基于文献的中国近代史知识图谱构建与实证研究*

曾桢 赵浩宇
（贵州财经大学信息学院，贵阳 550025）

摘要：历史文献资源的内容组织方式通常采用非结构化文本形式进行记录，缺点在于内容之间的系统性和语义性不足，在一定程度上阻碍历史文献资源的深层次利用和开发。因此本文提出中国近代史相关历史要素资源语义描述与知识组织的思路和方法，并基于这一思路构建中国近代史历史本体模型，在此基础上通过Flask框架开发前端平台，实现前端应用层服务，完成中国近代史知识图谱的实证研究。依托本体模型，通过对中国近代史知识图谱的实例构建，实现历史知识元之间的细粒度关联，完成知识图谱可视化展示和知识查询，方便学者和用户对相关资源的开发和利用，为进一步的深入研究提供参考。

关键词：中国近代史；文献资源；领域本体；知识图谱构建
中图分类号：G250.7　DOI：10.3772/j.issn.1673-2286.2022.04.005
引文格式：曾桢，赵浩宇. 基于文献的中国近代史知识图谱构建与实证研究[J]. 数字图书馆论坛，2022 (4)：35-42.

传统史书内容的组织方式往往只能揭示一个维度的信息而弱化了其他维度的信息[1]，这对专业学者和普通用户进行语义检索和分析对比造成了困难。从传统历史文献的知识组织方式来看，若以人物活动刻画历史，不利于历史事件整体维度的把握；若按历史时间划分，则会弱化人物活动维度。因此，仅用一种知识组织方式很难展现丰富多元的历史文献资源内容，不利于用户对其感兴趣的历史信息进行宏观把握和深层了解。

随着信息技术的快速发展，历史文献资源大都完成了数字化转型升级，但传统图情领域的知识组织方法面对海量的文献资源却显得力不从心[2]，如分类法、主题法所采用的传统知识组织方式相较于机器语言而言，其组织方式单一、语义表达性较差，很难发现知识资源之间隐含的复杂关系，因此一些有价值的信息被淹没在数字化的海洋里。此外，各种形式结构的中国近代史文献资源零散分布在不同的馆藏机构和互联网中，海量的数据成为封闭的孤岛[3]，难以充分发挥其潜在的价值。更智能地实现多源异构历史文献资源语义关联和深度融合的主要任务就是将现有的异构数据集成起来，让计算机能够自动识别和处理，所以必须建立统一的标准体系，即本体[4]。构建中国近代史本体有以下作用：①厘清历史概念之间的关系，扩充中国近代史本体词表；②对中国近代史知识进行多维度描述，将人物、事件、组织机构、地点等不同实体相互关联，有利于对历史内容的宏观把握和深层了解；③通过本体模型构建中国近代史知识图谱，以节点和边的形式对中国近代史知识进行细粒度的展示，实现中国近代史的可视化展示和知识查询。

因此，本文在借鉴国内外相关研究成果的基础上，提出中国近代史相关历史要素资源的语义描述与知识组织的思路和方法，并基于这一思路构建中国近代史的历史本体模型，在此基础上完成中国近代史知识图谱的实例展示，以期实现其可视化操作、复杂语义检索以及知识发现等应用层服务。

* 本研究得到国家自然科学基金项目“基于知识图谱的农产品价值链信息融合研究”（编号：2020XSXM）资助。

1 相关研究

本体源于哲学中的本体论，侧重于对“存在”进行抽象的刻画与描绘。Neches等[5]是人工智能领域最先为本体下定义的学者。Gruber[6]将本体定义为概念化的明确的规范说明。Borst[7]认为本体是一种共享的概念模型。计算机领域的本体侧重于模拟人类对世间万物认知的行为方式，展现出认知的概念体系，以及概念之间的语义关系，而提出本体的一个重要动机是知识的共享与复用，以及数据之间的互联互通。

