

存档编号_____

华北水利水电大学

North China University of Water Resources and Electric Power

毕 业 设 计

题目 中国姓氏文化可视化系统
的设计与实现

学 院 信息工程学院

专 业 软件工程

姓 名 苏玉恒

学 号 202018018

指导教师 _____

完成时间 _____

教务处制

独立完成与诚信声明

本人郑重声明：所提交的毕业设计（论文）是本人在指导教师的指导下，独立工作所取得的成果并撰写完成的，郑重确认没有剽窃、抄袭等违反学术道德、学术规范的侵权行为。文中除已经标注引用的内容外，不包含其他人或集体已经发表或撰写过的研究成果。对本文的研究做出重要贡献的个人和集体，均已在文中作了明确的说明并表示了谢意。本人完全意识到本声明的法律后果由本人承担。

毕业设计（论文）作者签名：

指导导师签名：

签字日期：

签字日期：

毕业设计（论文）版权使用授权书

本人完全了解华北水利水电大学有关保管、使用毕业设计（论文）的规定。特授权华北水利水电大学可以将毕业设计（论文）的全部或部分内容公开和编入有关数据库提供检索，并采用影印、缩印或扫描等复制手段复制、保存、汇编以供查阅和借阅。同意学校向国家有关部门或机构送交毕业设计（论文）原件或复印件和电子文档（涉密的成果在解密后应遵守此规定）。

毕业设计（论文）作者签名：

导师签名：

签字日期：

签字日期：

目录

North China University of Water Resources and Electric Power.....	I
目录.....	VII
摘 要.....	IX
Abstract	XI
第 1 章 绪论.....	2
1.1 研究背景及研究意义.....	2
1.2 研究现状.....	3
1.2.1 知识图谱研究现状.....	3
1.2.2 问答系统研究现状.....	5
1.3 本文主要研究内容.....	5
第 2 章 相关技术和理论基础.....	10
2.1 知识图谱构建技术概述.....	10
2.1.1 知识抽取.....	11
2.1.2 知识融合.....	11
2.1.3 知识加工.....	12
2.2 问答系统技术概述.....	12
第 3 章 中国姓氏文化知识图谱的构建.....	14
3.1 中国姓氏文化知识图谱的构建流程.....	14
3.2 数据采集.....	15
3.3 数据清洗.....	16
3.4 知识抽取.....	16
3.5 知识对齐.....	18
3.6 知识加工.....	18
3.7 知识存储.....	19
第 4 章 需求分析.....	22
4.1 可行性分析.....	22
4.2 系统需求分析.....	22
4.2.1 系统功能概述.....	22
4.2.2 系统目标用户.....	22
4.2.3 功能需求.....	22
4.2.4 数据需求.....	23
4.2.5 性能需求.....	23
4.2.6 运行需求.....	24
4.2.7 可靠性需求.....	24
第 5 章 系统的设计与实现.....	26
5.1 系统概要设计.....	26
5.2 系统详细设计与实现.....	27
5.2.1 主页.....	27
5.2.2 关系检索.....	27
5.2.3 关系全貌.....	28
5.2.4 问答系统.....	28

5.2.5 修改资料.....	29
5.3 系统展示.....	29
5.3.1 开发环境说明.....	29
5.3.2 系统功能展示.....	29
第 6 章 系统测试.....	34
6.1 功能性测试.....	34
6.2 非功能性测试.....	35
第 7 章总结与展望.....	36
7.1 总结.....	36
7.2 展望.....	36
参考文献.....	38
致 谢.....	40
附 录.....	42
外文文献原文.....	42
外文文献译文.....	64
任务书.....	72
开题报告.....	74

中国姓氏文化可视化系统的设计与实现

摘 要

中华文明有着五千年的历史，源远流长。从三皇五帝到春秋战国时期，再到秦汉魏晋时期，之后历经唐宋元明清，经历数千年未有之大变局，最终在毛主席的领导下建立了新中国。中华文明历经了繁荣和衰落，经历了无数个风云变换的岁月。从石器时代的刀耕火种，到当今世界上的互联网，中华文明已经历许多。国史、方志、家谱正是中华民族优秀传统文化传承的三大支柱，“家谱”这一支柱的背后正是中国姓氏文化^[3]。考虑到知识图谱在中国姓氏文化领域的空缺，本文通过研究构建了中国姓氏文化知识图谱，并在完成知识图谱的基础下，对中国姓氏文化可视化系统进行实现。

基于知识图谱的中国姓氏文化可视化系统使用了知识图谱的技术，以 Neo4j 图数据库作为数据库，使用 HTML、CSS、JavaScript 编写前端代码，使用 Flask 这一轻量级框架为基础，编写后端代码。

基于知识图谱的中国姓氏文化可视化系统包括关系检索、关系全貌、问答系统、修改资料四大功能，关系检索：本功能主要用于快速而准确地检索到姓氏与姓氏之间的关系。关系全貌：本功能主要用于查看所有用户之间的起源关系，哪一个姓氏是哪一个姓氏的祖先，哪一个姓氏是哪一个姓氏的祖先。问答系统：通过输入问题，可以查看各个姓氏的姓氏图腾、分布、历史来源、家族名人等细节信息，从而给用户节省不少时间和精力。修改资料：通过依照格式输入修改方法，可以对知识图谱进行修改，比如：添加 A 姓氏与 B 姓氏的关系、删除某个姓氏、修改某个姓氏的某个属性。

本文通过以上工作，构建了中国姓氏文化知识图谱，实现一个包括知识图谱、关系检索系统、问答系统以及修改功能的中国姓氏文化知识服务平台，让用户可以直观地了解姓氏的历史来源、家族名人、分布和姓氏图腾，增强文化认同感和归属感，从而更加深入地了解中国姓氏文化的丰富内涵。

关键词：知识图谱，姓氏文化，问答系统，实体关系抽取，命名实体识别

中图分类号：

Design and Implementation of a Visualization System for Chinese Surname Culture

Abstract

Chinese civilization has a history of five thousand years and a long history. From the Three Emperors and Five Emperors to the Spring and Autumn and Warring States periods, then to the Qin, Han, Wei, and Jin periods, and then to the Tang, Song, Yuan, Ming, and Qing dynasties, experiencing unprecedented changes for thousands of years, the New China was finally established under the leadership of Chairman Mao. The Chinese civilization has gone through prosperity and decline, as well as countless years of ups and downs. From slash and burn farming in the Stone Age to the Internet in the world today, Chinese civilization has experienced a lot. National history, local chronicles, and genealogy are the three pillars of the inheritance of excellent traditional Chinese culture, and behind this pillar is Chinese surname culture. Considering the gap of knowledge graph in the field of Chinese surname culture, this article constructs a knowledge graph of Chinese surname culture through research, and implements a visualization system of Chinese surname culture based on the completion of the knowledge graph.

The Chinese surname culture visualization system based on knowledge graph utilizes the technology of knowledge graph, using Neo4j graph database as the database, HTML, CSS, and JavaScript to write front-end code, and Flask as the lightweight framework to write back-end code.

The Chinese surname culture visualization system based on knowledge graph includes four major functions: relationship retrieval, relationship overview, question answering system, and data modification. Relationship retrieval: This function is mainly used to quickly and accurately retrieve the relationship between surnames. Relationship Overview: This function is mainly used to view the origin relationship between all users, which surname is the ancestor of which surname, and which surname is the ancestor of which surname. Q&A system: By entering questions, users can view detailed information such as surname totems, distribution, historical sources, and family celebrities for each surname, thereby saving them a lot of time and energy. Modifying data: By inputting modification methods according to the format, the knowledge graph can be modified, such as adding the relationship between A and B surnames, deleting a certain surname, and modifying a certain attribute of a certain surname.

Through the above work, this article constructs a knowledge graph of Chinese surname culture and implements a Chinese surname culture knowledge service platform that includes knowledge graphs, relationship retrieval systems, question answering systems, and modification functions. Users can intuitively understand the historical sources, family celebrities, distribution, and surname totems of surnames, enhance their sense of cultural identity and belonging, and thus gain a deeper understanding of the rich connotations of Chinese surname culture.

Key Words: knowledge graph; Surname culture; Q&A system; Entity relationship extraction; Named entity recognition;

第1章 绪论

1.1 研究背景及研究意义

中华文明有上下五千年的历史，可以说的上是源远流长。从三皇五帝到春秋战国，从秦汉魏晋到五代十国，之后历经唐宋元明清，经历数千年未有之大变局以及百年国耻，最终在以毛泽东为核心的中国共产党的领导下建立了新中国。中华文明饱经沧桑，历经了繁荣和衰落，经历了不知多少个风云变换的岁月。从石器时代的刀耕火种，到当今世界上的互联网，中华文明已经历许多。

国史、方志、家谱正是中华民族优秀传统文化传承的三大支柱，“家谱”这一支柱的背后正是中国姓氏文化。有一句俗语“姓者，统其祖考之所自出；氏者，别其子孙之所自分。”姓氏是一个人的家族血缘关系的标志。姓氏最早的作用主要是便于辨别部落中不同氏族的后代，便于不同氏族之间的通婚。^[3]因此姓氏的产生是原始人类从群婚制到以血缘关系的婚姻制转变的标志，对全人类的文明进步有里程碑式的意义。

通过对姓氏文化的研究，可以更好地弘扬中国优秀传统文化，提高中国文化软实力、加深海峡两岸之间的情感往来，提高中国的文化自信。

随着互联网的不断发展、人们生活水平的持续提高和人人之间文化交流的不断加深，中国姓氏文化的相关研究也越来越受到人们的重视。许多人希望自己能够更好的了解自己姓氏的历史来源、分布、名人等等，满足自己的好奇心。但是虽然搜索引擎可以帮用户轻松的找出相关资料，但是搜索引擎找出的资料良莠不齐，各式各样的关于中国姓氏文化信息相互交织。

所以用户最终要解决的问题就是怎样让用户快速且准确的得到有效中国姓氏文化信息。问答系统和关系检索系统的设计正是为了解决这一问题。问答系统以问答的形式，用户输入问题后，问题经过处理，分析出相关实体，匹配了相关 Neo4j 图数据库查询模板，对 Neo4j 图数据库进行知识检索，然后通过 Flask 框架将信息展现在浏览器上，不仅会以图形的方式展现用户所查询的姓氏与其他姓氏的关系，也将展现用户所查询的姓氏的属性。跟传统的搜索引擎比起来，最终方式可以更加准确地获取想要的中国姓氏文化知识。因此，中国姓氏文化问答系统以一问一答的形式使用户与经过整合的中国姓氏

文化知识图谱进行了直接的交互，中国姓氏文化关系检索系统将会直观地展现用户所查询的姓氏与其他姓氏的关系，问答系统和关系检索系统的设计帮助用户快速且准确的得到他们真正需要的信息。

知识图谱可以让信息用结构化的方式整合在一起并表达在浏览器上面，从而能够帮人们更详细的理解信息之间的关联和所有信息的完整性，不同的信息凭借知识图谱相互联结从而形成完整的、系统的知识网络。知识图谱是一个极其方便的辅助工具，在当今的时代，知识图谱与互联网一同被运用的了不同行业、不同领域，比如说历史^[21]、西藏文物^[17]、中华典籍^[6]、心理咨询^[7]、健康饮食^[8]、海洋减灾^[9]、作物种植^{[10][23]}、西藏旅游^[11]、糖尿病治疗^{[19][12]}、泥石流防治、高考志愿填报^[18]、法律法规^{[16][13]}、汽车、马铃薯病虫害^[15]、医疗^{[14][22]}等领域，来帮助使用者解决复杂问题，充当说明书的作用。

这项研究有着非常重要的意义，中国姓氏文化又是中华优秀传统文化的重要组成部分，它是中华民族的根。通过对不同姓氏的起源关系、家族名人、分布、姓氏图腾进行梳理，从而弘扬中国姓氏文化，增强中国的文化实力和民族自豪感，提升中国人民的凝聚力，增进海峡两岸的友好关系。

因此，本项研究使用爬虫对不同网站的信息进行了爬取，进行了数据的预处理，抽取实体、实体间的关系以及实体的多个属性，用得到的中国姓氏文化的相关数据完成了中国姓氏文化知识图谱的构建工作。在中国姓氏文化知识图谱构建完成的基础下，进一步把中国姓氏文化知识图谱问答系统和中国姓氏文化知识图谱关系检索系统设计出来，并且专门设计了一个网页来展现中国姓氏文化中不同姓氏的起源关系的全貌，而且，考虑到可能会对知识图谱进行一定的补充与修改，又设计出了资料修改的网页，在该网页可执行修改命令，从而实现对知识图谱的更新，这些才最终组成了中国姓氏文化可视化系统。通过使用中国姓氏文化可视化系统，让人们能够更加准确的、更加完整的、更加详细的了解中国姓氏文化，同时这也能够作为外国孩子了解中国姓氏文化的教学工具，也为探究中国姓氏文化提供了一点小小的帮助。

1.2 研究现状

1.2.1 知识图谱研究现状

随着信息时代的到来，Web 技术不断演化与发展，根据 W3C 中的解释，所谓的语义

网络，其实是一张以数据为节点、以关系为丝线构成的一张网络。用户真正从语义网络技术中得到的是一个特定的查询环境，它的精髓是向用户返回以图形的方式展现的被语义网络技术进行加工、融合、推理而得到的知识。知识图谱技术正是实现对语义网络进行智能语义检索的方法。^[4]

其实，传统搜索引擎技术是能够根据用户查询快速排序网页的，从而提高了信息检索的效率。尽管检索效率较高，但用户并不一定能够快速准确的获取用户真正想要的信息和知识，因此对于传统搜索引擎技术返回的大量的结果，往往需要进行人工筛选，这必将耗费人们大量的精力。伴随着信息时代的到来，互联网中的信息总量呈指数增长，传统搜索引擎技术已经很难满足人们对互联网信息的检索需求。知识图谱技术的出现给解决用户的信息检索需求开拓了新的道路。^[5]

早在 2015 年，知识图谱技术就已经应用到了工业界。为了提高搜索引擎的运行效率以及搜索内容的准确性，Google 公司在 2015 年五月十七日提出了知识图谱这一概念，并宣布要以此为基础创建新一代的更加智能化的搜索引擎，为此 Google 公司还收购了已经掌握了语义搜索核心技术的 Metaweb 公司，从而获得了该公司的语义搜索技术，该技术包括从因特网上抽取知识，分析实体及其属性的相关信息，建立实体与实体之间的联结关系，从而构建出一个知识网络，创造出一套 web 语义知识网络库。从而让用户不再只能通过简单的关键词匹配功能进行搜索，建立了一套全新的知识检索方式。^[4]

本质上，知识图谱就是结构化的语义知识库，是一种概念网络。知识图谱由两部分组成，第一部分是“实体—关系—实体”三元组，其中的实体是现实物理世界真实存在的物体（或者是概念），第二部分是实体与实体属性组成的键值对。实体与实体之间通过关系联结在一起，组成了一张语义网络，也就是知识图谱。知识图谱把一系列复杂的数据结构化，实现了真正的语义检索。基于知识图谱的搜索引擎可以以图形的形式把结构化的知识呈现在 Web 上面，从而反馈给用户，使得用户不必大量浏览网页，可以准确而高效的找到用户希望查询到的知识。^[5]

知识图谱的逻辑结构分为两个层次：第一个层次是数据层，数据层只有两种三元组，分别是“实体—关系—实体”三元组和“实体—属性—属性值”三元组，知识存储在图数据库的单位是实体。第二个层次是模式层。

尽管 2015 年知识图谱技术才应用到工业界，但其实知识图谱这一概念在 2005 年就在学术界被提出来了，当时被提出来的是一种叫做数据链接的思想，这种思想的提出当即掀起了一场数据链接的研究热潮，知识图谱这一概念就是从数据链接这一思想中升华

出来的。知识图谱这一概念建立在过去的种种研究成果之上。^[4]

最近这些年，世界各国的学者开始进行知识图谱的相关研究，在中国，有许多科研工作者开始了对知识图谱的研究，比如说，陈悦、刘则渊的《悄然兴起的科学知识图谱》^[20]，侯海燕的《国际科学计量学核心期刊知识图谱》，侯海燕《基于知识图谱的科学计量学进展研究》，到现在，知识图谱的研究工作已经有了雄厚的基础，比如说，对已经构建好的知识图谱修改后，知识图谱可以实时更新。可以预见，以后知识图谱的相关研究必然仍受到世界各地的科研工作者的持续关心，知识图谱的发展前景非常值得期待。

1.2.2 问答系统研究现状

基于知识图谱的问答系统已经成为最近五年的研究热点。基于知识图谱的问答系统的兴起，极大的方便了用户的信息搜索过程。在基于知识图谱的问答系统兴起之前，用户如果想要获取想要的知识，只能通过 Google、Bing 等传统搜索引擎，然后，呈现在用户面前的是相应的一长串 HTML 数据的列表。在基于知识图谱的问答系统兴起之后，中国姓氏文化问答系统以问答的形式使用户与经过整合的中国姓氏文化知识图谱进行了直接的交互，简洁高效，大大节省了用户的检索时间，省去了人工筛选最终答案过程中的耗费的时间与精力。

2016 年以来，国内的研究基于知识图谱的问答系统的科研工作者们也逐渐开始有针对性地进行更深入的研究，比如说，杜泽宇的《基于中文知识图谱的电商领域问答算法设计与系统实现》，杨燕的《面向电商领域的智能问答系统若干关键技术研究》，杜泽宇，杨燕等的《基于中文知识图谱的电商领域问答系统》等等。

基于知识图谱的问答系统的研究在软件工程的专业领域受到了极为热烈的关注，研究者通过搜索引擎模式分词，问句模板匹配充分的理解用户的种种意图，生成 Neo4j 查询语句，将查询结果生成为 JSON 文件，从而以文字和图像的形式呈现在浏览器上，即时对用户的请求做出回复，从而实现基于知识图谱的问答系统。

1.3 本文主要研究内容

本研究的主要研究内容是中国姓氏文化知识图谱的构建、中国姓氏文化关系检索系统的构建、中国姓氏起源关系全貌的展现、中国姓氏文化问答系统的构建、中国姓氏文化知识图谱修改功能的实现，这五项最终构建成一个完整的中国姓氏文化可视化系统。该系统的用户主要是一些中国姓氏文化研究人员和一些对中国姓氏文化与《百家姓》

感兴趣的普通大众，帮助他们解除一些关于中国姓氏文化的疑惑，帮助用户更加生动的了解不同姓氏的起源关系，帮助他们节省了搜集、整理、分析相关信息的时间和精力。整个中国姓氏文化可视化系统的设计思路如图 1-1 所示。

