

本课题的研究成果不仅能够一定程度上解决美国科研项目知识图谱的匮乏问题，还能借助语义网络分析、实体关系抽取等理论技术提供对美国科研项目的可视化功能，进而为相关科研机构管理部门及时处置提供科学的科研机构项目图谱参考，进而提高决策效率与准确率，降低人力、物力和财力的消耗。此外，本研究成果还可实际应用于万方数据、Media Wiki 等知识图谱网站，为其提供更加清晰完整的科研项目可视化分析与挖掘服务。

4、工作目的

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的语义网络分析，结合对专家人员的实体与关系抽取，构建美国科研机构项目知识图谱，并以 MediaWiki、SMV 等开源工具为基础搭建一个美国科研机构项目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，并通过美国科研机构项目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^[11]。

二、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]Liu, W., et al. Method of constructing knowledge graph network based on massive scientific research data, involves extracting subject word from title information of setting document of each topic as key technology, and clustering direction of topics, State Computer Network & Information Saf (Stcn-C).
- [3]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.
- [4]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).
- [5]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.

- [6]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.
- [7]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.
- [8]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..
- [9]李志民.美国科研机构概览[J].世界教育信息,2018,31(05):6-10.
- [10]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.
- [11]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.

三、设计（研究）内容和要求

1、课题研究内容

(1) 知识图谱相关概念：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，完成前期美国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。

2、对学生的具体要求

(1) 要求学生了解和熟悉知识图谱的相关概念，包括语义网络分析、实体关系抽取等相关概念；

(2) 要求学生能够分析、设计和完成美国科研机构项目知识图谱的基本框架，形成 RDF 文档；

(3) 要求学生熟练学习并掌握 Media Wiki、Semantic-Media Wiki 等软件的基本操作方法，完成知识图谱的构造，并利用主题模型等算法挖掘知识图谱价值；

- (4) 要求学生能够独立完成毕业论文，遵守学术规范，尊重他人学术成果；
- (5) 要求学生论文论点正确，重点突出，结构合理，层次清楚，论证深入有逻辑性，能够反映出一定的理论研究水平；
- (6) 要求学生毕业论文格式应符合《天津大学关于本科生学位论文统一格式规定》，并符合学校新工科毕设项目的基本要求；
- (7) 要求学生主动跟指导老师联系，保证毕业设计按时高效完成。

指导教师 (签字)

年 月 日

审题小组组长 (签字)

年 月 日