随着本体研究的逐步成熟，结合语义网技术开展相关研究日益成为图书情报领域所关注的焦点[8]。已有学者开展了中国近代史领域的本体和知识图谱的构建与应用。如陆伟忠[9]以“国共合作”为题材构建了国共合作历史本体，并实现了语义检索服务的本体应用。吴丽杰[10]以“东北抗战史”特色数据库为实例探讨特色数据库本体构建模式。梁恩平[11]对近代史研究者研究方向进行了梳理，利用Protégé构建了近代史研究者兴趣领域本体，并提出了历史档案资源的个性化推送策略。陈玖瑜[12]依托数字人文理论和语义网相关技术挖掘出了民国文献知识元之间的语义关联，设计了民国报纸本体，并以历史人物梅兰芳为实例完成了知识图谱的可视化展示，实现了民国时期报纸内容知识元的细粒度关联。孙辉等[13]探索了国史领域知识的特征，提出国史本体的构建步骤，实现了本体知识实例的可视化展示。王颖等[14]基于国史本体框架，利用Neo4j图数据库作为数据仓储，实现了国史知识的可视化展示和检索、问答等服务层应用，为国史领域知识的深度检索服务提供了重要参考。张云中等[15]在构建红色历史人物知识图谱schema基础上设计了知识问答服务架构，提升了用户的检索体验。王帅奇等[16]对中国革命历史档案资源进行开发，构建了革命战争知识图谱。刘伟丽[17]构建了中共一大人物知识图谱。葛勇文[18]构建了中国近代革命文物知识图谱，并实现了革命文物知识图谱的应用。可见，结合本体、知识图谱等语义网技术，深入挖掘中国近代史相关事件细粒度的语义特征，顺应了当前研究中国近代史的需要，具有很强的现实意义，但目前覆盖中国近代史文献资源全领域的本体建模相对较少，建模深度较浅，粒度较粗，本体开发的系统性和可扩展性有待提高。因此本研究将视角聚焦于中国近代史本体建模，实现中国近代史文献资源的关联与聚合，为中国近代史文本内容的知识组织和表示提供新方法，为中国近代史知识图谱的实证研究提供新思路。

2 数据来源及研究框架

2.1 数据来源

本研究所需要的数据为电子形式的中国近代史文献资料，主要选取历史名人数据、中国近代史历史大事记等具有历史典型特征的文本数据作为本研究的基础支撑。其中文本形式的资料以《简明中国近代史读本》《中国近代史》《中国近现代名人生平暨生卒年录（1840—2000）》《中国近代人物录》等著作内容为主。历史人物数据主要来自国家图书馆人物专题数据库、孙中山故居纪念馆相关人物专题库、维基百科以及百度百科等。历史大事记主要来自网络论坛、开放数据集、垂直站点等多种数据源。其中，开放数据集是结构化数据的主要来源，专题数据库和百科是半结构化数据的来源，从中国近代史书籍和垂直站点获取的是非结构化形式的文本内容。针对以上数据源主要采用网络爬虫、人工筛选、自然语言处理等方式获取相关数据。

2.2 研究设计

研究的主要工作是实现中国近代史文献资源内容细粒度知识元的语义化表示，因此设计了中国近代史本体模型，并基于此模型完成知识图谱实证研究，实现知识内容可视化展示和知识检索等应用。研究思路如图1所示，分三步实现知识图谱构建。

首先，进行模式层的搭建，通过系统的调研分析，确定中国近代史知识图谱所需要的具体数据；其次，通过深入剖析文本内容特征以及结合领域专家知识来设计相关概念、关系及属性，运用Protégé构建中国近代史本体，完成“中国近代史”知识建模；再次，基于设计好的本体库，利用自然语言处理技术扩充实例数据，根据不同形式的数据类型采取不同的方法对其进行抽取；最后，将抽取得到的实例知识进行整合处理，将其导入Neo4j中，并通过Web前端完成知识的可视化呈现，实现中国近代史知识图谱的实例构建。整个构建过程具有非领域性和非针对性，因此该方法不仅适用于中国近代史领域本体构建，而且适用于其他领域本体模型的构建。