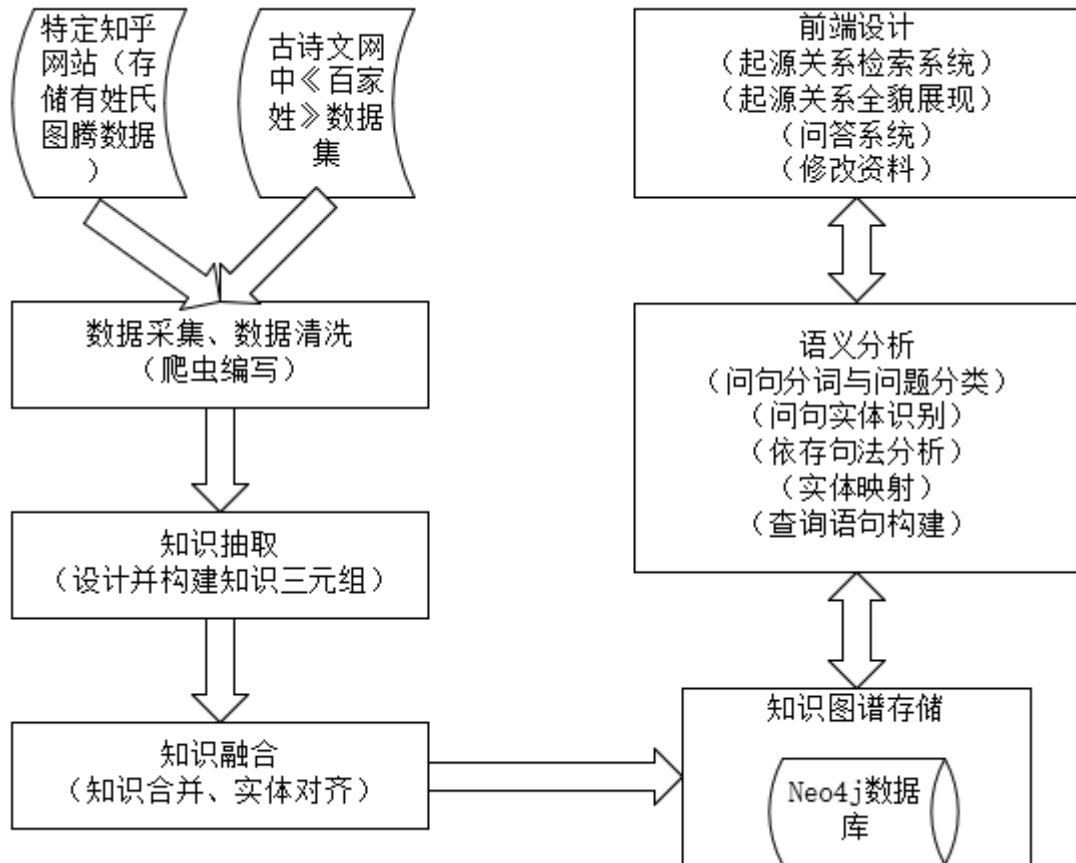


图 1-1 系统整体设计思路图

因此，本研究主要分为四个方面：

（1）构建中国姓氏文化知识图谱

第一步，信息搜集。调查研究许多与中国姓氏文化相关的网站，构建中国姓氏文化知识图谱。中国姓氏文化知识图谱的实验数据^{[1][2]}主要源于各类网络平台。通过编写网络爬虫，爬取信息，对全书文字进行分门别类，不同姓氏，有着不同的历史来源、许多个家族名人、零散的分布，然后从分门别类后的文字中提取数据信息。

信息清洗。比如某些姓氏没有姓氏图腾，要补充上“暂无资料”，比如说某些名人的生卒年月不完整，要全部去掉。

信息抽取。主要是抽取其中的实体，也就是姓氏。

知识融合。两部分数据都有姓氏作为实体，以此为基础，对齐数据即可，中国姓氏

文化知识图谱的构建已完成。

最后，使用 py2neo 登录 Neo4j，将已经构建完成的中国姓氏文化知识图谱存储到 Neo4j 图数据库里面。

（2）语义分析

语义分析分为三类：中国姓氏起源关系检索系统、中国姓氏文化问答系统、修改中国姓氏文化资料的功能。中国姓氏起源关系全貌的展现并不需要输入，只需要鼠标点击即可。对于以上三类语义分析，都有着基本相似的步骤：第一步，对输入的问题进行分类，对输入的文本分词。第二步是识别输入的文本里面的实体，第三步，进行依存句法分析，第四步是要把第二步识别出的实体映射到代码层面，确认第二步中识别出的实体是否存在，第五步，要整合前四步的内容，构建出需要运行的 Neo4j 图数据库专用语句并进行运行。

（3）前端设计

完成了中国姓氏文化知识图谱的构建以及语义分析工作之后，应当以被存储到 Neo4j 图数据库的中国姓氏文化知识图谱为数据源，基于知识图谱的中国姓氏文化可视化系统的前端部分主要是由 HTML、CSS 和 Java Script 组成，使用这些技术实现在网页上的设计和信息交互逻辑，CSS 和 Java Script 主要存储在 static 文件夹里面，HTML 文件主要存储在 templates 文件夹里面。基于知识图谱的中国姓氏文化可视化系统的后端部分主要使用 Python 语言的 Flask 框架搭建 Web 服务，接收来自前端的请求并与 Neo4j 图数据库进行交互，从而将整个系统进行封装，把中国姓氏文化关系检索系统、中国姓氏起源关系全貌的展现、中国姓氏文化问答系统、中国姓氏文化知识图谱修改功能展现这些都封装到 web 程序里面，除了上述四项功能每一项都设计一个网页，再设计一个主页，呈现在浏览器上面，用户也已直接用浏览器打开 <http://127.0.0.1:5000> 进行访问，从而实现了基于知识图谱的中国姓氏文化可视化系统的多项功能。

（4）系统的有四项主要功能

构建完整的中国姓氏文化关系检索系统。在上述研究的基础上，在浏览器中请求访问输入姓氏或点击姓氏，与构建好的中国姓氏文化知识图谱从而进行实体映射，组成完整的查询语句，然后从构建好的中国姓氏文化知识图谱进行知识检索，并将得到的数据转化为 JSON 格式。将得到的结果通过 Flask 框架及工具对进行封装，然后把数据消息发送给浏览器，从而实现查询知识图谱关系检索功能。

展现完整的中国姓氏起源关系全貌。在浏览器中请求访问，从构建好的中国姓氏文化知识图谱调取 Neo4j 图数据库的全部关系数据，并将得到的数据转化为 JSON 格式。将得到的 JSON 数据通过 Flask 框架对进行封装，然后展现在 Web 端，从而实现展现完整的中国姓氏起源关系全貌功能。

构建完整的中国姓氏文化问答系统。通过分析用户感兴趣的问题，设定一系列问题模板，然后通过模板匹配的方法将中国姓氏文化问答系统构建出来。具体的说，第一步，要对用户问句进行搜索引擎模式的中文分词，一步一步地分析用户提出的问题，对各种问题进行分类并构建出相应的问题查询语句模板。然后对询问句中的实体（姓氏或姓氏的属性）进行识别，与已经构建好的中国姓氏文化知识图谱进行实体映射，匹配相应的问题查询语句模板，两者组成完整的查询语句，然后从构建好的中国姓氏文化知识图谱进行知识检索，并生成 JSON 文件。将得到的结果通过 Flask 框架对进行封装，传输到浏览器中展现出来，从而实现中国姓氏文化问答系统的问答功能。

实现中国姓氏文化知识图谱修改功能。通过分析用户可能对知识图谱进行修改，设定一系列资料修改模板，然后通过模板匹配的方法修改中国姓氏文化知识图谱。具体的说，第一步，要对用户的修改命令进行中文分词，一步一步地分析用户发送的修改命令，对不同修改命令进行分类并构建出相应的修改命令语句模板。然后识别修改命令中的实体（姓氏或姓氏的属性），与已经构建好的中国姓氏文化知识图谱进行实体映射，匹配相应的修改命令语句模板，两者组成完整的修改命令语句，然后修改构建好的中国姓氏文化知识图谱，并生成相应的 JSON 文件。将得到的结果通过 Flask 框架对进行封装，传输到浏览器中展现出来，从而实现中国姓氏文化知识图谱修改功能。

第 2 章 相关技术和理论基础

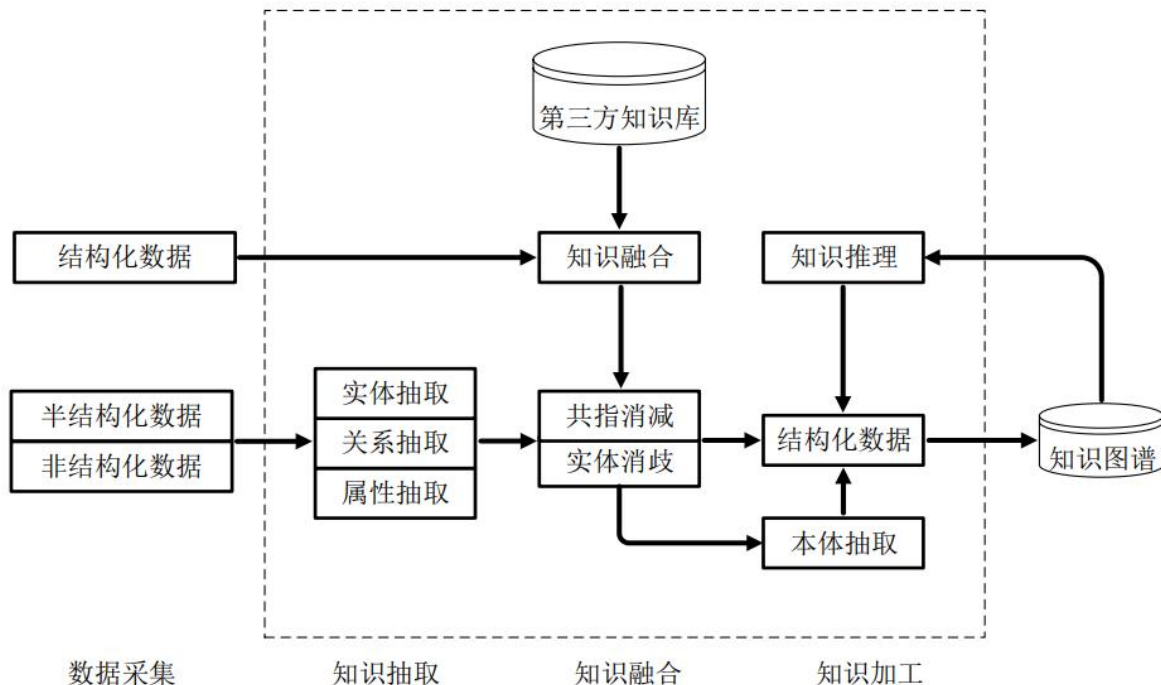
本章将深入介绍构建知识图谱所涉及的技术，例如，知识抽取、知识融合、知识存储、知识加工、知识推理等，以及构建智能问答系统的相关技术和理论方法。

2.1 知识图谱构建技术概述

本质上，知识图谱是一种有知识构成的语义网络，以图的形式揭示实体与实体之间的关系，存储在图数据库里面。知识图谱常常用三元组表示，有两种三元组，分别是“实体—关系—实体”三元组和“实体—属性—属性值”三元组，知识存储在图数据库的最基本的单位是实体。关系可以描述实体与实体之间的关联。

知识图谱也被分为两大类，分别是通用知识图谱和行业知识图谱。通用知识图谱覆盖范围较广，常被用作个性化推荐、智能搜索等领域。专用知识图谱覆盖范围较小，主要依靠特定行业的数据进行构建，对于该特定的行业极具实用性，比如玉米育种知识图谱、东周列国历史知识图谱、家用药箱知识图谱、建筑法规知识图谱、文物知识图谱等等。

知识图谱的技术架构如图 2-1 所示：



2.1.1 知识抽取

知识抽取是知识图谱构建过程中最重要也是最复杂的一个步骤。它主要是对半结构化和非结构化的数据进行抽取，抽取出信息的微小单元，包括实体、关系以及属性三类，并初步组织成“实体—关系—实体”三元组和“实体—属性—属性值”三元组的形式，从而初步获得结构化的数据，当然，如果已经是结构化的数据，显然不需要抽取，只需稍加整理即可，比如说，本项研究中姓氏图腾的数据是结构化的数据。知识抽取又根据抽取内容的不同分为三个大类，分别是实体抽取、属性抽取以及关系抽取。

在知识图谱研究的早期，实体抽取又被叫做命名实体识别，它要从原始数据里面分析出原始数据的排布规律，然后编写爬虫提取出相应实体，实体抽取的准确率将会对后续工作产生极大的影响，因此实体识别是知识抽取中最为基础也是最为关键的一部分操作。实体抽取分为三种：第一种是基于规则与词典的实体抽取方法，这种方法需要基于目标实体编写出一段模板，跟语料库进行匹配。第二种是基于深度学习的实体抽取方法，主要是通过机器学习的方法，使用原始语料进行训练，训练好一个模型去进行识别。第三种是面向开放域的实体抽取方法，这种实体抽取需要面对整个因特网。本项研究中主要使用第一种方法基于规则与词典的实体抽取方法进行训练。

完成实体抽取后，还要进行关系抽取，将实体与实体之间的关系抽取出来，这样才能形成一套完整的网状知识网络。在知识图谱发展早期，关系抽取主要是通过人工构造语义规则以及模板的方法才能识别出实体与实体之间的关系，显然这需要拥有极高的语言学造诣，关系抽取又分为开放式实体关系抽取和基于联合推理的实体关系抽取两大类，这两类中后者较为复杂。

对于属性抽取，大量的属性数据都存在在于非结构化、半结构化的百科网站中，可以把属性当成一种特殊的实体来去提取。

2.1.2 知识融合

通过知识抽取对半结构化和非结构化的数据进行抽取，抽取出信息的微小单元，包括实体、关系以及属性三类，并初步组织成“实体—关系—实体”三元组和“实体—属性—属性值”三元组的形式，已经初步获得了获得结构化的数据。但知识抽取得到的结构化数据仍然包括大量的重复冗余或者错误的信息，信息良莠不齐，关联不明确，因此必须进行知识融合，让来自不同地方的不同结构的知识被整合到同一个结构和格式里面，消除其中的错误和重复，形成一个高质量的完整的知识图谱。

知识融合分为 2 个步骤，第一步叫做是实体对齐，第二步叫做知识合并。

实体对齐主要是针对有时在上一步知识抽取过程中，可能会遇到知识抽取得到的许多实体都指代的是同一个对象，比如说，班级里面出现了同名同姓的同学。显然，如果不进行实体对齐会让查询知识图谱时的数据质量降低。

知识合并是指不同知识图谱进行合并的过程，知识合并能够让知识图谱结构更加完整，让知识图谱中正确的链接指向正确的实体，使知识图谱更具系统性。

2.1.3 知识加工

知识加工分为本体构建和知识推理两步。知识抽取和知识融合让半结构化数据和结构化数据转变成为了结构化数据。但抽取得来的数据仍然存在一些错误数据，本体构建过程就可以让知识图谱的错误内容以及低质量内容删除。

知识推理分为基于逻辑的知识推理和基于图的知识推理，知识推理是在已有知识图谱的基础上更深一步的挖掘隐藏的知识和潜在的关联。

2.2 问答系统技术概述

问答系统是根据用户提出的问题进行语义分析，然后到数据库中进行信息检索，进行答案抽取，最终给出准确的结果。

问答系统的构建过程中主要有三个问题要解决，分别是语义分析、信息检索和答案抽取。问答系统常常被用于网络客服的工作中。

语义分析是在用户给定问句的情况下，将用户的问句与事先设定的模板匹配后进行分类，然后提取问句里面与知识图谱相关的实体或者属性。一旦用户给出问句，就用已经事先设定好的问题模板与问题进行匹配，从而确定问题的分类。问题分类确定后，就可以依据模板提取出问句里面的主题词汇。

信息检索是根据语义分析中提取到的主题词汇，在知识图谱之中进行查询，去除无关内容，缩小查询范围。答案抽取即是把信息检索里面得到的答案传输出来，传递到用户面前，从而实现问答系统与用户的交互，

这一类基于规则的问答系统相比于基于深度学习的问答系统效率更高，准确性更高，但问题模板依赖人工标记，如果问题模版与用户的问句不能够匹配的时候，往往不能给出正确的答案。问题模板需要大量精力去预设，难以真正理解用户问题。

第3章 中国姓氏文化知识图谱的构建

构建中国姓氏文化知识图谱能够给基于知识图谱的中国姓氏文化可视化系统提供数据支撑。知识图谱能够将半结构化的数据和非结构化的数据转变成结构化的数据，让信息资源变得更加易于理解、更加方便检索。

构建知识图谱通常有两种方式，分别是自顶向下构建知识图谱和自底向上构建知识图谱，其中自顶向下构建知识图谱是知识图谱构建中较为常用的方式，这种方式是先收集所有数据，然后数据清洗，知识融合，最后定义模式层，把实体添加到概念里面，定义数据模式。自底向上是另一种常见方法，首先对底层实体总结为概念，然后逐步向上抽象出上层概念。

本章将对中国姓氏文化知识图谱的构建进行研究。首先，梳理中国姓氏文化知识图谱的构建流程；然后，研究中国姓氏文化知识抽取、知识融合和知识加工方法；最后，利用Neo4j图数据库存储中国姓氏文化关系三元组，完成中国姓氏文化知识图谱的构建，为基于知识图谱的中国姓氏文化可视化系统提供数据基础。

3.1 中国姓氏文化知识图谱的构建流程

中国姓氏文化知识图谱的构建采用自顶向下的方式，中国姓氏文化知识图谱的构建过程包括以下几个阶段：

- （1）数据采集，编写网络爬虫对中国姓氏文化的数据进行收集。
 - （2）数据清洗，对收集回来的中国姓氏文化数据进行清洗。
 - （3）知识抽取，对收集到的中国姓氏文化数据进行知识抽取。
 - （4）知识对齐，对不同来源的不同结构的数据进行实体对齐。
 - （5）知识加工，补全中国姓氏文化知识图谱，提高中国姓氏文化知识图谱的可靠性。
 - （6）知识存储，将构建好的中国姓氏文化知识图谱存储到 Neo4j 图数据库里面。
- 中国姓氏文化知识图谱的构建流程如图 3-1 所示：

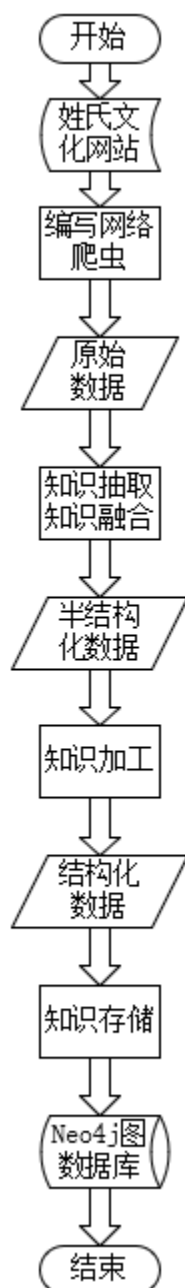


图 3-1 中国姓氏文化知识图谱的构建流程

3.2 数据采集

构建中国姓氏文化知识图谱需要用到大量数据。数据采集正是为构建中国姓氏文化知识图谱搜集大量数据。中国姓氏文化知识图谱的实验数据主要源于各类网络平台，通过编写网络爬虫，收集信息，对全书文字进行分门别类，不同姓氏，有着不同的历史来源、许多个家族名人、零散的分布，然后从分门别类后的文字中提取数据信息。