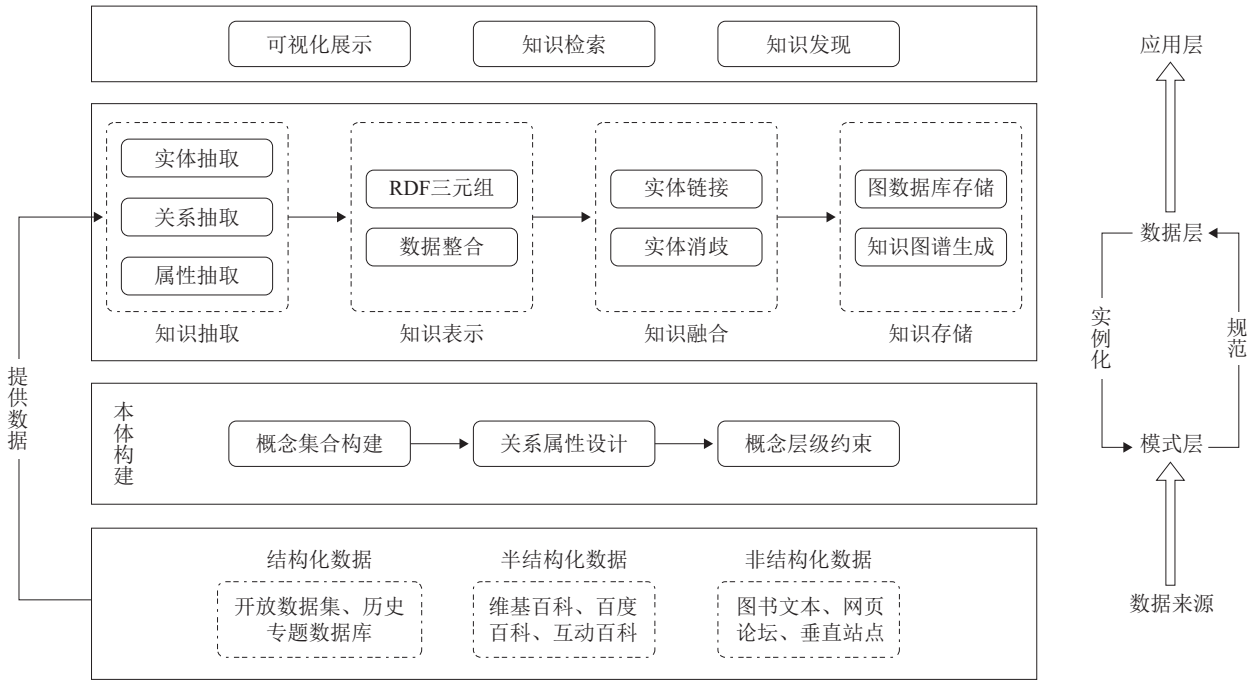

图1　研究思路

3　中国近代史知识图谱构建

3.1　模式层构建

模式层即知识图谱的本体模型[19]，是对数据层的约束和规范，通过本体模型刻画出中国近代史的核心概念体系。构建本体的根本目的在于为某一特定领域提供一套被广泛接受、认可和共享重用的概念体系[20]，使得领域知识能够被重用，避免“重复造轮子”的情况。本研究根据文献资料的关键词，结合历史领域相关学者的专业知识，考虑实际情况提炼出最具代表性的核心概念作为中国近代史本体的核心类目，使用Protégé工具并结合“七步法”构建中国近代史本体模型。具体步骤如下。

第一步，确定中国近代史本体的构建范围。根据需求分析确定构建本体的对象，以中国近代史文献资源为主要参考资料，确定以中国近代史内容要素为研究对象。

第二步，寻找可复用的本体。通过DAML、Ontolingua、Protege本体库，调研可复用本体的类和属性并进行引用，发现Foaf、Time Ontology、DC terms、EventKG、Org、CIDOC-CRM、BIBFRAME等本体中的相关概念和属性可复用，但是目前可复用的本体模型不能完全满足中国近代史细粒度知识描述的需要，因此笔者在基于关联数据发布准则的基础上，根据需要自定义类和属性，构建中国近代史本体模型CMH（China’s Modern History），用缩写“:cmh”作为前缀名称定义中国近代史本体的描述词汇。

第三步，列举出中国近代史内容中的重要术语。与相关历史专家和学者进行交流沟通，认真听取其意见，并结合网上调研，对中国近代史的相关知识内容做了系统的梳理分析，最终凝炼出10个最具概括性的核心概念作为中国近代史本体的一级类目。

第四步，定义本体分类体系。根据中国近代史历史知识元素和重要术语，对其进行归纳分类，确定本体模型中包含的类及其层次关系，逐渐构建完整的层级体系。在最顶级owl:Thing类目下面设置“历史人物”“历史事件”“历史文献”“地点”“时间实体”“历史时期”“思想理念”“领域”“行为主体”“组织机构”10个核心概念。通过对概念的层次体系进行构建，能够较好地抽象出中国近代史知识的概念体系，更真实地还原历史细节。

第五步，定义本体属性及关系。定义中国近代史本体数据属性可以丰富对历史实例的描述，扩展实例含义；类之间的相互关联通过定义对象属性来完成，对象属性的建立可以方便中国近代史知识图谱进行语义关

联和知识发现。例如，“历史事件”类通过sem:hasActor属性与“历史人物”类相互关联，“历史文献”类通过dc:creator属性与“历史人物”类相互关联。本体类之间的部分关联如图2所示。

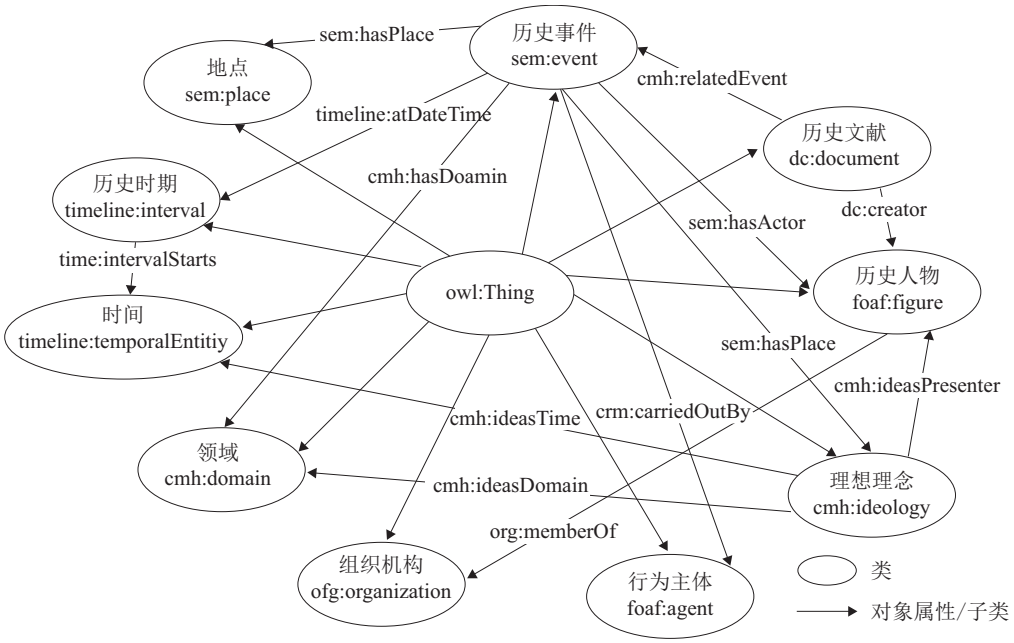

图2 中国近代史本体模型图

第六步，定义本体属性约束。本体属性约束就是对属性添加语义关系约束，具体包括三方面：一是定义属性的定义域和值域；二是定义属性的特性；三是定义属性的限制。属性约束的目的在于减少语义冲突，提升本体推理能力。

第七步，建立中国近代史本体实例。完成中国近代史本体类、对象属性、数据属性的添加后，需要利用Protégé对中国近代史相关人物、事件、地点、时间等实例和属性值进行添加，形成中国近代史知识组织体系。