对于互联网里面的数据，本研究使用的 requests、bs4 等软件包编写网络爬虫对网页内部的元素进行解析并且获取其中的中国姓氏文化数据。第一步先分析百家姓的姓

氏列表，从形式列表中解析出不同姓氏的详情页的 URL，并将它们添加到 URL 数组中，程序会依次从 URL 数组中取出链接，进行信息解析，显然古诗文网中《百家姓》里面的数据的数据结构结构单一，信息解析方式也较为单一，并把爬到的数据保存到文件里面。使用爬虫脚本获取了约 400 个页面，以半结构化数据为主。

3.3 数据清洗

对获取到的中国姓氏文化知识图谱进行数据清洗，在爬取到的数据中，可能会出现不规范的奇怪文字、冗余信息、特殊符号、中英符号格式或者乱码等。使用 python 自带的字符串处理方式之后，数据还要进行格式化处理，使用 python 创建字典，从而提升数据质量。

3.4 知识抽取

经过了上述操作，得到了中国姓氏文化的数据结构清晰的数据。通过观察可以发现，本项研究在以上几步获取的数据，上述使用 python 创建的字典，每一个字典元素的键为一个姓氏，值为一个小字典，每一个小字典的都有四个项，分别是“历史来源”、“家族名人”、“分布”。

中国姓氏文化是以姓氏为实体，通过分析中国姓氏文化的“历史来源”数据，可以得到姓氏间的关系。由于中国古代的宗法制度，宗族中分为大宗和小宗，当小宗的人数达到一定程度，小宗可能会更改姓氏，不同姓氏之间只有一种关系：后代。每个姓氏都有各自的属性“历史来源”、“家族名人”、“分布”和“姓氏图腾”4 类属性。

显而易见，两类知识三元组已经形成。

本项研究设置有一类实体：姓氏，如表 3-1 所示，据统计，本项研究一共建立了 535 个实体。

表 3-1 实体信息及其示例

实体类型	示例
姓氏	杨

本项研究中姓氏之间只有一种关系：后代，如表 3-2 所示，据统计，本项研究一共建立了 534 个关系。

表 3-2 姓氏关系信息及其示例

关系类型	示例
后代	<杨, 后代, 茹>

本性研究中，每个姓氏的属性分为 4 种：“历史来源”、“家族名人”、“分布”和“姓氏图腾”，如表 3-3 所示。

表 3-3 姓氏属性信息及其示例

属性类型	示例
历史来源	<p><杨, 历史来源, “杨源出一出自姬姓。以国为氏。据《元和姓纂》云。周武王分封其第叔虞于唐邑山西翼城。出公子齐, 生伯侨, 天子封为杨侯, 以国为氏。</p> <p>二为五胡十六国时氏族姓氏, 据《晋书》云氏有杨氏, 世居仇池甘肃仇池山。</p> <p>三据《魏书官氏志》所载, 北魏有代北姓莫胡芦氏, 后改为汉字单姓杨氏。</p> <p>四出自扬姓。以邑为氏。古时有扬氏邑。”></p>
家族名人	<p><杨, 家族名人, “杨坚</p> <p>杨广</p> <p>杨玉环</p> <p>杨行密</p> <p>字化源。五代吴国开国君主。庐州合肥人。少为群盗, 占据庐州, 唐昭宗拜为淮南节度使, 封吴王, 悉有淮南江东地, 在位十五年。后其子溥称帝, 追尊为太祖武皇帝。</p> <p>杨业</p> <p>宋朝太原人。初事北汉, 骁勇善战, 赐姓刘, 名继业。北汉亡, 归宋, 刺代州, 积功迁云州观察使, 后潘美以忠武军节度使</p>

	<p>北征，业副之，连拔应、寰、朔各州。会契丹国母萧氏率众南下，复陷寰州，业等在朔州，蔚州刺史王侁等偕业赴敌，既战而败，援兵不至，力竭被擒，不食三日死。后杨氏几代杨延昭，杨文广一直为北宋王朝捍卫疆土，杨家将的忠勇事为后人传诵。</p> <p>杨秀清</p> <p>原名嗣龙，清朝广西桂平人。因与洪秀全约为兄弟，故改名秀清。与秀全同起事，屡败清兵，封东王，太平军号令几全出自秀清，权势颇盛，故世并称洪杨，后为韦昌辉所杀。”></p>
分布	<p><杨，历史来源，“河南弘农郡，甘肃天水郡，河南河内郡。” ></p>
姓氏图腾	<p><杨，姓氏图腾，“图腾由来自远古三苗之一的柯挪耶劳，祖先是兵主蚩尤。”杨”意为太阳，由”木”和”易”组成，表示扶桑树生长在东方大海上的汤谷。” ></p>

3.5 知识对齐

中国姓氏文化知识图谱有着多个不同结构的数据源，知识质量良莠不齐，对于同一个实体，有着不同的表达这必将影响知识图谱的知识质量。

本项研究中，有一个数据源存储有不同姓氏“姓氏图腾”的属性，因此要将该数据源中的数据转入知识图谱。

3.6 知识加工

到目前为止，知识图谱已经存储有大量的数据，但知识图谱并不完整，比如说：有的姓氏实体的“分布”属性为空，有的姓氏实体的“姓氏图腾”属性为空，有的姓氏实体的“家族名人”属性为空，因此中国姓氏文化知识图谱仍然不够完整，需要进行进一步的补充。

因此，内容为空的属性，应该填充“暂无资料”四个字。

3.7 知识存储

中国姓氏文化知识图谱的知识存储，就是要分析清楚整个知识图谱的结构特点，在非结构化数据和半结构化数据彻底转化为结构化数据之后，将所有数据存储到数据库里面。由于图数据库擅长处理复杂的关系、灵活性和可扩展性较强、查询效率极高，在知识图谱的构建中，人们总是使用 Neo4j 图数据库作为数据库，Neo4j 图数据库被广泛地使用在学术领域，Neo4j 图数据库还有专门的数据库语言，Cypher Query Language(CQL)，与大多数数据库使用的 SQL 不同，CQL 是专用于图数据库的数据库查询语言，可以使用户轻松的理解与操作图形数据结构。另外 python 还有专门支持 Neo4j 图数据库的软件包 py2neo，这可以让人使用轻松的编写操作图数据库的代码。

Neo4j 图数据库支持构建超大规模知识图谱，导入节点数量大、速度快、效率高，它本质上是一个复杂网络，无论是查询节点，还是去查询节点与节点之间的关系，具有极高的查询效率。因此，Neo4j 图数据库是构建中国姓氏文化知识图谱的最佳选择。

本项研究中，在经过以上一系列处理后，形式为“实体—关系—实体”的三元组数据存储在文件 fangji_spo.txt 里面和形式为“实体—属性—属性值”的三元组数据存储在文件 zhongyao_spo.txt 里面。

先使用控制台运行 Neo4j 图数据库，再使用 python 的 py2neo 包提供的诸多 API 登录 Neo4j 图数据库后，使用 `graph.run("MERGE(p: Entity{cate:'%s', Name: '%s'})" % (rela_array[3], rela_array[0]))` 创建一个节点，并添加节点的名字，再创建节点之间的关系，以及给节点导入“历史来源”、“家族名人”、“分布”和“姓氏图腾”等属性。

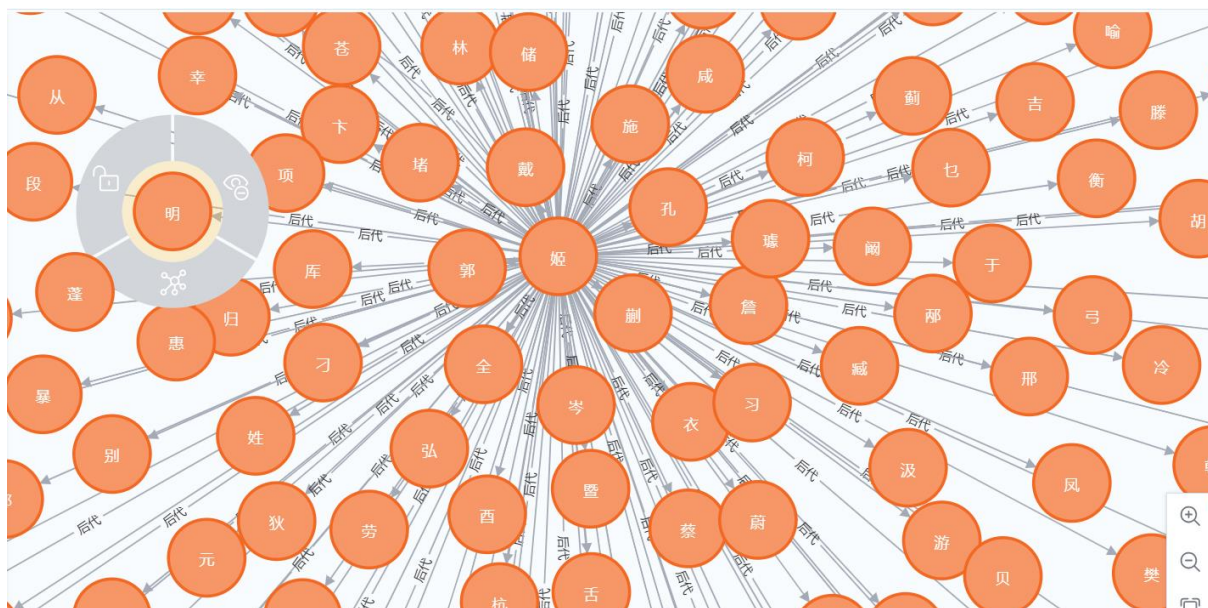


图 3-2 部分中国姓氏文化知识图谱可视化

第4章 需求分析

在第三章中，本研究已经实现了中国姓氏文化知识图谱的构建，本章将会以第三章中建立在 Neo4j 图数据库中的中国姓氏文化知识图谱为知识库，完成基于知识图谱的中国姓氏文化可视化系统的设计与实现。

4.1 可行性分析

基于知识图谱的中国姓氏文化可视化系统使用了知识图谱的技术，以 Neo4j 图数据库作为数据库，使用 HTML、CSS、JavaScript 编写前端代码，使用 Flask 这一轻量级框架为基础，编写后端代码。

关键技术与资源均已具备，因此该系统可行性极高。

4.2 系统需求分析

4.2.1 系统功能概述

基于知识图谱的中国姓氏文化可视化系统可以帮助用户更充分的理解中国姓氏文化，更加方便地查询各个姓氏的“历史来源”、“家族名人”、“分布”和“姓氏图腾”等信息，更有利于弘扬中华优秀传统文化。

基于知识图谱的中国姓氏文化可视化系统主要包括中国姓氏文化关系检索系统、中国姓氏起源关系全貌的展现、中国姓氏文化问答系统以及中国姓氏文化知识图谱修改资料功能这四项功能。

4.2.2 系统目标用户

该系统的用户主要是一些中国姓氏文化研究人员和一些对中国姓氏文化与《百家姓》感兴趣的普通大众，从而帮助他们解除一些关于中国姓氏文化的疑惑。

4.2.3 功能需求

- （1）关系检索：本功能主要用于快速而准确地检索到姓氏与姓氏之间的关系。
- （2）关系全貌：本功能主要用于查看所有用户之间的起源关系，哪一个姓氏是哪一个姓氏的祖先，哪一个姓氏是哪一个姓氏的祖先。
- （3）问答系统：通过输入问题，可以查看各个姓氏的姓氏图腾、分布、历史来源、家族名人等细节信息，从而给用户节省不少时间和精力。

（4）修改资料：通过依照格式输入修改方法，可以对知识图谱进行修改，比如：添加 A 姓氏与 B 姓氏的关系、删除某个姓氏、修改某个姓氏的某个属性。

4.2.4 数据需求

（1）静态数据：

1. 使用 Neo4j 图数据库存储姓氏与姓氏之间的关系、姓氏的类型、姓氏的历史来源、姓氏的分布、姓氏的家族名人、姓氏的姓氏图腾。
2. 前端使用到的姓氏图片、华北水利水电大学 logo 图标、主页封面图片。
3. 暂时记录姓氏关系的 JSON 文件，放置在 static 文件之下。

（2）动态数据

1. 关系检索功能中输入的姓氏、回复到网页端的 JSON 文件、图片。
2. 问答系统功能中输入的句子、回复到网页端的 JSON 文件、图片。
3. 修改资料功能中输入的修改内容、回复到网页端的 JSON 文件、图片、对 Neo4j 图数据库进行的修改。

（3）数据字典，如表 4-1 所示。

表 4-1 基于知识图谱的中国姓氏文化可视化系统数据字典

知识图谱	
数据名	数据类型
姓氏	Text
姓氏类型	Text
姓氏关系	Text
姓氏图腾	Text
历史来源	Text
家族名人	Text
分布	Text

4.2.5 性能需求

（1）数据精度

一般

（2）时间特性

系统响应时间 $\leq 100\text{ms}$

更新处理时间 $\leq 2000\text{ms}$

数据传送时间 $\leq 100\text{ms}$

数据计算时间 $\leq 100\text{ms}$

4.2.6 运行需求

（1）用户界面

要求每个界面简单大气，配色合理、简洁清新，方便操作，有比较好的用户体验。

（2）软件接口

服务器运行在 Flask、Neo4j 的环境里面。

4.2.7 可靠性需求

（1）系统容错性

基于知识图谱的中国姓氏文化可视化系统应当可以依据 Neo4j 图数据库进行恢复。

（2）数据备份与恢复

时常进行中国姓氏文化知识图谱的数据备份，系统故障时，快速恢复中国姓氏文化知识图谱数据。

第5章 系统的设计与实现

5.1 系统概要设计

基于知识图谱的中国姓氏文化可视化系统具有这四个功能：关系检索、关系全貌、问答系统、修改资料。基于知识图谱的中国姓氏文化可视化系统采用三个层次来构建，分别是：显示层、业务层、数据层。系统概要设计如图所示：

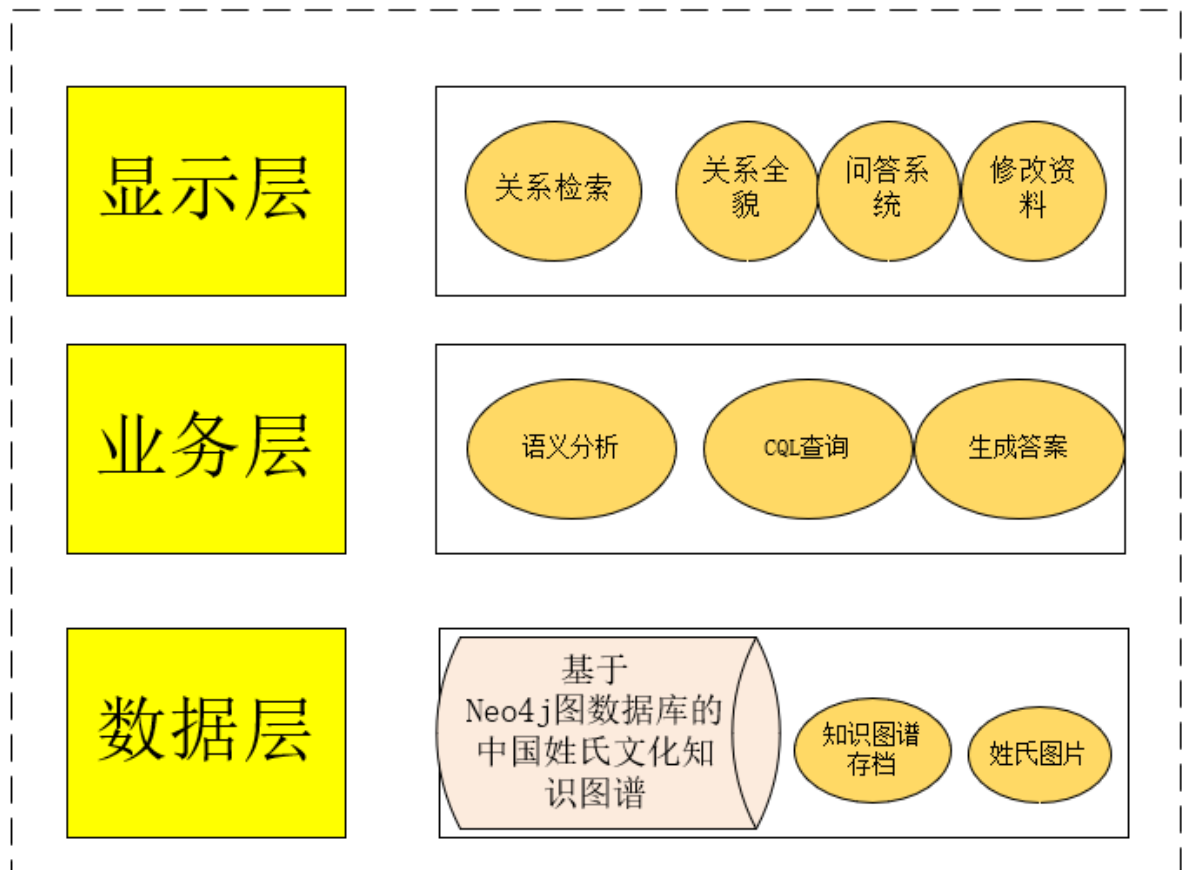


图 5-1 基于知识图谱的中国姓氏文化可视化系统 概要设计图

（1）数据层

数据层主要完成系统与 Neo4j 图数据库的交互，进行针对 Neo4j 图数据库的增删改查，负责对数据的持久化处理，并完成数据的备份与存档。本项研究将使用爬虫对网络数据爬取，将半结构化数据转化为结构化数据，存入文本文件作为原始数据库，当用户对中国姓氏文化知识图谱进行增删改查时，不仅会直接修改 Neo4j 图数据库里面的信息，也会生成 2 个 JSON 格式的文件，作为数据库存档，一个 JSON 文件存储姓氏与姓氏之间的关系，另一个 JSON 文件存储姓氏及其多个属性的属性值。

（2）业务层

业务层是位于数据层和显示层中间的一层，用来接受从显示层传来的用户要求，判断出用户的具体意图，然后把要求分门别类，然后按照用户要求，把需要做出的增删改查操作以及相关信息全部传递给数据层，数据层完成任务后，再接收从数据层发来的数据，再将来自数据层的数据做出进一步包装，使数据符合显示层的接受要求，从而完成该系统的业务。本项研究的业务层主要由 Flask 框架完成，Flask 框架扩展性强，比较灵活，特别适合小型项目。

（3）显示层

显示层将直接面对用户，以 web 界面呈现结果。负责与用户进行直接的交互，负责接收用户给出的要求，把用户的要求转变为数据发送给业务层，业务层与数据库将数据处理完成，再把数据按照显示层的要求发送回来，显示层将数据以图形、图片和文字的形式呈现给用户。本项研究的显示层主要由 HTML、CSS、JavaScript 完成，除了主页之外，主要有四个界面：关系检索、关系全貌、问答系统、修改资料。

5.2 系统详细设计与实现

本节是系统详细设计，将以中国姓氏文化可视化系统的 5 个页面为出发点，对中国姓氏文化可视化系统的详细设计进行仔细地介绍。

5.2.1 主页

打开控制台显示的网站即可进入主页，主页的 HTML 文件存储在 templates 文件夹内，命名为 index.html。与主页相关的所有 JavaScript、CSS、JPG 文件均存储在 static 文件夹下面，主页封面存储在 static 文件夹下面 images 文件夹里面。本页面及其涉及到的功能较为简单，不再赘述。点击“进入页面”，运行 app.py 里面的 search 函数即可进入系统。

5.2.2 关系检索

进入主页后，点击“进入页面”，将会进入中国姓氏文化可视化系统的关系检索功能（search.html）。

有两种搜索方式，

第一种，可以通过在搜索框输入姓氏并点击搜索。

第二种，直接点击关系检索页面右下方的动态球状物体中的汉字。

这两种搜索方式都将把一个字符串传递到 Flask 框架里面的 app.py 文件中。运行

search_name 函数，该函数接受了传递来的字符串，调用 query_graph.py 中的 query 函数处理该字符串，首先，检验字符串中的姓氏是否存在，则继续，否则输出“查无此姓”，然后，运行相应 CQL 查询代码，将查询结果合并为列表，再调用 get_json_data 函数转化为能够显示在网页的表现为知识图谱的图形，当前仅仅是表现为知识图谱的图形，仍需 jsonify 函数转化为真正的 JSON 格式才可以传入前端，最终显示为被查询姓氏的祖先和后裔的关系的图形。

5.2.3 关系全貌

该页面（all_relation.html）将显示知识图谱中所有关系。在创建知识图谱时，程序也将为知识图谱进行存档，其中一份关于所有姓氏与姓氏之间关系的存档就储存再 static 文件夹中，以 JSON 格式存储。

该页面的前端将调用该存档，显示在关系全貌这一页面中。

该页面的后端，在 app.py 中，还将再一次读取中国姓氏文化知识图谱的 Neo4j 图数据库，并更新一次存档。

5.2.4 问答系统

该页面（KGQA.html）将拥有问答系统的功能，输入内容，并点击“询问”。若不存在被问到的姓氏，将显示“查无此姓”，若存在被询问的姓氏，则在页面左侧，显示表达姓氏起源关系的知识图谱图形，在页面右侧显示该姓氏被问到的内容，例如介绍、历史来源、姓氏图腾、家族名人等。

点击“询问”后，输入的字符串信息发送到后端，开始执行 app.py 中的 KGQA_answer 函数，该函数接收到字符串信息后，先使用 KGQA 文件夹中的 ltp.py 中的 get_target_array 函数进行分词，通过分词，确定了用户的意图，查明了用户想要查询哪一个姓，接着把结果转化为数组，传递给 neo_db 文件夹里面的 query_graph.py 中的 get_KGQA_answer 函数，该函数查明用户想要查询哪一种属性，并且该函数运行 CQL 查询语句，进行查询，将查询结果合并为列表，再调用 get_json_data 函数转化为能够显示在网页的表现为知识图谱的图形，当前仅仅是表现为知识图谱的图形，仍需 jsonify 函数转化为真正的 JSON 格式才可以传入前端，最终显示为被查询姓氏的祖先和后裔的关系的图形，以及被查询的姓氏属性，以及该姓氏的图片。

5.2.5 修改资料

该页面（CRUD.html）可以执行一些命令，包括添加不同姓氏的关系、删除姓氏或姓氏属性、修改姓氏属性。执行这些命令时必须遵守该页面的小助手中的格式设置。否则将会失败。因为问答系统的语义分析何以较为轻松的设计出来，但针对修改资料功能的语义分析，涉及到深度学习，难度极大，暂时无法实现，只能设定一个标准格式

按照输入格式，输入命令，点击“执行”后，输入的字符串信息发送到后端，开始执行 app.py 中的 CRUD_answer 函数，该函数接收到字符串信息后，先使用 KGQA 文件夹中的 ltp.py 中的 get_CRUD_array 函数进行分词，通过分词，确定了用户的意图，查明了用户想要修改哪一个姓，接着把结果转化为数组，传递给 neo_db 文件夹里面的 query_graph.py 中的 get_CRUD_answer 函数，该函数查明用户想要执行怎样的命令，并且该函数运行相应的 CQL 查询语句，进行添加、删除、修改，将查询结果合并为列表，再调用 get_json_data 函数转化为能够显示在网页的表现为知识图谱的图形，当前仅仅是表现为知识图谱的图形，仍需 jsonify 函数转化为真正的 JSON 格式才可以传入前端，最终显示为与命令相关的姓氏的祖先和后裔的关系的图形，以及与命令相关的姓氏属性，以及该姓氏的图片。

5.3 系统展示

5.3.1 开发环境说明

基于知识图谱的中国姓氏文化可视化系统是在 python 环境下基于 Flask 进行开发，python 的 py2neo 软件包可以去连接 Neo4j 图数据库，并进行创建节点、关系、节点属性。可以在添加、删除、修改、查询节点、关系、节点属性。

5.3.2 系统功能展示

（1）主页：如图 5-2 所示：



图 5-2 基于知识图谱的中国姓氏文化可视化系统 主页

(2) 关系检索:

有两种搜索方式，第一种，可以通过在搜索框输入姓氏并点击搜索。第二种，直接点击关系检索页面右下方的动态球状物体中的汉字。这两种方式都可以查询到姓氏与姓氏之间的关系。如图 5-3 所示：

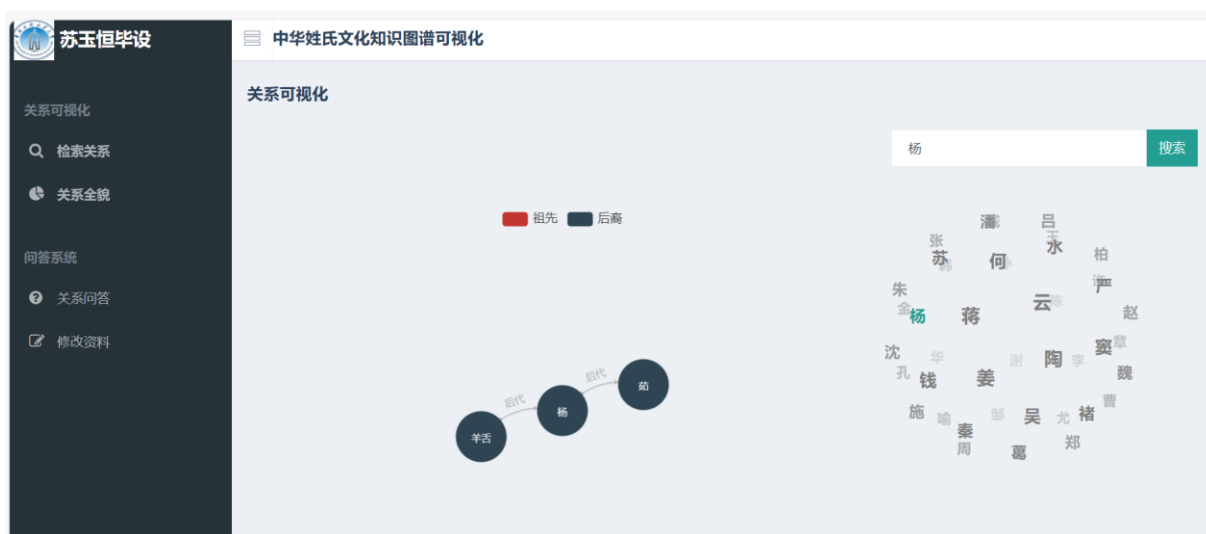


图 5-3 基于知识图谱的中国姓氏文化可视化系统 关系检索页面

(3) 关系全貌: 如图 5-4 所示:

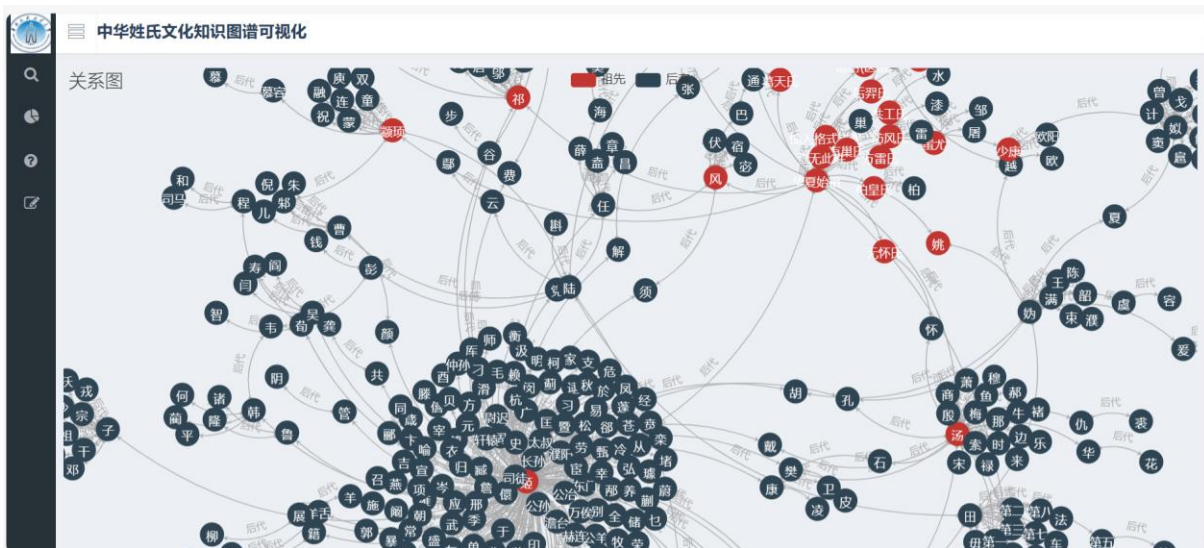


图 5-4 基于知识图谱的中国姓氏文化可视化系统 关系全貌页面

（4）问答系统：

如果被询问的姓氏真的存在，则在页面左侧，显示表达姓氏起源关系的知识图谱图形，在页面右侧显示该姓氏的图片以及被问到的内容，例如介绍、历史来源、姓氏图腾、家族名人等。如图 5-5 所示：



图 5-5 基于知识图谱的中国姓氏文化可视化系统 问答系统页面

（5）修改资料：

本页面可以执行一些命令，包括添加不同姓氏的关系、删除姓氏或姓氏属性、修改姓氏属性。当然，命令应当按照预定输入格式执行。

注意：执行“添加-杨-8”命令后，左侧的知识图谱可视化的图形，如图 5-6 所示：



图 5-6 基于知识图谱的中国姓氏文化可视化系统 修改资料页面 1

注意：执行“修改-8-历史来源-出自杨姓”命令后，左侧的知识图谱可视化的图形和右侧资料，如图 5-7 所示：



图 5-7 基于知识图谱的中国姓氏文化可视化系统 修改资料页面 2

第 6 章 系统测试

本节进行对于基于知识图谱的中国姓氏文化可视化系统的测试。测试分为两块，功能性测试和非功能测试。

6.1 功能性测试

基于知识图谱的中国姓氏文化可视化系统的各个功能模块的测试方法、标准、和结果如表 6-1 所示：

表 6-1 基于知识图谱的中国姓氏文化可视化系统 功能测试结果

功能模块	测试方法	测试标准	测试结果
主页显示	在 Google、Edge、QQ 浏览器等浏览器打开本系统	正常显示界面且点击“进入页面”后，进入关系检索功能	与预期结果完全相同
关系检索	输入姓氏并搜索、点击姓氏	正常显示姓氏与姓氏之间的关系	与预期结果完全相同
关系全貌	进入页面	正常显示所有的姓氏与姓氏之间的关系	与预期结果完全相同
关系问答 1	询问某姓氏的介绍	正常显示该姓氏的介绍	与预期结果完全相同
关系问答 2	询问某姓氏的历史来源	正常显示该姓氏的历史来源	与预期结果完全相同
关系问答 3	询问某姓氏的家族名人	正常显示该姓氏的家族名人	与预期结果完全相同
关系问答 4	询问某姓氏的分布	正常显示该姓氏的分布	与预期结果完全相同
关系问答 5	询问某姓氏的姓氏图腾	正常显示该姓氏的姓氏图腾	与预期结果完全相同
修改资料 1	添加姓氏与姓氏之间的关系	正常显示新添加的关系	与预期结果完全相同

修改资料 2	删除姓氏	问答系统中查不到该姓氏	与预期结果完全相同
修改资料 3	修改姓氏属性	正常显示新修改的姓氏属性	与预期结果完全相同

6.2 非功能性测试

非功能性测试主要用于测试基于知识图谱的中国姓氏文化可视化系统的鲁棒性，本项目研究要进行安全性测试、兼容性测试以及响应时间测试。如表 6-2 所示：

表 6-2 基于知识图谱的中国姓氏文化可视化系统 非功能测试结果

测试内容	测试方法	测试标准	测试结果
系统安全性	提交错误姓氏	显示“输入错误”	与预期结果完全相同
兼容性	在 Google、Edge、QQ 浏览器等多种浏览器打开本系统	系统运行正常	与预期结果完全相同
响应时间	记录 20 次修改资料执行命令的平均响应时间	小于等于 3 秒	与预期结果完全相同

第7章总结与展望

7.1 总结

本文基于对中国姓氏文化的研究，设计并实现了一个中国姓氏文化可视化系统，该系统通过构建姓氏知识图谱，实现了对姓氏文化的可视化展示与探索。在系统的设计与实现过程中，通过收集整理姓氏数据，构建知识图谱，并设计相应的可视化界面，成功地实现了对中国姓氏文化的直观展示和探索。

在本文的研究中，发现了中国姓氏文化的丰富多样性，每个姓氏背后都蕴含着深厚的历史文化底蕴。通过对姓氏起源、分布、文化特点等方面的分析，可以更加全面地了解中国姓氏文化的内涵与特点，促进姓氏文化的传承与发展。

在系统的实现过程中，采用了知识图谱的技术手段，将姓氏相关的知识进行了组织和关联，通过可视化的方式呈现给用户。用户可以通过系统，轻松地浏览、搜索和探索各种姓氏的相关信息，从而深入了解中国姓氏文化的内涵与特点。

本研究主要内容如下：

（1）知识图谱构建：收集各类网络平台数据，使用网络爬虫分类整理姓氏文化信息，提取数据并进行清洗和知识抽取，生成知识三元组存储在文本文件中，最终构建中国姓氏文化知识图谱存储于 Neo4j 图数据库。

（2）关系检索系统：基于构建好的知识图谱，开发完整的姓氏文化关系检索系统。用户可在浏览器中输入姓氏或点击姓氏，系统将组合查询语句并从知识图谱中检索，将结果转化为 JSON 格式并通过 Flask 框架传输至浏览器展示。

（3）姓氏起源关系展示：用户可通过浏览器访问，从知识图谱中提取 Neo4j 数据库的全部关系数据并展示在 Web 端，呈现完整的中国姓氏起源关系全貌。

（4）姓氏文化问答系统：分析用户提问，设定问题模板并进行中文分词和实体识别，从知识图谱中匹配相应查询语句并生成 JSON 文件，通过 Flask 框架封装展示在浏览器中，实现姓氏文化问答功能。

（5）知识图谱修改功能：分析用户修改命令，设定修改模板并进行中文分词和实体识别，匹配相应修改命令语句模板并修改知识图谱，生成相应 JSON 文件并通过 Flask 框架展示在浏览器中，实现姓氏文化知识图谱修改功能。

7.2 展望

尽管本文已经设计并实现了一个基于知识图谱的中国姓氏文化可视化系统，但仍然存在一些可以改进和完善的方面，同时也存在一些未来可以开展的工作：

当前系统的数据源主要来自古诗文网，《百家姓》数据作为起始点进行可视化展示。未来可以进一步完善数据资源，增加更多的姓氏数据，并丰富数据的内容和质量，以提升系统的可用性和可靠性。

除了姓氏知识图谱展示外，可以进一步拓展系统的功能模块，如姓氏起源探索、姓氏分布地图展示、姓氏文化特点分析等，以满足用户多样化的需求，并提升系统的吸引力和实用性。

在系统设计方面，可以进一步提升用户体验，优化界面设计和交互方式，使用户能够更加轻松

地使用系统，提高用户满意度和使用率。

除了在学术研究领域的应用外，还可以进一步拓展系统在其他领域的应用，如文化旅游、教育培训等领域，为相关行业提供姓氏文化方面的数据支持和服务，推动姓氏文化的传承和发展。

参考文献

- [1] 族谱网. (2023). <https://zhuanlan.zhihu.com/p/654821433>
- [2] 古诗文网. so.gushiwen.cn/guwen/book_46653FD803893E4FE03CBAE75DE61AB8.aspx
- [3] 宗洋. 姓名背后的文化内涵[J]. 中国民族博览, 2024, (01): 41-43.
- [4] 刘峤, 李杨, 段宏, 等. 知识图谱构建技术综述[J]. 计算机研究与发展, 2016, 53(03): 582-600.
- [5] 徐增林, 盛泳潘, 贺丽荣, 等. 知识图谱技术综述[J]. 电子科技大学学报, 2016, 45(04): 589-606.
- [6] 张朦. 基于知识图谱的中华典籍问答系统研究[D]. 中北大学, 2023. DOI:10.27470/d.cnki.ghbgc.2023.000112.
- [7] 陈志波. 基于知识图谱的心理咨询问答系统分析与设计[D]. 阜阳师范大学, 2023. DOI:10.27846/d.cnki.gfysf.2023.000209.
- [8] 赵同明, 钱佳琛, 王翔, 等. 健康饮食知识图谱和问答系统构建研究[J]. 现代信息科技, 2024, 8(06): 7-10. DOI:10.19850/j.cnki.2096-4706.2024.06.002.
- [9] 宋强, 宋子豪, 刘江杰, 等. 海洋综合减灾知识图谱及问答系统的研究与实现[J]. 信息与电脑(理论版), 2024, 36(03): 130-132.
- [10] 李成林, 赵珍威, 李国厚, 等. 基于知识图谱的农作物良种问答系统的设计与实现[J]. 河南科技学院学报(自然科学版), 2024, 52(01): 48-59.
- [11] 王博, 群诺, 王京博, 等. 基于知识图谱的西藏旅游问答系统设计与实现[J]. 信息技术与信息化, 2023, (12): 181-184.
- [12] 赵伟航. 基于知识图谱的糖尿病智能问答系统关键技术研究及应用[D]. 西京学院, 2023. DOI:10.27831/d.cnki.gxjxy.2023.000075.
- [13] 侯礼灏. 基于知识图谱的法律问答系统的设计与研究[D]. 西京学院, 2023. DOI:10.27831/d.cnki.gxjxy.2023.000108.
- [14] 刘星翔. 基于知识图谱的医疗问答系统研究与实现[D]. 南昌大学, 2023. DOI:10.27232/d.cnki.gnchu.2023.004678.
- [15] 赵赛, 杨婉霞, 王巧珍, 等. 基于马铃薯病虫害知识图谱的问答系统[J]. 农业工

- 程, 2023, 13(08):29-37. DOI:10.19998/j.cnki.2095-1795.2023.08.006.
- [16] 侯礼灏, 迟玉红, 谢虹铭. 基于知识图谱的法律问答系统研究[J]. 电脑编程技巧与维护, 2023, (08):106-108. DOI:10.16184/j.cnki.comprg.2023.08.010.
- [17] 王博, 王泽辉, 张坦, 等. 基于知识图谱的西藏文物问答系统构建与实现[J]. 电脑知识与技术, 2023, 19(18):30-32. DOI:10.14004/j.cnki.ckt.2023.0854.
- [18] 张浪浪. 基于知识图谱的高考志愿填报辅助问答系统的设计与实现[D]. 浙江师范大学, 2023. DOI:10.27464/d.cnki.gzsfu.2023.002084.
- [19] 吴涛. 基于知识图谱的肝脏疾病问答系统的研究与实现[D]. 中北大学, 2023. DOI:10.27470/d.cnki.ghbgc.2023.001462.
- [20] 陈悦, 刘则渊. 悄然兴起的科学知识图谱[J]. 科学学研究, 2005, (02):149-154. DOI:10.16192/j.cnki.1003-2053.2005.02.002.
- [21] 刘星宇. 基于知识图谱的东周列国问答系统的研究与实现[D]. 中北大学, 2023. DOI:10.27470/d.cnki.ghbgc.2023.001033.
- [22] 汤澜. 基于知识图谱的医疗问答系统设计与实现[D]. 西安石油大学, 2023. DOI:10.27400/d.cnki.gxasc.2023.000554.
- [23] 庞文婷. 基于知识图谱的玉米育种智能问答系统研发[D]. 西北农林科技大学, 2023. DOI:10.27409/d.cnki.gxbnu.2023.000457.