按照七步法构建完毕后，中国近代史本体模型共有10个一级类、53个二级类、88个三级类，以及包括数据属性和对象属性在内的95个属性约束。有关中国近代史领域范畴本体构建的研究，大多数学者都围绕某一特定题材或专注于特定历史要素进行语义建模，而涵盖中国近代史全领域的本体构建研究相对较为缺乏。本研究构建的本体模型，涵盖中国近代史这一特定历史时期内的人物、事件、文献、地点、机构等重要实体概念，使单一的历史要素之间融合成相互关联的有机整体，拓宽了先前学者所构建的本体范围，补充了领域本体术语词表，加深了中国近代史历史要素之间的关联性和系统性。目前，中国近代史本体主要涵盖历史人物和历史事件的基本信息、人际关系、历史事件的因果关系、历史文献的著述信息及思想内容等多维度信息。此外，还可以依据本体构建生命周期理论，根据需求的变化而动态扩充实体、关系和属性。中国近代史本体模型的构建为知识图谱应用层的搭建提供了基础支持。

3.2 知识获取

知识获取是将半结构化和非结构化数据转换为构建知识图谱数据层所需要的实体和关系的过程。因此根据数据来源的不同，本文通过网络爬虫、模式匹配、包装器适配等方法，采集相关数据。通过使用HanLP、Jiagu等自然语言处理工具包完成实体识别、信息抽取等任务，抽取所需要的实体、关系和属性，并将其转换成实体关系三元组。

结构化数据具有良好的层次结构，通常存储在数据库中。本文从中文开放知识图谱（OpenKG.CN）中获取“中国近代历史人物知识图谱”开放数据集，包含近1 300位中国近代史人物的结构化数据。

中国近代史人物实体属性的来源通常是百科网站中的Infobox模块的半结构化数据。因页面格式基本固定，遂采用包装器方法对网站内容进行解析实现数据

自动采集，并将其存储到关系数据库。

非结构化数据通常是文本资源，其内容完整，数据丰富，它是知识图谱实例数据的主要来源，也是抽取任务的难点。鉴于选择的文本数据缺乏大规模词性标注数据集，因此本文采用规则和深度学习神经网络模型相结合的方式来抽取三元组知识。经过文献调研发现，中国近代史文本内容中包含大量的著作、条款、会议、事件、条约、日期等内容，其特点是表达形式比较固定，规则性较强，易于提取知识元素，因此针对此类型的数据主要采取基于模式匹配的方法抽取。最直接的方式就是将文本内容视为字符序列，构造正则表达式的字符模式，实现抽取。其余实体的抽取主要通过Hanlp开源工具包、自定义词典和规则相结合的方法自动抽取文本中的实体。hanlp工具对特定领域中的实体识别具有较高的准确度[21]。实体的属性和关系利用jiagu深度学习神经网络开源模型进行抽取，抽取的结果以三元组的形式表达出来。Jiagu深度学习神经网络开源模型是使用大规模语料训练而成，并且提供中文分词、词性标注、命名实体识别、关系抽取等常用自然语言处理功能，得益于已训练好的模型，其使用时无须对数据进行标注。

通过以上方法采集的数据大多需要逐条筛选进行二次过滤，剔除无关、重复数据，完善缺省数据，保障收集到的数据具有较高的质量。本文利用上述方法半自动获取人物实体及其属性5 507个，获取中国近代史中具有重要意义的历史事件及其属性177个，人物间关系7万余对。