致 谢

行文至此，落笔为终。四年的学习生活，始于金秋，终于盛夏，简短的八个字很难去描述这一段很长的路。至此，学生生涯也接近尾声，我也曾欢声笑语，也曾跌跌撞撞，纵有千般不舍，也即将奔赴下一场山海。在人生中的青春年华，收获了最宝贵的财富，感谢所有相遇！

岁月如歌，师恩难忘。首先感谢我的毕业设计指导老师，从论文的选题到提纲设计，再到最终定稿，都离不开老师的帮助和指导。老师为人谦和，知识储备渊博，实践经验丰富，治学态度严谨。感谢您的耐心和包容。在写毕业论文的几个月中，给予我莫大的帮助，在一次次组会和交谈中，不仅鼓励我不断探索新知识，激发更好的灵感，同时还让我学会了如何以积极乐观的态度去审视世界，这些都对我今后的道路起到重要作用。最后在这里感谢曾经所有的授课老师，使我能够学一点知识、明一丝道理、寻一条大道。

山河不足重，重在遇知己。感谢我的朋友们，你们在生活和学业上给我很多的帮助，在科研遇到难题时，为我指点迷津；在忙碌之余，会陪我一起放松休闲；在心情烦闷之时，会陪我彻夜长聊。你们的陪伴与支持使我的研究生生活充实且快乐，在此衷心祝愿我的室友和朋友们前程似锦。

父母之爱子，则为之计深远。感谢我的家人，在过去的二十多年里，你们是我的港湾，替我遮风挡雨；又是我启蒙老师，教会我为人处事的道理；更是一束光，为我指亮前程，亦照亮归途。养育之恩，无以为报，我会继续努力，让你们过上更幸福的生活。

愿你们永远健康快乐，万事顺遂。

附 录

附录包含外文原文（近 5 年与毕业设计（论文）题目相关的期刊论文）、外文译文（不少于 2000 字）、任务书、开题报告、设计书、图表、程序等（附录部分装订顺序按照此顺序进行）。

外文文献原文

A Review of Knowledge Graph Research in Military Domain

Xudong Guo , Linxiu Chen , Yuan Li , Qinglin Wang , Weili Guan

Abstract. With the development of information technology as well as the continuous improvement of military informatization level, the complex and changing battlefield situation put forward a huge challenge on the military's massive data processing capability. This paper presents the technical framework related to the construction of knowledge graph in military field, summarizes different aspects of the application of knowledge graph in military field, and reveals the directions for improvement in view of the current situation of the application of knowledge graph in military field.

Keywords: Military Domain, Knowledge Graph, Ontology

1. INTRODUCTION

In the context of modern intelligent warfare, the widespread application of battlefield information technology and sensing technology has led to a dramatic expansion and interconnection of battlefield space and an explosion of data. Intelligence, reconnaissance, surveillance and massive data have "4V" characteristics: large scale (Volume), very fast change (Velocity), a variety (Variety), and value (Value). The huge amount of data makes the battlefield situation assessment and analysis have far needs exceeded the traditional computer-aided assessment and decision processing capabilities based on manual experience and cognition[1]. It is desirable to introduce powerful knowledge description and management methods into military application.

The knowledge graph originated from the concept of semantic web and was first proposed by Google in 2012[2]. Knowledge graphs integrate technical approaches

such as databases, information retrieval, data mining, natural language processing, and graph computational analysis[3]. Knowledge graphs can integrate knowledge of a subject area and provide services such as visual presentation. The application of knowledge graphs has significantly improved the information retrieval capability and demonstrated the powerful function of knowledge graphs in information integration.

As the forefront of science and technology in the military field , the construction of proprietary knowledge graphs can largely improve the efficiency of operations, mainly in the following aspects[4]: knowledge graphs can integrate a large amount of scattered and isolated intelligence; knowledge graphs can allow computers to understand the semantics of intelligence; knowledge graphs can correlate a large amount of intelligence at the semantic level; and knowledge graphs can mine new knowledge in correlated intelligence. Therefore, it is necessary to apply knowledge graphs to the military field.

A knowledge graph consists two levels: the schema layer and the data layer[5]. The schema layer is used to regulate the data layer and to formulate the rules of data types and hierarchies followed by the data layer, and it is the core of the knowledge graph. The schema layer usually uses an ontology library to manage the schema layer of the knowledge graph. The data layer stores the knowledge specific to the knowledge graph and can be seen as an instantiation of the schema layer.

In this paper, the construction of knowledge ontology in military domain is introduced, and the knowledge service of knowledge graph in military domain in three directions, and conclusion with a summary and outlook.

2. CONSTRUCTION OF KNOWLEDGE ONTOLOGY IN MILITARY FIELD

Knowledge graphs are constructed in two ways, the top-down and bottom-up approaches[6]. The top-down approach usually starts by defining a data schema for the knowledge graph, which is constructed from the topmost concept and gradually refined downward to form a well-structured taxonomic hierarchy, and then entities are added to the concept. The bottom-up approach organizes entities

inductively to form the bottom-level concepts and then gradually abstracts upward to form the top-level concepts.

The military domain is a typical specific domain with a well-defined conceptual hierarchy, however, part of the expertise exists in the minds of domain experts and may not be available inductively and abstractly from the data. Due to the confidentiality of military data and the special nature of military equipment, many documents and information show absolute hindrance to the establishment of knowledge graph, resulting in less structured data and difficult data collection. Therefore, the knowledge graph in this domain is more suitable for a top-down approach at the beginning of its construction. After the basic structure (or conceptual hierarchy) of the knowledge graph has been established and the number of its category nodes or relationships has reached a certain scale, and the bottom-up approach can be used to expand the instance and attribute data for the graph[6]. With above analysis, the ontology should be constructed first due to the special nature of the knowledge graph of military equipment.

2.1. Ontology

Ontology is originally a philosophical definition[7], which is the essential description of objective things. With the development of artificial intelligence (AI), knowledge graph (KG) and other fields, the concept of ontology has been gradually emphasized and widely used in these fields. In 1991, Neches proposed the definition of ontology, "the basic terms and relations that constitute the vocabulary of the domain in question, and the rules that define the extension of these terms using these terms and relations" [8]. In 1993, Gruber proposed that "ontology is a shared, conceptualized, and formalized specification"[10], and the definition was widely accepted by academics.

In the domain of knowledge graph, an ontology is a specification obtained by abstracting and constraining different entity concepts, attributes and their interrelationships in the domain, and is the most central part of the knowledge graph, which is mainly described as a tree structure[11]. Ontologies are the

conceptual templates of structured knowledge bases, and the knowledge bases formed through ontology repositories are not only stronger in hierarchical structure but also less redundant[12]. The main purpose of building ontology is to define a clear data schema for the knowledge graph, and the correct data schema can ensure that the entities in the knowledge graph have strong systematic relationships, which is the basis for building the knowledge graph[13].

2.2. Ontology Construction Principles

Ontology construction is the main way to construct knowledge models following the following five accepted principles[13][14], the principles of Clarity and Objectivity, Completeness, Coherence, Maximum Monotonic Extendibility, and Minimal Ontological Commitments.

On clarity and objectivity, the ontology must be able to effectively convey the meaning of the defined term with a definition that is objective and independent of context. On completeness, the definitions of terms given in the ontology are complete and fully express the meaning of the described terms.

The axioms defined by the ontology, and the documents that describe them in natural language, should be consistent.

On coherence, the inferences drawn from the terms are compatible with the meaning of the terms themselves, i. e., they support reasoning that is consistent with their definitions and does not create contradictions; the axioms defined and the documentation illustrated in natural language should also be consistent.

On maximum monotonic extendibility, adding generic or specialized terms to an ontology does not require modification of its existing concept definition and content, and supports the definition of new terms based on existing concepts.

On minimal ontological commitments, ontology commitments should be minimal and treat modeling objects with as few constraints as possible. Ontology commitment means a consensus on how to use the shared vocabulary in a consistent and compatible way. In general, ontology conventions sufficiently satisfy specific knowledge sharing needs, which can be ensured by defining the least constrained axioms and defining only the vocabulary needed for communication.

2.3. Ontology Construction Methods

There are three main methods of ontology construction, manual construction, reuse of existing ontologies (semi-automatic construction) and automatic construction[15]. Manual construction methods mainly determine the knowledge content and relationships through domain experts, so that the ontology quality and accuracy are high. The military domain requires high knowledge completeness and accuracy, so the manual construction method is applicable to the construction of military knowledge ontologies. Adopting the idea of manual construction of general ontology, military ontology can apply top-down manual construction, firstly constructing the top-level ontology of military knowledge, starting to construct the knowledge framework of each sub-domain on this basis, and gradually completing the construction of the whole military knowledge ontology[16].

There are five relatively influential ontology manual construction methods in China and abroad[17][18]: the skeleton method, the TOVE method, the IDEF5 method, the Methontology method, and the seven-step method. Compared with the other methods, the seven-step method is relatively mature. It includes seven steps of identifying the domain and scope of expertise, examining the possibility of reusing existing ontologies, listing important terms in the domain, defining classes and class hierarchy levels, defining class attributes, defining attribute facets, and creating instances[19]. However, the seven-step approach also has some limitations; this method does not have a complete life cycle[20] and lacks a model revision process for responding to requirements adjustments or feature additions in the later stages of development.

2.4. Military Field Ontology Structure

In the field of military equipment, the literature[13] divides military equipment into eight categories, such as artillery, ships and vessels, and flying vehicles, and defines attributes such as lethality and maximum speed for equipment entities. This equipment knowledge system is biased toward the scientific description of military equipment and ignores the expression of knowledge oriented to military combat needs. Liu et al [21] designed a

demand-based spiral feedback method to construct a military equipment knowledge model, defining one top-level core concept military equipment, four related concepts such as combat activities, and seven generic concepts such as country, in which equipment is divided into eight major categories and 148 subcategories. The model includes 202 entity attributes, 6 entity relationships, covering 5,800 weapon-type entities, 18 combat activities, and 86 countries.

From the above classification, it can be seen that the focus of ontology construction for different tasks in the military domain varies, but they can be basically divided into the following aspects.

3. CONSTRUCTION OF KNOWLEDGE GRAPH IN MILITARY DOMAIN

Based on the knowledge ontology, the knowledge graph can be constructed. The construction and application of military domain knowledge graph is a system engineering, and its construction process is divided into six stages, which can be called the whole life cycle of domain knowledge graph, and the key technical processes of its design mainly contain knowledge representation, knowledge storage, knowledge extraction, knowledge fusion, knowledge evolution, and knowledge application[22].

3.1. Knowledge Representation

Knowledge representation studies how to achieve the modeling of things in the real world and things related to each other, providing data logical information that conforms to human expressions and enabling barrier-free communication between humans and computers[23]. The knowledge representation in the knowledge graph defines the basic cognitive framework of the domain, clarifies the basic concepts and the semantic associations between concepts, and provides the basic machine cognitive skeleton and data structure to achieve a rational organization of knowledge[24]. Knowledge graphs are mainly based on the representation of semantic networks, and knowledge representation is performed in the form of triad (three elements set).

In the military domain, some application scenarios add expression elements such as spatio-temporal dimensions that go far beyond the simple associated facts that can be expressed by the triad of knowledge graphs. Although fine-grained knowledge representation is one of the strong requirements for military domain applications, it does not mean that all application scenarios need fine-grained representation, and the finer the granularity the stronger the expression capability, the higher the acquisition cost.

These problems can be addressed in ways that enhance the representation of complex knowledge.

3.2. Knowledge Storage

The constructing the knowledge graph depends on a framework for storing the knowledge graph. The construction, representation and application of the knowledge graph on a perfect knowledge graph storage framework will greatly improve the efficiency of using the knowledge graph[4].

In the military field, the knowledge involved in the knowledge graph has the characteristics of multi-source, heterogeneity, time-series, real-time, redundancy, strong correlation, and closed-loop. At the same time, the applications of complex scenarios need to meet the requirements of high efficiency query and high concurrency[22].

To address these problems, Wu et al[35] used the open-source graph database system Neo4j graph database to store domain knowledge graphs oriented to simulation derivation, and used the convenient graph query language Cypher for data query to solve the problem of constructing knowledge graphs for text-based content. Xing et al[22] proposed a framework that employs multiple types of storage media in order to achieve hybrid storage of military knowledge, which includes six parts: source data storage manager, data processor, data extraction engine, message queue, storage and indexing interface and subscription service, including relational database, graph database, in-memory database, file

indexing system, etc. The unstructured and semi-structured data are distributed in the knowledge graph in HDF format and RDF format distributed graph storage and annotation, effective fusion of multi-source heterogeneous data through memory database and relational database, unified storage management of multi-source heterogeneous knowledge, and storage of video, picture, sound and other documents using file storage system. Ge et al[4] proposed a chunking management framework based on multiple constraints to optimize multiple factors affecting query efficiency. The framework improves the query efficiency by optimizing the minimum number of cut edges and achieves the purpose of load balancing by optimizing the total amount of communication.

3.3. Knowledge Extraction

The knowledge extraction is to populate the knowledge instances of the knowledge framework and to identify, discover and extract concepts, types, facts and their related relationships, constraint rules, etc. from the heterogeneous and complex data using relevant techniques. Knowledge extraction can be divided into entity extraction, relationship extraction and attribute extraction.

In the military field, large military information systems of all levels and types generate a large amount of military data under the complex usage environment of multiple tasks, and such data have the characteristics of different data sources and diverse data formats, which require key means to improve the accuracy and recall rate of knowledge extraction designed for the characteristics of different data. For example, military databases have the problems of inconsistent data representation, data cannot complete aggregation and intelligent retrieval; combat instruments have high complexity, many special data and strong data regularity, which need to inject corresponding military rules in advance. In addition, due to the extreme seriousness of application scenarios in the military field, the quality requirements for knowledge acquisition are very demanding, and the knowledge acquisition scheme has to

balance the requirements and costs to avoid the problems of missing, wrong, and obsolete information in the knowledge base[35][37].

In solve these problems, Xing et al[22] proposed the principle of minimizing cost while ensuring quality: First, specify the data sources, and try to select data sources with better quality and higher structured degree according to the principle of taking refuge in simplicity, such as “Globe News”, “The Encyclopedia Entries”, and “The Great Military Word Sea”, etc. After that, gradually expand the rule set through unstructured data and conduct knowledge of cross-domain migration and reuse similar domain knowledge.

For unstructured data extraction methods, supervised machine learning methods are generally used. Liu[38] et al. used HanLP, a Chinese classification system in the open-source Java toolbox, for entity extraction and a maximum entropy model combining CRF and syntactic analysis tree for relationship extraction. Zhao et al[39] used a bidirectional LSTM-CRF model for entity extraction, syntactic analysis and hierarchical Dirichlet process (HDP) clustering to achieve relationship word extraction, and a bidirectional RNN model to implement an entity relationship extraction algorithm.

3.4. Knowledge Fusion

Due to the diverse data sources and different extraction methods, there may be a large number of redundancy, noise, non-completeness and uncertainty in the knowledge obtained by knowledge extraction. At the same time, there is no hierarchy and clear logical relationship between data, and such problems cannot be solved by the cleaning session of knowledge extraction, and there is an urgent need to build an entity-relationship representation model based on integration[22]. Therefore, it is necessary to complete semantic standards and links for massive heterogeneous data resources as well as entity disambiguation and entity alignment for different domains, data patterns, and application requirements to achieve cross-domain semantic fusion of multiple domain mappings and form a global unified knowledge base, i.e., knowledge fusion[39][41]. The schema layer of the military domain knowledge graph is constructed with manual

assistance and high accuracy; the main thing to be dealt with is the fusion of the data layer. [42]

For the problem of data-level knowledge fusion, techniques such as entity disambiguation, entity alignment, attribute fusion and value normalization can be used. Entity disambiguation inferred the specific entity to which the current word refers by measuring the correlation between candidate entities and the context. Entity alignment is to disambiguate multiple feature data in multiple graphs by representing them as the same entity. Wu et al[35] used cosine similarity to cluster entities, thus realizing the functions of entity disambiguation and alignment. A method based on quality estimation is used with the best-valued knowledge source as the valid knowledge of conflicting entities, thus realizing functions such as attribute fusion.

3.5. Knowledge Computing and Evolution

In order to improve the practical application value of knowledge graphs in the military domain, it is necessary to mine the large number of implicit relationships that exist between entities.

In the military domain, the entities of the knowledge graph are more densely distributed, and the related knowledge of a single entity covers a wide area, so it is necessary to complete long-distance inference by fast loading of incremental knowledge and rules for the characteristics of deeper granularity and depth of its knowledge representation, in order to guarantee the extension and correctness of inference and to mine more related information and knowledge at the semantic level[43]. At the same time, there is also the problem of small samples in military domain knowledge graphs, and the inference problem of knowledge graphs with data sparsity characteristics needs to be solved. In addition, the construction of military knowledge map is a process of constant iterative updating, and the operational schemes, strategies, and means are constantly changing, which requires timely confidence assessment of knowledge and the improvement of new knowledge.