3.3 知识表示

知识表示[22]是把人类知识表示成机器可以理解的数据结构和系统控制结构的策略，知识表示是知识组织的前提和基础。知识表示的形式大致可分为3种：三元组的形式、图结构的形式以及低维稠密向量表示的形式。本文使用RDF数据模型对中国近代史本体概念和关系进行形式化表示，使计算机能够理解数据模型。由于RDF三元组是由“点-边-点”组成的有向语义网络图，本质上属于图形模式的数据结构，因此可以与图结构数据相互映射。例如，三元组中每个实体对应Neo4j图中的一个节点，属性和关系对应图中的有向边。数据层以实例数据为对象，为方便下文使用Neo4j作为知识图谱的数据仓储，因此对RDF进行格式转换，以备导入Neo4j图数据库中使用，并完成从“实体-属性-属性值”或“实体-关系-实体”的三元组形式到Neo4j的对应。将本体中的类映射为图中的实体节点，本体的类间关系映射为图中节点的边，本体属性映射为图中节点的属性，从而实现本体模型到Neo4j的映射。在Neo4j图数据库中，数据属性以键值对的形式作为对节点特征的描述，对象属性作为节点和边的关联形式进行表示。图3为李鸿章人物信息属性图，数据属性表示为<李鸿章-民族-汉族>，对象属性表示为<李鸿章-任职机构-清政府>等。

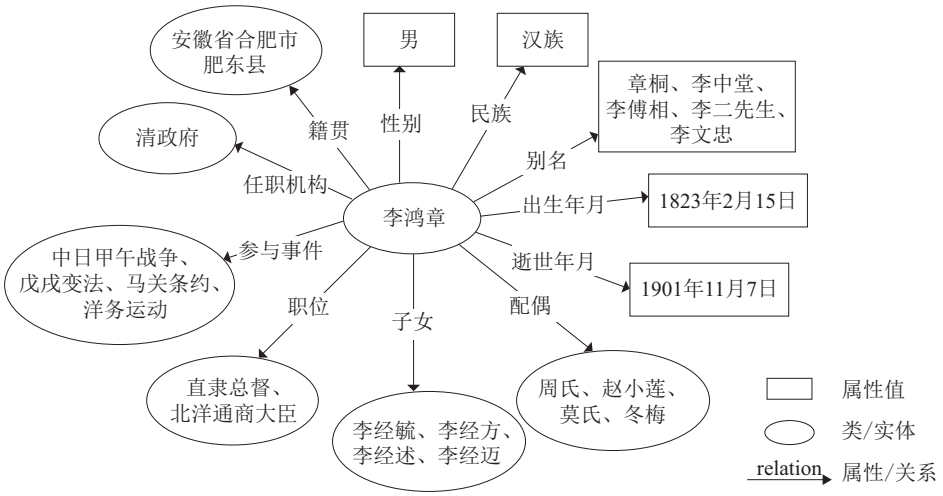

图3 李鸿章人物信息属性图

3.4 知识融合

中国近代史知识的融合包括本体层中概念、关系、属性的融合，以及数据层中实例、属性值的融合。概念层的融合即将中国近代史本体模型与其他相关本体中等价类或属性建立映射关系，实现模式层的语义融合，不同本体相同的类和属性用owl:equivalentClass和owl:equivalentProperty进行关联。例如本文中“历史文献”类所包含的“文献名称”概念在文献组件本体（DoCO）中表示为doco:title，而在书目框架本体（BIBFRAME）中描述为bf:workTitle，因此可以使用owl:equivalentClass属性完成不同本体之间相同类的映射。通过概念层的融合可以发现更多相似的本体，实现本体概念和属性的扩展，让本体得到充分的共享，发挥本体的最大作用。数据层的融合包括实体链接和实体消歧。实体链接是将中国近代史实体实例与本体中的概念进行相互映射，例如，若两个含义相同的不同实例进行相互关联，则使用owl:sameAs属性来表示两者身份的同一性，owl:sameAs属性表示两个不同URI的引用实际上指的是同一事物，两个实体具有相同的“身份。crm:isEqualInTimeTo属性用来融合两个不同的时间表示方式，以此来表示两个不同的时间表示方式指的是同一个时间点或时间段。例如，清帝退位时间按照皇帝年号纪年是“宣统三年十二月二十五日”，而用公元纪年法则是“1912年2月12日”，因此用crm:isEqualInTimeTo属性来表示二者指代同一天。

实体消歧旨在解决不同名称的实体含义相同的问题，消除实体的多样性和歧义性。同一地点在古代和近代往往存在不同的名称，为了实现地名的统一，本文利用中国省市县区域划分开放数据集作为实体链接的标准数据，然后将已抽取的中国近代史相关地名与其进行实体链接，完成实体对齐工作。其他实例，包括“历史人物”“历史事件”“历史文献”等赋予唯一标识符URI并使用“别名”等属性进行辅助识别，完成实体消歧。例如，在中国近代史文献中，“直隶”往往指皇帝所在的心腹之地，又称“京师”，现今为“河北省”的管辖范围。使用OWL语言进行表示为：

```
<rdf:Description rdf:about="#直隶/京师">
        <owl:sameAs rdf:resource="#河北"/>
</rdf:Description>；
```

3.5 知识存储

中国近代史本体的构建，标志着知识图谱模式层的完成。本文在中国近代史本体框架的基础上增加相关实例，完成中国近代史知识图谱的实例构建。知识图谱的可视化呈现是通过图形化的形式表现出来，因此选择合适的存储方式至关重要。本文使用Neo4j作为数据仓储，完成知识图谱的实例构建。Neo4j为多种语言提供了API接口[23]，如Java、Python、C#等。下面将通过Python语言和Cypher命令句对Neo4j进行读写操作。首先将上文抽取得到的数据进行整合处理，转换成CSV格式文件并存储到Neo4j根目录下的import文件，使用Cypher命令语句LOAD CSV将人物、事件节点及其属性导入图数据库Neo4j中，然后再将人物和事件所对应的关系导入其中。其次，因为人物间关系数量较多，且存储格式为三元组的形式，所以选择更快捷方便的Python第三方库Py2neo将其导入Neo4j中，完成知识的可视化呈现，为中国近代史知识图谱实证研究打下基础。完成知识存储后，数据库中共有包括人物、事件、地点、职位、作品、毕业学校等各类实体节点数量11 768个，各类关系在内的16 592条边。

4 知识图谱实证研究

知识图谱最重要的作用就是把知识以图的形式展现出来，图中的节点和关系一目了然，得益于边与边之间的相互链接，可以沿着相邻节点依次发现相互关联的新知识，最大程度地为用户节省时间和精力。本文以中国近代时间段内的相关人物和事件为例进行知识图谱的实证研究，以期探寻人物之间的深层关系和人物与事件的参与关系。本研究使用HTML+CSS+D3（jQuery）技术构建前端展示平台，使用基于Python的Flask框架搭建后端服务，并利用Neovis.js可视化组件与Neo4j图数据库进行连接并对其进行操作。该平台立足于中国近代史领域，以相关历史要素为核心，构建一个包含浏览与检索功能的展示平台，实现中国近代史知识图谱的可视化展示和相关应用。

（1）知识图谱的可视化展示。中国近代史知识图谱的展示功能体现在两个方面：一是浏览功能，即以图的形式对知识元进行部分或全部展示，并且支持节点的放大、缩小以及节点属性详情的浏览，让用户能够从宏观层面把握中国近代史知识脉络；二是词云展示功

能，即通过对中国近代史文本内容进行分词和词频统计的直观展现，并生成相应的词云图。

（2）知识图谱的相关应用。知识检索是知识图谱应用层的一项基本功能，中国近代史知识图谱的检索功能可实现历史人物和历史事件的查询，人物知识图谱能直观地了解人物间关系，方便发掘人物之间的隐含关系，事件知识图谱可深入挖掘事件之间错综复杂的关联关系，能更好地把握历史事件发展的趋势和脉络。

知识图谱的检索功能不但可以迅速返回结果，而且可以根据已存在的逻辑关系发现新的实体间关系，实现对隐性知识的挖掘。Neo4j图数据库使用的是Cypher查询语言，形式与SQL查询语言较为相似，它是一种声明性模式匹配语言，可以通过简单的语法规则进行非常复杂的查询。例如，在前端历史人物知识图谱中查询与“陈独秀”相关的节点，查询结果如图4所示，从图中可以看出陈独秀和李大钊共同参与了新文化运动，和胡适共同参与了“五四运动”等历史事件，从图中也能看出与陈独秀相关的属性信息及其丰富的人际关系等。