To address the problems such as knowledge inference, Xing et al[22] proposed a knowledge inference method with a hybrid remote supervision and deep learning under collaborative filtering mechanism, which intervenes in the knowledge inference process to build a dataset by adding external knowledge through remote supervision, and constructs and trains a bidirectional LSTM neural network to extract new triads through a deep learning method, for improving the performance of knowledge inference, while solving the small sample problem and achieving the effect of knowledge map complementation. Wu, et al[35] used a probabilistic logic inference method incorporating numerical computation to obtain candidate answers first and then perform probabilistic logic inference, thus realizing the function of entity relationship complementation. Ge et al [4] proposed an exact query framework based on shortest path and a fuzzy query framework based on similarity calculation.

For the problem of updating and maintaining the knowledge graph, Zhao et al[39] gave an update algorithm process framework for new data, which determines whether to enter the database by template matching and other algorithms, and judges whether to perform knowledge update by means of confidence calculation. Ge et al[4] gave a semi-automated knowledge graph updating process. Firstly, distributed crawlers is used to collect data sources with high confidence level, on which the entities are extracted, while the relationships involved in the field of military intelligence can be manually operated due to the relatively fixed and clear relationships. The newly generated triads are validated by multiple data sources, and finally the existing triads within the military knowledge graph are updated and maintained. The relationship determination in this process can be achieved by using machine learning methods plus manual review, which can improve the efficiency of knowledge base updating.

3.6. Knowledge Application

On knowledge applications, technical maturity should be considered, and carefully selected for application scenarios. Knowledge graph, as an open knowledge management mechanism oriented to human-computer collaboration, has

specific values and costs, and it is not always the best solution for different application problems[44]. The application scenarios of military domain knowledge graphs are much more complex than those of general-purpose knowledge graphs, and most of the current domain knowledge systems are built with the basic concepts and entities of domain data in mind, which has a natural defect in understanding the user requirements in complex application scenarios. How can we better design complex combat application scenarios and bridge this semantic gap is a problem that needs to be focused on and solved in the knowledge application session. This section will be specifically developed in the next section.

4. KNOWLEDGE SERVICES FOR KNOWLEDGE GRAPH IN MILITARY DOMAIN

4.1. Information Service

Based on the knowledge graph, basic information services such as data visualization and intelligent Q&A can be provided. Zhao et al[39] used Javaneese, Tomcat graph display and Bootstrap technology to design a knowledge graph retrieval and display service of topic \rightarrow event \rightarrow entity in a hierarchical manner with regional partitions and topics of concern as aggregation points, which improved the efficiency of sub-domain related hotspot retrieval, analysis and collation. On the basis of the constructed knowledge graph of military equipment, Dou et al[44] implemented a knowledge Q&A based on template matching[46].

4.2. Intelligence Reconnaissance and Mining

Based on the characteristics of knowledge graph supporting fast retrieval, efficient storage, and long-chain reasoning, knowledge graphs can accomplish intelligence collection and collation, etc. in intelligence reconnaissance mining[47].

Palantir was the first company to use knowledge graph technology in the defense sector to provide intelligence analysis services for counterterrorism and combat operations, has played an important role in both government and military decision-making and command. The Defense Advanced Research Projects

Agency (DARPA)[48] started the XDATA program in March 2012 to develop computational techniques and software tools for analyzing large amounts of semi-structured and unstructured data, and the Deep Exploration and Filtering of Text(DEFT) program, launched in the same year, more explicitly proposes to use deep learning techniques to uncover the implicit, practical features of large volumes of structured text, as well as the ability to further integrate the processed information, and on this basis, to use these techniques in operational assessment, planning, and forecasting to support decision making. DARPA's "AIDA" project, released in 2017, investigates the mapping of automatically acquired knowledge from multiple media sources to common semantic representations, known ontology extension techniques, and other techniques to transform fragmented information and achieve deep knowledge mining.

4.3. Battlefield Situational Awareness

Based on the powerful information integration and knowledge mining capabilities of knowledge graphs, in battlefield situational awareness, knowledge graphs can facilitate military decision makers to better understand the battlefield situational awareness by improving the performance of knowledge fusion algorithms and enhancing the operational efficiency of knowledge processing tools.

Yi et al[49] proposed an intelligent recognition reasoning framework, which first constructs a knowledge graph database using knowledge graph and graph database technologies, and then constructs multiple intelligences with domain knowledge reasoning capabilities, and uses classifier reasoning techniques for integrated recognition reasoning, which can achieve the function of intelligent target recognition.

4.4. Combat Command and Control

Based on the strong knowledge expression ability of knowledge graph, the accuracy and rapidity of knowledge extraction, in the aspect of combat command and control, knowledge graph can help for efficient command and rapid decision making.

In 2012, the U. S. Army's Data to Decision (D2D) program used key technologies such as knowledge graph to efficiently distill knowledge from data, obtain the information warfighters and decision makers need to guide operations, and integrate it with relevant contextual and situational information through enhanced data fusion to provide a clear picture of threats, options, and consequences [47]. Simulation push is the process of rehearsing and analyzing the deployment, operational objectives, or outcomes of different phases of an operational scenario before or during the execution of the operation, in accordance with the operational intent, sequence, and process specified in the operational scenario. Knowledge graph also provides knowledge and data support for simulation rehearsal and provides structured storage for massive heterogeneous data, thus improving the efficiency of the simulation rehearsal system[35].

4.5. Cyberspace Security

Based on the capabilities of knowledge graph data retrieval, data management, and knowledge inference, knowledge graphs have important significance in enhancing network situational awareness, improving network offensive and defensive capabilities, and maintaining national cyberspace security in the context of cyberspace security[49].

The National Security Agency's (NSA) Prism program extracts the call records of millions of users from telecommunication companies every day and obtains information from the central servers of nine web giants, including Microsoft, Google, Yahoo, Facebook, and Apple, to perform correlation analysis and reasoning through information fusion and knowledge acquisition technologies, and to generate high-quality intelligence products. The information fusion and knowledge acquisition technologies can be used to correlate, analyze, and reason, and produce high-quality intelligence products that pose a great threat to the security of other countries' cyberspace[50]. The U. S. Army's "Cyber Attack Automation Unconventional Sensor Environment" program seeks to develop predictive methods for cyber attack behavior and effective methods for detecting

impending cyber phenomena to help cyber defenders respond to cyber-attacks, develop and validate unconventional, multidisciplinary sensing technologies that can predict cyber-attacks and perform existing advanced intrusion detection capabilities[51].

5. SUMMARY AND PROSPECT

This paper summarizes the technical framework related to the construction of knowledge graphs in the military domain and shows different aspects of the application of knowledge graphs in the military domain. It can be seen that knowledge graphs provide a powerful method for improving intelligence in the military domain by employing semantic networks of enormous scale to describe concepts, entities, attributes and the relationships between them in the objective world.

In order to fully utilize the advantages of knowledge graph technology in areas such as resource integration and intelligent computing in the military domain, research can be enhanced in the following areas.

First, the basic technology research of knowledge graph should be strengthened. At present, most of the research on knowledge graph in the military field is still in the exploration stage, and there is still a certain distance from the real landing. Research on key technologies of knowledge graph should be improved by making full use of artificial intelligence and other related technologies, which can improve the representation capability, algorithm efficiency and intelligence level of knowledge graph.

Second, the digital level of the army should be improved. Command information systems are basically still in the stage of auxiliary "labor" rather than auxiliary "intelligence", and they cannot give good solutions to the problems requiring high intelligence such as situational understanding, decision support or confrontation rehearsal, which are urgently needed by commanders [11]. The rapid application of knowledge graph to battlefield situational awareness and operational command and control can greatly improve the operational level and operational efficiency of the military.

Third, the application of knowledge graph in the military field should be expanded. Knowledge graph technology has made significant progress in digital medical, digital library, search engine and other fields, but its application in the military field is still relatively limited. Giving full play to the application of knowledge graph in the military field will yield great military benefits.

Finally, improve the application capability based on knowledge graph. With the increasingly severe of current international situation and complicated of international relation, it is difficult to speculate on combat motives and analyze the battlefield situation clearly based on a single knowledge graph. Therefore, we need to establish complex situational analysis models and methods with integrating structured knowledge from multiple fields and combining multifaced information, which is an important way to better play the role of knowledge graph in the field of battlefield. So it puts higher requirements for the application system based on knowledge graph, such as comprehensive information collection capability, efficient information fusion capability, and powerful information reasoning capability.

Knowledge graph will continue to be a hot issue for cutting-edge research in the coming period. It is believed that in the near future, through the combination of big data technology and artificial intelligence technology, knowledge graph can add to the improvement of military level.

Acknowledgements

This work was supported by Open Foundation of Science and Technology on Complex System Control and Intelligent Agent Cooperation Laboratory (No. 212002).

REFERENCES:

- [1] Zhu Feng; Hu Xiaofeng, Review of deep learning-based battlefield

- situational assessment and research outlook, [J], Military Operations Research and Systems Engineering, 2016, 24-29, 24-29. (In Chinese)
- [2] Pujara J, Miao H, Getoor L, et al. Knowledge graph identification[C]//International semantic web conference. Springer, Berlin, Heidelberg, 2013: 542-557.
- [3] Juanzi L I , Hou L . Reviews on Knowledge Graph Research[J]. Journal of Shanxi University(Natural Science Edition), 2017.
- [4] Ge B; Tan Z; Zhang Chong; Xiao Weidong, Military knowledge graph construction techniques, [J], Journal of Command and Control, 2016, 42-48, 42-48. (In Chinese)
- [5] Yufeng MA, Nan XIANG, Yajie DOU, Jiang JIANG, Kewei YANG, Yuejin TAN. Application and research of knowledge graph in military system engineering[J]. Systems Engineering and Electronics, 2022, 44(1): 146-153.
- [6] Xu Zenglin, Sheng Yongpan, He Lirong, et al. A review of knowledge graph techniques [J]. Journal of University of Electronic Science and Technology, 2016, 45(4) : 589 -606. (In Chinese)
- [7] Gu, Danyang; Li, Mingqian; Quan, Jichuan; Liu, Yong; Luo, Chen, Ontology-based knowledge mapping construction for main combat weapons and equipment, [J], Command Control and Simulation, 2021, 18-24, 18-24. (In Chinese)
- [8] Meng Rongrong, An analysis of the connotation of "ontology" and "ontology" in Chinese philosophy, [J], Journal of Heihe College, 2018, 30-31, 30-31. (In Chinese)
- [9] ECHES R, FIKES R E, FININ T, et al. Enabling technology for knowledge sharing[J]. AI Magazine, 1991, 12(3) : 36-56.
- [10] RUBER T R. A translation approach to portable ontology specifications [J]. Knowledge Acquisition, 1993, 5 (2) : 199-220. [11]
- Liu, Yang Li, Hong Duan, et al. A review of knowledge graph construction techniques [J]. Computer Research and Development,

- 2016, 53(3) : 582 -600. (In Chinese)
- [12] Yufeng MA, Nan XIANG, Yajie DOU, Jiang JIANG, Kewei YANG, Yuejin TAN. Application and research of knowledge graph in military system engineering[J]. Systems Engineering and Electronics, 2022, 44(1): 146-153.
- [13] Jinli Che; Liwei Tang; Shijie Deng; Xujun Su, Construction and application of military equipment knowledge graph based on encyclopedic knowledge, [J], Journal of Weaponry Equipment Engineering, 2019, 154-159, 154-159. (In Chinese)
- [14] Zhu Wenyue, Liu Zongtian. Event ontology-based domain knowledge modeling of emergent events [J]. Computer Engineering and Applications, 2018, 54(21) : 148-155. (In Chinese)
- [15] Li W. T. Research on the construction of ontology library based on biological domain [J]. Shandong Industrial Technology, 2018(15) : 229. (In Chinese)
- [16] Ren Fei-Liang, Shen J-Kun, Sun Bin-Bin, Zhu Jing-Bo. A review of techniques for building domain ontologies from text[J]. Journal of Computer Science, 3(2019)654-676. (In Chinese)
- [17] Wang Hongyu; Yang Chaohong; Zhou Yuwei, Research on the content and methods of ontology construction in military domain, [J], Information Communication, 2020, 153-155, 153-155. (In Chinese)
- [18] Yue Lixin, Liu Wenyun. A comparative study of domestic and foreign domain ontology construction methods [J]. Intelligence Theory and Practice, 2016, 39(8) : 119-125. (In Chinese)
- [19] Zhang YF, Guo Y, Li KW, et al. An ontology modeling approach for emergent event domain [J]. Information Systems Engineering, 2020(5) : 134-136. (In Chinese)
- [20] Ma XM, Wang HR. Ontology construction methods and

- applications [J]. Information and Computer (Theory Edition) , 2018(5) : 33-35. (In Chinese)
- [21] Dai T., Zhou L., Yu Q. Y., et al. A cyclic iterative ontology construction method based on requirements assessment response [J]. Computer Applications, 2020, 40(9) : 2712-2718. (In Chinese)
- [22] Liu, Mengchao; Wang, Yumei; Wu, Yafei; Zang, Yihua; Liang, Jia, Ontology-based knowledge modeling and analysis of military equipment, [J], Computers and Modernization, 2021, 80-84, 80-84. (In Chinese)
- [23] Xing Meng; Yang Chaohong; Bi Jianquan, Construction and application of knowledge graph in military domain, [J], Command Control and Simulation, 2020, 7-13, 7-13. (In Chinese)
- [24] Ou Yanpeng, A review of knowledge graph technology research, [J], Electronic World, 2018, 56+58, 56+58. (In Chinese)
- [25] Sil A, Kundu G, Florian R, et al. Neural Cross-Lingual En-tity Linking [C] // Processings of the 32nd AAAI Conference on Artificial Intelligence, 2018: 5464-5472.
- [26] Bordes A , Usunier N , Garcia-Duran A , et al. Translating Embeddings for Modeling Multi-relational Data. Curran Associates Inc. 2013.
- [27] Wang Z , Zhang J , Feng J , et al. Knowledge Graph Embedding by Translating on Hyperplanes. AAAI Press, 2014.
- [28] Shaozhi, Da i, Yanchun, et al. Learning Entity and Relation Embeddings with Entity Description for Knowledge Graph Completion[C]// 2018.
- [29] Jia Y, Wang Y, Jin X, et al. Knowledge graph embedding: A locally and temporally adaptive translation-based approach[J]. ACM Transactions on the Web (TWEB), 2017, 12(2): 1-33.
- [30] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[J]. Advances in neural

information processing systems, 2013, 26.

[31] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv:1301.3781, 2013.

[32] Lin T Y , Roychowdhury A , Maji S . Bilinear CNN Models for Fine-grained Visual Recognition[J]. 2015.

[33] Trouillon T, Welbl J, Riedel S, et al. Complex embeddings for simple link prediction[C]//International conference on machine learning. PMLR, 2016: 2071–2080.

[34] Dettmers T, Minervini P, Stenetorp P, et al. Convolutional 2d knowledge graph embeddings[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2018, 32(1).

[35] Wu, Yunchao; Mao, Shaojie; Zhou, Fang, Domain knowledge graph construction techniques for simulation-based extrapolation, [J], Command Information Systems and Technology, 2019, 36-40+86, 36-40+86. (In Chinese)

[36] Fetahu B, Gadiraju U, Dietze S. Improving Entity Retrieval on Structured Data [C] // International Semantic Web Conference. Springer International Publishing, 2015.

[37] Yan Fan, Chengyu Wang, Guomin Zhou, etc. An Architecture for Building a Domain Knowledge Graph from Scratch [C] // International Conference on Database Systems for Advanced Applications. Springer, Cham, 2017.

[38] Liu C , Yu Y , Li X , et al. Application of Entity Relation Extraction Method Under CRF and Syntax Analysis Tree in the Construction of Military Equipment Knowledge Graph[J]. IEEE Access.

[39] Zhao Yu; Chen Zhikun; Yang Chun, An open source data-based knowledge graph construction method for military domain, [J], Command Information Systems and Technology, 2019, 68-73, 68-73. (In Chinese)

- [40] Sil A, Kundu G, Florian R, et al. Neural Cross-Lingual Entity Linking [C] // Processings of the 32nd AAAI Conference on Artificial Intelligence, 2018: 5464-5472.
- [41] Wang Ting , Di Ruihua , Song Jicheng . A Novel Online Encyclopedia-oriented Approach for Large-scale Knowledge Base Construction [J]. Journal of Software, 2014, 9 (2) :482-489.
- [42] Xuezheng Y I N, Hui Z, Junbao Z, et al. Multi-neural network collaboration for Chinese military named entity recognition[J]. Journal of Tsinghua University (Science and Technology), 2020, 60(8): 648-655.
- [43] Wang Z, Chen T, Ren J, et al. Deep Reasoning with Knowledge Graph for Social Relationship Understanding [C]. International Joint Conference on Artificial Intelligence: 1807. 00504, 2018.
- [44] Li Baozhen, Su Jing. Construction of domain knowledge graph based on expert-generated content [J]. Intelligence Science, 2018, 36(10) : 13-19. (In Chinese)
- [45] Hou X, Zhu C, Li Y, et al. Question answering system based on military knowledge graph[C]//International Conference on Electronic Information Engineering and Computer Communication (EIECC 2021). SPIE, 2022, 12172: 33-39.
- [46] ZHENG W, ZOU L, LIAN X, et al. How to build templates for RDF question / answering: an uncertain graph similarity join approach [C] / / ACM SIGMOD International Conference on Management of Data. ACM, 2015: 1809 — 1824.
- [47] Lin Wangqun; Wang Miao; Wang Wei; Wang Chongnan; Jin Songchang, Current status of knowledge graph research and military applications, [J], Chinese Journal of Information, 2020, 13-20, 13-20. (In Chinese)
- [48] DARPA. Department of Defense Fiscal Year (FY) 2012-2016President’ s Budget Submission (Unclassified) [R].

DefenseWide Justification Book.

[49] Zuo Yi; Zhang Guilin; Wu Wei; Wang Jing; Li Xiaodong; Wang Huijuan, Knowledge graph application for battlefield sea and air target identification, [J], Command Information System and Technology, 2019, 5-9+28, 5-9+28. (In Chinese)

[50] Wang Limin. Construction of vulnerability knowledge graph and research on vulnerability situational awareness technology [D]. University of Chinese Academy of Sciences (School of Artificial Intelligence, University of Chinese Academy of Sciences), 2020. (In Chinese)

[51] Chen Sheng. Uncovering the "prism" [J].

East-West-North-South, 2013 (16): 12-12. (In Chinese)

[52] Nana. Arms company BAE Systems receives \$11.4 million for cyber attack program [EB/OL].

<https://www.aqniu.com/industry/20508.html>. (In Chinese)

外文文献译文

军事领域知识图谱研究综述

郭旭东、陈林秀、袁莉、王庆林、关伟丽

摘要随着信息技术的发展和军事信息化水平的不断提高，复杂多变的战场形势对军队海量数据处理能力提出了巨大挑战。本文提出了构建军事领域知识图谱的相关技术框架，总结了知识图谱在军事领域应用的不同方面，并针对知识图谱在军队领域应用的现状，揭示了改进方向。

关键词：军事领域，知识图谱，本体

1.简介

在现代智能战争的背景下，战场信息技术和传感技术的广泛应用导致了战场空间的急剧扩展和互联，数据爆炸。情报、侦察、监视和海量数据具有“4V”特征：规模大、变化快、多样性和价值。庞大的数据量使得战场态势评估和分析的需求远远超过了传统的基于人工经验和认知的计算机辅助评估和决策处理能力[1]。将强大的知识描述和管理方法引入军事应用是可取的。

知识图谱起源于语义网络的概念，由谷歌于2012年首次提出[2]。知识图谱集成了数据库、信息检索、数据挖掘、自然语言处理和图计算分析等技术方法[3]。知识图谱可以集成主题领域的知识，并提供视觉呈现等服务。知识图谱的应用显著提高了信息检索能力，展示了知识图谱在信息集成中的强大功能。

作为军事领域的科技前沿，专有知识图谱的构建可以在很大程度上提高作战效率，主要体现在以下几个方面[4]：知识图谱可以集成大量分散和孤立的情报；知识图谱可以让计算机理解智能的语义；知识图谱可以在语义层面上关联大量的智能；知识图谱可以挖掘相关智能中的新知识。因此，有必要将知识图谱应用于军事领域。

知识图谱由两个层次组成：模式层和数据层[5]。模式层用于规范数据层，并制定数据层所遵循的数据类型和层次结构的规则，它是知识图谱的核心。模式层通常使用本体库来管理知识图谱的模式层。数据层存储特定于知识图谱的知识，并且可以被视为模式层的实例化。

本文介绍了军事领域知识本体的构建，并从三个方面对军事领域知识图谱的知识服务进行了总结和展望。

2.军事领域知识本体的构建

知识图谱有两种构建方式，自上而下和自下而上[6]。自上而下的方法通常首先为知识图谱定义一个数据模式，该模式从最顶层的概念开始构建，并逐步向下细化，形成一个结构良好的分类层次，然后将实体添加到概念中。自下而上的方法归纳组织实体，形成底层概念，然后逐渐向上抽象，形成顶层概念。

军事领域是一个典型的特定领域，具有明确的概念层次，然而，部分专业知识存在于领域专家的脑海中，可能无法从数据中归纳和抽象地获得。由于军事数据的保密性和军事装备的特殊性，许多文件和信息对知识图谱的建立具有绝对的阻碍作用，导致数据结构较少，数据收集困难。因此，该领域的知识图谱在构建之初更适合采用自上而下的方法。在知识图谱的基本结构已经建立，其类别节点或关系的数量已经达到一定规模后，可以使用自下而上的方法来扩展图的实例和属性数据[6]。通过以上分析，鉴于军事装备知识图谱的特殊性，首先需要构建本体。

2.1. 本体论

本体论最初是一个哲学定义[7]，是对客观事物的本质描述。随着人工智能、知识图谱等领域的发展，本体的概念逐渐得到重视，并在这些领域得到广泛应用。1991年，Neches提出了本体论的定义，“构成所讨论领域词汇的基本术语和关系，以及使用这些术语和关系定义这些术语的扩展的规则”[8]。1993年，Gruber提出“本体是一种共享的、概念化的、形式化的规范”[10]，该定义被学术界广泛接受。

在知识图谱领域,本体是通过抽象和约束领域中不同实体概念、属性及其相互关系而获得的规范,是知识图谱最核心的部分,主要被描述为树结构[11]。本体是结构化知识库的概念模板,通过本体库形成的知识库不仅层次结构更强,而且冗余度更低[12]。构建本体的主要目的是为知识图谱定义一个清晰的数据模式,而正确的数据模式可以确保知识图谱中的实体具有强大的系统关系,这是构建知识图谱的基础[13]。

2.2. 本体构建原则

本体构建是构建知识模型的主要方式,遵循以下五个公认原则[13][14],即清晰与客观、完整性、连贯性、最大单调可扩展性和最小本体承诺原则。

在清晰度和客观性方面,本体论必须能够以客观且独立于上下文的定义有效地传达所定义术语的含义。在完备性上,本体中给出的术语定义是完整的,并充分表达了所描述术语的含义。

本体论定义的公理,以及用自然语言描述它们的文档,应该是一致的。

在连贯性方面,从术语中得出的推论与术语本身的含义是一致的,即它们支持与其定义一致的推理,不会产生矛盾;自然语言中定义的公理和说明的文档也应该是一致的。

在最大单调可扩展性上,向本体添加通用或专门术语不需要修改其现有概念定义和内容,并且支持基于现有概念定义新术语。

关于最小本体承诺,本体承诺应该是最小的,并在尽可能少的约束下处理建模对象。本体承诺意味着就如何以一致和兼容的方式使用共享词汇达成共识。一般来说,本体约定充分满足特定的知识共享需求,这可以通过定义约束最小的公理和仅定义通信所需的词汇来确保。

2.3. 本体构建方法

本体构建主要有三种方法,手动构建、重用现有本体(半自动构建和自动构建[15])。人工构建方法主要通过领域专家确定知识内容和关系,使本体质量和准确性较高。军事领域对知识的完整性和准确性要求很高,因此人工构建方法适用于军事知识本体的构建。采用通用本体手工构建的思想,军事本体可以应用自上而下的手工构建,首先构建军事知识的顶层本体,在此基础上开始构建每个子域的知识框架,逐步完成整个军事知识本体的构建[16]。

国内外比较有影响力的本体手册构建方法有五种[17][18]:骨架法、TOVE法、IDEF5法、本体方法和七步法。与其他方法相比,七步法比较成熟。它包括七个步骤,即确定专业知识的领域和范围,检查重用现有本体的可能性,列出领域中的重要术语,定义类和类层次结构级别,定义类属性,定义属性方面,以及创建实例[19]。然而,七步走的方法也有一些局限性;这种方法没有一个完整的生命周期[20],并且缺乏在开发后期对需求调整或功能添加做出响应的模型修订过程。

2.4. 军事领域本体结构

在军事装备领域,文献[13]将军事装备分为火炮、舰艇和飞行器等八类,并为装备实体定义了杀伤力和最大速度等属性。这种装备知识体系偏向于对军事装备的科学描述,忽视了面向军事作战需求的知识表达。刘等人[21]设计了一种基于需求的螺旋反馈方法来构建军事装备知识模型,定义了一个顶级核心概念——军事装备、四个相关概念(如作战活动)和七个通用概念(如国家),其中装备分为八大类和148个子类。该模型包括202个实体属性、6个实体关系,涵盖5800个武器类型实体、18个作战活动和86个国家。

从以上分类可以看出,军事领域不同任务的本体构建重点各不相同,但基本可以分为以下几个方面。

3. 军事领域知识图谱的构建

基于知识本体,可以构造知识图谱。军事领域知识图谱的构建和应用是一项系统工程,其构建过程分为六个阶段,可称为领域知识图谱全生命周期,其设计的关键技术过程主要包括知识表示、知识存储、知识提取、知识融合、知识进化和知识应用[22]。

3.1. 知识表示

知识表示研究如何实现对现实世界中的事物和相互关联的事物的建模,提供符合人类表达的数据逻辑信息,并实现人与计算机之间的无障碍通信[23]。知识图谱中的知识表示定义了领域的基本认知框架,阐明了基本概念和概念之间的语义关联,并提供了基本的机器认知框架和数据结构,以实现知识的合理组织[24]。知识图谱主要基于语义网络的表示,知识表示以三元集的形式进行。

在军事领域,一些应用场景添加了时空维度等表达元素,这些元素远远超出了知识图谱三元组所能表达的简单关联事实。尽管细粒度的知识表示是军事领域应用的强大需求之一,但这并不意味着所有应用场景都需要细粒度的表示,而且粒度越细,表达能力越强,获取成本就越高。

3.2. 知识存储

知识图谱的构建依赖于用于存储知识图谱的框架。在一个完美的知识图谱存储框架上构建、表示和应用知识图谱,将大大提高知识图谱的使用效率[4]。

在军事领域,知识图谱所涉及的知识具有多源性、异构性、时间序列性、实时性、冗余性、强相关性、闭环性等特点。同时,复杂场景的应用需要满足高效查询和高并发的要求[22]。

为了解决这些问题,Wu 等人[35]使用开源的图数据库系统 Neo4j 图数据库来存储面向模拟推导的领域知识图谱,并使用方便的图查询语言 Cypher 进行数据查询,以解决为基于文本的内容构建知识图谱的问题。Xing 等人[22]提出了一种采用多种存储介质实现军事知识混合存储的框架,该框架包括六个部分:源数据存储管理器、数据处理器、数据提取引擎、消息队列、存储和索引接口以及订阅服务,包括关系数据库、图数据库、内存数据库、文件索引系统等。非结构化和半结构化数据以 HDF 格式和 RDF 格式分布在知识图谱中,分布式图存储和注释,通过内存数据库和关系数据库有效融合多源异构数据,对多源异构知识进行统一存储管理,使用文件存储系统存储视频、图片、声音等文档。Ge 等人[4]提出了一种基于多约束的分块管理框架,以优化影响查询效率的多个因素。该框架通过优化最小切边数来提高查询效率,并通过优化通信总量来达到负载均衡的目的。

3.3. 知识提取

知识提取是为了填充知识框架的知识实例,并使用相关技术从异构和复杂的数据中识别、发现和提取概念、类型、事实及其相关关系、约束规则等。知识提取可以分为实体提取、关系提取和属性提取。

在军事领域,各级各类大型军事信息系统在多种任务的复杂使用环境下生成大量军事数据,这些数据具有数据来源不同、数据格式多样的特点,这就需要针对不同数据的特点,采用关键手段来提高知识提取的准确性和召回率。例如,由于军事领域应用场景的极端严重性,知识获取的质量要求非常高,知识获取方案必须平衡需求和成本,以避免知识库中信息丢失、错误和过时的问题[35][37]。为了解决这些问题,Xing 等人[22]提出了在保证质量的同时最大限度地降低成本的原则:首先,指定数据源,并尽量按照简单中避难的原则选择质量更好、结构化程度更高的数据源,如《环球报》等。之后,通过非结构化数据逐步扩展规则集,进行跨领域知识迁移和重用相似领域知识。

对于非结构化数据提取方法,通常使用有监督的机器学习方法。刘[38]等人使用开源 Java 工具箱中的中文分类系统进行实体提取,并使用结合 CRF 和句法分析树的最大熵模型进行关系提取。赵等人[39]使用双向 LSTM-CRF 模型进行实体提取、句法分析和分层 Dirichlet 过程聚类来实现关系词提取,并使用双向 RNN 模型来实现实体关系提取算法。

3.4. 知识融合

由于数据来源的多样性和提取方法的不同,通过知识提取获得的知识可能存在大量的冗余、噪声、不完整性和不确定性。同时,数据之间没有层次结构和明确的逻辑关系,这些问题无法通过知识提取的清理会话来解决,迫切需要一个基于集成的实体关系表示模型[22]。因此,需要完成海量异

构数据资源的语义标准和链接,以及不同领域、数据模式和应用需求的实体消歧和实体对齐,实现多领域映射的跨领域语义融合,形成全球统一的知识库,即知识融合[39][41]。军事领域知识图谱的模式层采用人工辅助构建,精度高;要处理的主要问题是数据层的融合。[42]

对于数据级知识融合问题,可以使用实体消歧、实体对齐、属性融合和值归一化等技术。实体消歧通过测量候选实体和上下文之间的相关性来推断当前单词所指的特定实体。实体对齐是通过将多个图中的多个特征数据表示为同一实体来消除它们的歧义。吴等[35]利用余弦相似度对实体进行聚类,从而实现了实体消歧和对齐的功能。使用基于质量估计的方法,将最有价值的知识源作为冲突实体的有效知识,从而实现属性融合等功能。

3.5. 知识计算与进化

为了提高知识图谱在军事领域的实际应用价值,有必要挖掘实体之间存在的大量隐含关系。

在军事领域,知识图谱的实体分布更为密集,单个实体的相关知识覆盖范围更广,因此有必要通过快速加载增量知识和规则来完成远程推理,以满足其知识表示粒度和深度更深的特点,从而保证推理的可扩展性和正确性,并在语义层面挖掘更多相关信息和知识[43]。同时,军事领域知识图谱还存在小样本问题,需要解决具有数据稀疏性特征的知识图谱的推理问题。此外,军事知识图谱的构建是一个不断迭代更新的过程,作战方案、战略和手段也在不断变化,这就需要及时对知识进行信心评估,并改进新知识。

为了解决知识推理等问题,Xing 等人[22]提出了一种基于协同过滤机制的远程监督与深度学习相结合的知识推理方法,通过远程监督,通过添加外部知识,介入知识推理过程构建数据集,并构建和训练双向 LSTM 神经网络,通过深度学习方法提取新的三元组,以提高知识推理的性能,同时解决小样本问题,达到知识图谱互补的效果。吴等[35]采用了结合数值计算的概率逻辑推理方法,首先得到候选答案,然后进行概率逻辑推理,从而实现了实体关系互补的功能。Ge 等人[4]提出了一种基于最短路径的精确查询框架和一种基于相似度计算的模糊查询框架。

针对知识图谱的更新和维护问题,赵等人[39]给出了新数据的更新算法过程框架,通过模板匹配等算法确定是否进入数据库,通过置信度计算判断是否进行知识更新。Ge 等人[4]给出了一个半自动化的知识图谱更新过程。首先,分布式爬虫用于收集高置信度的数据源,在这些数据源上提取实体,而军事情报领域涉及的关系由于相对固定和清晰,可以手动操作。新生成的三元组通过多个数据源进行验证,最后更新和维护军事知识图谱中现有的三元组。该过程中的关系确定可以通过使用机器学习方法加上手动复习来实现,这可以提高知识库更新的效率。

3.6. 知识应用

在知识应用方面,应考虑技术成熟度,并仔细选择应用场景。知识图谱作为一种面向人机协作的开放式知识管理机制,具有特定的价值和成本,并不总是解决不同应用问题的最佳方案[44]。军事领域知识图谱的应用场景比通用知识图谱复杂得多,而当前的领域知识系统大多是在构建时考虑到领域数据的基本概念和实体,这在理解复杂应用场景中的用户需求方面存在天然缺陷。如何更好地设计复杂的作战应用场景,并弥合这一语义鸿沟,是知识应用会话中需要关注和解决的问题。本节将在下一节中具体阐述。

4. 面向军事领域知识图谱的知识服务

4.1. 信息查询

基于知识图谱,可以提供数据可视化、智能问答等基础信息服务。赵等[39]利用 Javaneese、Tomcat 图形显示和 Bootstrap 技术,设计了主题的知识图谱检索与显示服务→事件→以区域划分和关注主题为聚合点,以分层的方式对实体进行分类,提高了子领域相关热点的检索、分析和整理效率。在构建的军事装备知识图谱的基础上,Dou 等人[44]实现了基于模板匹配的知识问答[46]。

4.2. 情报侦察与采矿

基于知识图谱支持快速检索、高效存储和长链推理的特点,知识图谱可以完成情报侦察挖掘中的情

报收集和整理等工作[47]。

Palantir 是第一家在国防部门使用知识图谱技术为反恐和作战行动提供情报分析服务的公司，在政府和军事决策和指挥中发挥了重要作用。国防高级研究计划局（DARPA）[48]于 2012 年 3 月启动了 XDATA 计划，以开发用于分析大量半结构化和非结构化数据的计算技术和软件工具，同年启动的文本深度探索和过滤（DEFT）计划更明确地建议使用深度学习技术来揭示大量结构化文本的隐含、实用特征，以及进一步整合处理后的信息的能力，并在此基础上，将这些技术用于作战评估、规划和预测，以支持决策。DARPA 于 2017 年发布的“AIDA”项目研究了从多个媒体源自动获取的知识到通用语义表示、已知本体扩展技术和其他技术的映射，以转换碎片信息并实现深度知识挖掘。

4.3. 战场态势感知

基于知识图谱强大的信息集成和知识挖掘能力，在战场态势感知中，知识图谱可以通过提高知识融合算法的性能和提高知识处理工具的操作效率，帮助军事决策者更好地理解战场态势感知。

Yi 等人[49]提出了一种智能识别推理框架，该框架首先利用知识图谱和图数据库技术构建知识图谱数据库，然后构建具有领域知识推理能力的多元智能，并利用分类器推理技术进行集成识别推理，可以实现智能目标识别的功能。

4.4. 作战指挥与控制

基于知识图谱强大的知识表达能力以及知识提取的准确性和快速性，在作战指挥控制方面，知识图谱有助于高效指挥和快速决策。

2012 年，美国陆军的数据到决策计划使用知识图谱等关键技术，有效地从数据中提取知识，获得作战人员和决策者指导作战所需的信息，并通过增强的数据融合将其与相关的上下文和态势信息相集成，以提供威胁、选项和后果的清晰画面[47]。仿真推送是根据作战场景中指定的作战意图、顺序和过程，在作战执行之前或执行过程中，对作战场景的部署、作战目标或不同阶段的结果进行预演和分析的过程。知识图谱还为仿真预演提供知识和数据支持，并为海量异构数据提供结构化存储，从而提高仿真预演系统的效率[35]。

4.5. 网络空间安全

基于知识图谱的数据检索、数据管理和知识推理能力，知识图谱在网络空间安全背景下，对增强网络态势感知、提高网络攻防能力、维护国家网络空间安全具有重要意义[49]。

美国国家安全局的棱镜计划每天从电信公司提取数百万用户的通话记录，并从微软、谷歌、雅虎、脸书和苹果等九家网络巨头的中央服务器获取信息，通过信息融合和知识获取技术进行相关分析和推理，生成高质量的情报产品。信息融合和知识获取技术可用于关联、分析和推理，并产生对其他国家网络空间安全构成巨大威胁的高质量情报产品[50]。美国陆军的“网络攻击自动化非常规传感器环境”计划旨在开发网络攻击行为的预测方法和检测即将发生的网络现象的有效方法，以帮助网络防御者应对网络攻击，开发和验证能够预测网络攻击的非常规、多学科传感技术，并执行现有的先进入侵检测能力[51]。

5、总结与展望

本文总结了军事领域知识图谱构建的相关技术框架，并展示了知识图谱在军事领域应用的不同方面。可以看出，知识图谱通过使用大规模的语义网络来描述客观世界中的概念、实体、属性及其之间的关系，为提高军事领域的情报提供了一种强大的方法。

为了充分利用知识图谱技术在军事领域资源整合、智能计算等领域的优势，可以加强以下方面的研究。

首先，要加强知识图谱的基础技术研究。应充分利用人工智能等相关技术，提高知识图谱的表示能力、算法效率和智能化水平，加强对知识图谱关键技术的研究。

第二，提高军队数字化水平。指挥信息系统基本上仍处于辅助“劳动”而非辅助“智能”阶段，无法很好地解决指挥官急需的态势理解、决策支持或对抗预演等需要高智能的问题[11]。将知识图谱快

速应用于战场态势感知和作战指挥控制，可以大大提高军队的作战水平和作战效率。

第三，应扩大知识图谱在军事领域的应用。知识图谱技术在军事领域的应用仍然相对有限。充分发挥知识图谱在军事领域的应用，将产生巨大的军事效益。

最后，提高了基于知识图谱的应用能力。随着当前国际形势的日益严峻和国际关系的日益复杂，基于单一的知识图谱很难推测作战动机，也很难对战场态势进行清晰的分析。因此，我们需要建立复杂的态势分析模型和方法，整合来自多个领域的结构化知识，结合多种信息，这是更好地发挥知识图谱在战场领域作用的重要途径。因此，它对基于知识图谱的应用系统提出了更高的要求，如全面的信息收集能力、高效的信息融合能力和强大的信息推理能力。

知识图谱将继续是未来的热点问题。相信在不久的将来，通过大数据技术和人工智能技术的结合，知识图谱可以为军事水平的提高添砖加瓦。

参考文献：

- [1] 朱峰; 胡晓峰, 深度学习战场综述态势评估与研究展望[J]. 军事行动研究与系统工程, 2016, 24-29, 24-29. (中文)
- [2] Pujara J, Miao H, Getoor L 等. 知识图谱概述 identification[C]国际语义网会议. 施普林格, 柏林, 海德堡, 2013:542-557.
- [3] Juanzi LI, Hou L. 知识图谱研究综述[J]. 山西大学学报 (自然科学版), 2017.
- [4] Ge B; 谭 Z; 张冲; 肖卫东, 军事知识图构造技术综述[J]. 指挥与控制学报, 2016, 42-48, 42-48. (中文)
- [5] 马玉凤、项楠、窦亚杰、姜、杨克伟、, 跃进潭. 知识图谱在中的研究军事系统工程领域的应用[J]. 系统工程与电子, 2022, 44(1): 146-153.
- [6] 徐增林, 盛永攀, 何丽蓉, 等, 知识图谱技术综述[J]. 电子大学学报科学与技术, 2016, 45 (4): 589-606. (中文)
- [7] 顾, 丹阳; 李明倩; 权继川; 刘、勇; 罗, 陈, 基于本体的知识图谱构建针对主战武器装备, [J], 司令部控制与仿真, 2021, 18-24, 18-24. (中文)
- [8] 孟荣荣: “本体论”的内涵分析中国哲学中的“本体论”[J]. 黑河学院学报, 2018, 30-31, 30-31. (中文)
- [9] ECHES R、FIKES R E、FININ T 等. 使能技术 foR 知识动摇[J]. AI 杂志, 1991,12 (3): 36-56.
- [10] RUBER T.R.可移植本体的翻译应用 Roach 规范 (J) .知识获取, 1993,5 (2): 199-220.[11]刘, 杨丽, 洪端, 等.知识图谱综述施工技术[J]. 计算机研究与开发, 2016, 53(3): 582 -600. (中文)
- [12] 马玉凤、项楠、窦亚杰、姜、科伟杨, 谭跃进. 的应用和研究军事系统工程中的知识图谱[J]. 系统工程与电子, 2022, 44 (1): 146-153.
- [13] 金丽车; 唐立伟; 邓世杰; 苏旭军, 建筑军事装备知识图谱及其应用基于百科全书式的知识, [J]武器装备工程, 2019, 154-159, 154-159. (中文)
- [14] 朱文跃, 刘宗天. 基于事件本体的领域突发事件的知识建模[J]. 计算机工程和应用, 2018, 54 (21): 148-155. (中文)
- [15] 基于的本体库构建研究生物结构域[J]. 山东工业技术, 2018 (15): 229. (中文)
- [16] 任飞亮, 沈建坤, 孙斌, 朱景波从文本构建领域本体的技术[J]. 的期刊计算机科学, 3 (2019) 654-676. (中文)

- [17] 王洪宇; 杨朝红; 周玉伟军事领域本体构建的内容与方法[J], 信息通信, 2020, 153-155, 153-155。
(中文)
- [18] 岳立新、刘文云。国内与国外的比较研究国外领域本体构建方法[J]。智力理论与实践, 2016, 39 (8): 119-125。(中文)
- [19] 张, 郭勇, 李 KW, 等。本体建模突发事件域的方法[J]。信息系统工程, 2020 (5): 134-136。
(中文)
- [20] 马 XM, 王 HR.本体构建方法与方法应用[J]。信息与计算机(理论版本), 2018 (5): 33-35。(中文)
- [21]戴, 周, 于 Q.Y., 等.循环迭代本体论基于需求评估响应的施工方法[J]。计算机应用, 2020, 40 (9): 2712-2718。(中文)
- [22]刘, 孟超; 王, 玉梅; 吴, 亚飞; 臧义华; 梁、贾、, 基于本体的军事知识建模与分析设备, [J], 计算机与现代化, 2021, 80-84, 80-84。(中文)
- [23]邢萌; 杨朝红; 毕建全, 建筑与知识图谱在军事领域中的应用[J].司令部控制与仿真, 2020, 7-13, 7-13。(中文)
- [24]欧艳鹏, 知识图谱技术研究综述, [J], 电子世界, 2018, 56+58, 56+58。(中文)
- [25]Sil A, Kundu G, Florian R, 等神经跨语言研究链接 (C) 第 32 届 AAAI 会议进程人工智能, 2018:55464-5472。
- [26]Bordes A, Usunier N, Garcia Duran A, et al.翻译用于建模多关系数据的嵌入。Curran Associates 股份有限公司 2013 年。
- [27]王, 张, 冯, 等.知识图谱的嵌入在超平面上进行翻译。AAAI 出版社, 2014 年。
- [28]邵志, 大一, 延春, 等.学习实体与关系知识图谱的实体描述嵌入完成[C].2018。
- [29]贾, 王, 金, 等.知识图谱嵌入: 局部以及基于时间自适应翻译的方法[J]。ACM 网络交易 (TWEB), 2017, 12 (2): 1-33。
- [30]Mikolov T, Sutskever I, Chen K 等。分布式表示词和短语及其合成性[J]。神经研究进展信息处理系统, 2013, 26。
- [31]Mikolov T, Chen K, Corrado G 等.词的有效估计向量空间中的表示[J]。arXiv 预印本 arXiv:1301.3781, 2013。
- [32]林, 罗伊乔杜里。的双线性 CNN 模型细粒度视觉识别[J]。2015。
- [33]Trouillon T, Welbl J, Riedel S 等。简单的复杂嵌入链接预测[C]机器学习国际会议。PMLR, 2016 年: 2071-2080。
- [34]Dettmers T, Minervini P, Stenetorp P 等, 卷积 2d 知识图谱嵌入[C]AAAI 论文集人工智能会议。2018, 32(1)。
- [35]吴, 云超; 毛、少杰; 周, 方, 领域知识图谱基于仿真的外推构造技术[J], 指挥信息系统与技术, 2019, 36-40+86, 36-40+86。(中文)
- [36]Fetahu B, Gadiraju U, Dietze S.改进实体检索结构化数据 (C) 国际语义网会议。施普林格国际出版社, 2015 年。
- [37]严帆, 王承宇, 周国民等.中国建筑从头开始构建领域知识图谱 (C) I高级数据库系统国际会议应用。施普林格, 商会, 2017。
- [38]刘 C, 余 Y, 李 X, 等.实体关系提取的应用 CRF 下的方法和语法分析树在的构建军事装备知识图谱[J]。IEEE 接入。
- [39]赵于; 陈志坤; 杨春, 开源的基于数据的知识图谱构建方法军事领域, [J], 指挥信息系统和技术, 2019, 68-73, 68-73。(中文)
- [40]Sil A, Kundu G, Florian R, 等神经跨语言实体链接 (C) I第 32 届 AAAI 会议进程人工智能, 2018:55464-5472。

- [41]王婷, 地 R 宋继成. 网络小说面向百科全书的大规模知识库方法建筑〔J〕. 软件学报, 2014, 9 (2): 482-489.
- [42]Xuezhen Y I N, Hui Z, Junbao Z, 等. 多神经网络中国军方命名实体识别合作[J]. 清华大学学报 (科技版), 2020, 60(8): 648-655.
- [43]王 Z, 陈 T Ren J, et al. Deep R 用知识思考社交关系图 R 关系理解〔C〕. 国际人工智能联合会议: 1807 年. 00504, 2018.
- [44]李宝珍, 苏静. 基于领域知识图谱的构建关于专家生成的内容[J]. 智能科学, 2018, 36 (10): 13-19. (中文)
- [45]侯 X, 朱 C, 李 Y, 等. 基于军事知识图谱电子信息工程与计算机通信 (EIECC 2021). 斯皮, 2022, 12172: 33-39.
- [46]郑 W, 邹 L, LIAN X, 等. 如何为 RDF 问答: 一个不确定的图相似连接方法〔C〕ACM SIGMOD 国际会议数据管理. ACM, 2015:1809-1824.
- [47]林王群; 王淼; 王伟; 王崇南; 金宋昌, 知识图谱研究现状与军事应用, [J]. 中国信息学报, 2020, 13-20, 13-20. (中文)
- [48]DARPA. 国防部财政年度 2012-2016 年居民预算提交 (未分类) [R]. 防守范围内的辩护书.
- [49]左一; 张桂林; 吴伟; 王晶; 李晓东; 王慧娟, 知识图谱在海空战场中的应用目标识别, [J], 指挥信息系统和技术, 2019, 5-9+28, 5-9+28. (中文)
- [50]王利民. 漏洞知识图谱的构建脆弱性态势感知技术研究[D]. 中国科学院大学 (人工学院) 中国科学院大学情报), 2020 年. 在里面中国人
- [51]陈生. 揭开“棱镜”的面纱[J]. 东西-南北, 2013 (16): 12-12. (中文)
- [52]娜娜. 军火公司 BAE 系统公司收到 1140 万美元用于网络攻击程序[EB/OL]. <https://www.aqniu.com/industry/20508.html>. (中文)

任务书

毕业设计任务书

中国姓氏文化可视化系统的设计与实现

一、毕业设计目的

该设计旨在以直观、生动的方式展示中国姓氏文化的丰富内涵和独特魅力。通过这一系统，用户可以深入了解自己的姓氏起源、发展历程和家族故事，增强文化认同感和归属感。同时，该系统也可作为研究、教育工具，促进中外文化交流与传播，为中华文化的传承与创新注入新的活力。

二、主要内容

1. 数据收集与处理模块：负责从各种来源收集姓氏数据和相关信息，并进行必要的清洗、整理和分类。
2. 数据存储与管理模块：用于存储和管理姓氏数据和信息，提供高效的数据检索和访问功能。
3. 数据可视化模块：利用各种可视化工具和技术，将姓氏数据以直观、生动的方式呈现给用户，如动态图表、交互式地图、时间线等。
4. 查询模块：根据用户输入的姓氏，提供相关的查询结果和推荐信息，满足用户的个性化需求。
5. 系统管理模块：负责对整个系统进行管理和维护，包括数据更新与维护等。

三、重点解决的问题

- 1、任务模块的整体划分
- 2、前端展示界面的设计

3、数据的完整性

四、主要技术指标或主要参数

系统响应时间：一个好的预约系统应该能够快速响应用户的操作，提供流畅的用户体验。

并发用户数：保证一定数量用户同时访问时系统的稳定性和可用性。

可视化效果：系统采用的数据可视化技术需要提供清晰、直观、生动的可视化效果，使用户能够快速获取信息。

五、基本要求

完成软件设计、代码编写与模块测试，形成包含上述功能模块的软件原型产品；

完成 1 篇英文文献翻译；

完成不少于 15000 字的毕业论文一篇

六、其它（包括选题来源）

实践

指导教师： 2024 年 1 月 10 日

开题报告

华北水利水电大学本科毕业生毕业设计（论文）开题报告

学生姓名	苏玉恒	学号	202018018	专业	软件工程
题目名称	中国姓氏文化可视化系统的设计与实现				
题目来源	自选				
研究或设计 概述 (500 字左右)	<p>该系统是中国姓氏文化可视化系统的设计与实现，该设计旨在以直观、生动的方式展示中国姓氏文化的丰富内涵和独特魅力。通过这一系统，用户可以深入了解自己的姓氏起源、发展历程和家族故事，增强文化认同感和归属感。同时，该系统也可作为研究、教育工具，促进中外文化交流与传播，为中华文化的传承与创新注入新的活力。</p> <p>北宋刘恕所撰的《通鉴外纪》里面说：姓者，统其祖考之所自出；氏者，别其子孙之所自分；“姓氏者，标示家族血缘之符号也”。根据中国的历史研究，据研究，姓氏起源于图腾崇拜，它是氏族徽号或标志，大量的古代文献佐证了这一点。</p> <p>姓氏一直是代表中国传统的宗族观念的主要的外在表现形式，以一种血缘文化的特殊形式记录了中华民族的形成，在中华民族文化的同化和国家统一上曾起过独特的民族凝聚力的作用。</p> <p>中国人遵循“礼乐教化”的道德传统,重视祖先的教导,强调宗族的延续,血脉的传承。甚至可以通过姓氏将身在异乡、远离故土的炎黄子孙凝聚在一起,激励每个中华儿女,作为民族的一分子,血脉延续的一个环节,为祖国效忠,为民族争光。</p> <p>在该系统中，该系统将着重讲述不同姓氏的姓氏起源、发展历程和家族故事，力图该系统成为一个对中国姓氏的研究工具和教育工具。</p>				
主要内容	<p>中国姓氏文化可视化系统主要有 6 个模块：</p> <ol style="list-style-type: none"> 1. 数据收集与处理模块：负责从各种来源收集姓氏数据和相关信息，并进行必要的清洗、整理和分类。 2. 数据存储与管理模块：用于存储和管理姓氏数据和信息，提供高效的数据检索和访问功能。 3. 数据可视化模块：利用各种可视化工具和技术，将姓氏数据以直观、生动的方式呈现给用户，如动态图表、交互式地图、时间线等。 4. 查询与推荐模块：根据用户输入的姓氏，提供相关的查询结果和推荐信息，满足用户的个性化需求。 5. 社区互动模块：为用户提供一个交流和分享的平台，可以发布自己的姓氏故事、家族历史等，与其他用户进行互动。 6. 系统管理模块：负责对整个系统进行管理和维护，包括用户管理、权限管理、数据更新与维护等。 				

<p>主要参考文献（不少于 10 篇）</p>	<p>[1]肖锐.论中国姓氏文化研究意义[J].中南民族大学学报(人文社会科学版),2015,35(04):63-66.</p> <p>[2]田亚岐,倪景杰.我国姓氏的起源与发展[J].西安教育学院学报,2001,(01):43-47.</p> <p>[3]黄修明.中国古代姓氏文化变革中的社会政治因素[J].西南民族学院学报(哲学社会科学版),2000,(06):127-133.</p> <p>[4]周淑芸.我国历史上的姓氏变化与民族关系[J].太原师范学院学报(社会科学版),2004,(04):91-92.</p> <p>[5]冯丽娟,刘建荣.中华姓氏文化的内涵与当代价值研究[J].桂林师范高等专科学校学报,2019,33(04):61-65.DOI:10.16020/j.cnki.cn45-1302/z.2019.04.014.</p> <p>[6]邵妍.《百家姓》文化研究及其教学策略[D].山西大学,2013.</p> <p>[7]潘守蛟.姓氏的社会文化内涵[J].山东省农业管理干部学院学报,2003,(04):117-118.</p> <p>[8]袁义达.中国姓氏对传统文化和血缘寻根的影响[J].福建省社会主义学院学报,2012,(04):4-6.</p> <p>[9]刘青.姓氏中的文化现象解读[J].福建论坛(社科教育版),2009,(06):71-72.</p> <p>[10]李志前.中华姓氏的起源与形成[J].法制与社会,2012,(06):179-180.DOI:10.19387/j.cnki.1009-0592.2012.06.087.</p> <p>[11]钱文忠.《百家姓》蕴含的文化与文明[J].天津政协,2013,(04):57-58.</p> <p>[12]王泉根.姓氏谱系与中国文化之根[J].寻根,1994,(02):8-12.</p> <p>[13]张弛.漫谈中国历史上的改姓[J].丝绸之路,2004,(S2):19-20.</p> <p>[14]李振翼.中国姓氏文化的郡望之源[J].天水师范学院学报,2006,(04):24-26.</p> <p>[15]卢美松.姓氏文化与血缘认同[J].福建省社会主义学院学报,2014,(04):64-65.</p> <p>[16]王泉根.论姓氏谱系与中国文化[J].传统文化与现代化,1996,(04):45-54.</p> <p>[17]朱国林.漫谈姓氏文化的当代价值[J].寻根,2015,(06):16-18.</p> <p>[18]海各.中国的姓氏文化[J].承德大学,1996,(03):36.</p> <p>[19]叶姣.中日姓氏文化内涵比较[J].才智,2014,(35):329.</p> <p>[20]陈学芹.中国姓氏来源的故事图形化设计与研究[D].深圳大学,2019.DOI:10.27321/d.cnki.gszdu.2019.000244.</p> <p>[21]汪惠萍.中华姓氏新知[J].出版广角,2005,(05):53.</p>
<p>采取的主要技术路线或方法</p>	<p>中国姓氏文化可视化系统是一个基于知识图谱的系统，主要编程语言是 python，同时使用语言技术平台（LTP）。</p>
<p>时间安排</p>	<p>第 1 周：提交开题报告。</p> <p>第 2 周：完成中国姓氏文化可视化系统第 1 个模块的代码编写。</p> <p>第 3 周：完成中国姓氏文化可视化系统第 2 个模块的代码编写。</p> <p>第 4 周：完成中国姓氏文化可视化系统第 3 个模块的代码编写。</p> <p>第 5 周：提交论文中期检查。</p> <p>第 6 周：完成中国姓氏文化可视化系统第 4 个模块的代码编写。</p> <p>第 7 周：完成中国姓氏文化可视化系统第 5 个模块的代码编写。</p> <p>第 8 周：完成中国姓氏文化可视化系统第 6 个模块的代码编写。</p> <p>第 9 周：运行中国姓氏文化可视化系统，检查并清除系统运行中的漏洞。</p> <p>第 10 周：测试系统响应时间和并发用户数等重要参数，测试不同模块的运行情况，并重新改进代码，从而提高系统的可视化情况。</p> <p>第 11 周：提交论文初稿。</p> <p>第 12 周：提交论文外文译文。</p>

	<p>第 13 周：提交论文文献综述。</p> <p>第 14 周：提交论文指导记录</p> <p>第 15 周：提交毕业设计。</p> <p>第 16 周：完成毕业答辩。</p>
指导教师意见	<p style="text-align: right;">签 名： 年 月 日</p>
备注	