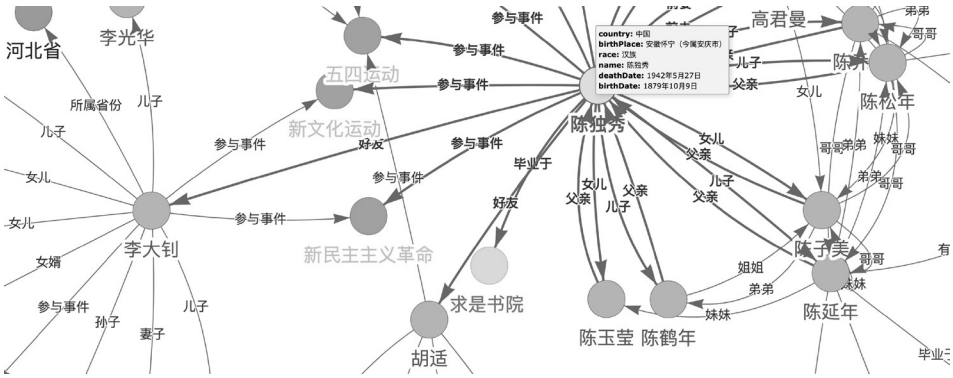

图4　知识检索可视化展示

本文构建了十分丰富的人物关系知识图谱，包括父母、姐弟、战友、好友、师生、领导等在内的117个人物间关系类型。研究中国近代史很重要的一点在于厘清历史人物之间错综复杂的关系，发现人物之间的隐性关系，深入挖掘人物的潜在历史价值。

在不知道两个人物之间有何关系时，可以通过多深度关系节点查询来发现人物节点之间的关系。当需要实现中国近代史相关人物的多深度关系节点查询时，可以使用Cypher语法中的深度运算符来完成查询操作，Neo4j可以快速地对实体节点完成图遍历，并且可以计算出各节点的路径，利用路径关系推导出节点间的联系。

综上所述，历史内容通过书籍或网页形式的非结构化数据进行展示时，会浪费用户大量的时间和精力去挖掘、揭示各实体间的隐含关系，而知识图谱能以最直观的形式为历史爱好者提供相关人物与事件的知识查询，为了解历史人物和事件之间的复杂关系提供新的视角和方法。

5　结语

本研究从历史文献资源的开发利用入手，以本体和知识图谱等语义网技术为手段，从文本资料中筛选出相关概念及概念间关系，构建了能够揭示细粒度知识元之间语义关系的中国近代史本体模型，完成了中国近代史知识图谱的实例构建，实现了各实体属性的细粒度知识关联，又以具体的历史人物与事件为样例进行查询验证，在理论与实践上证明了知识图谱技术在中国近代史研究上的可行性，并形成了较为完备的研究思路。

中国近代史知识图谱的构建，为相关学者探究中国近代史知识提供了便利，为历史人文研究数字化提供了技术支持，在一定程度上丰富了历史文献资源的开发利用，因此本研究既是一次有价值的尝试，也为后续相关学科交叉研究提供借鉴和参考。为确保数据的准确性，利用人工手段对爬取的数据进行清洗，数据质量高但效率较低，同时对于存在于非结构化文本中的人物和事件实体识别的能力和方法有待进一步完善，下一步的研究需要根据中国近代史文献的内外部特征，建立整个中国近代史文本标注的语料库，以期从海量的文本中更准确地获取数据，降低人工参与度，扩充知识图谱实体数量，补充实体属性，从而为用户提供更完善的智能推荐、知识推理、语义问答等应用层服务。

参考文献

[1] 纪姗姗，赵炳容，刘峥. 国外知识组织体系管理工具比较分析与启示 [J]. 图书情报工作，2020，64（24）：73-83.

[2] 贾琼，王萍. 基于关联数据的历史档案资源聚合研究 [J]. 图书情报工作，2021，65（10）：105-112.

[3] 王世伟. 重新认知中国智慧图书馆发展的历史方位 [J]. 图书馆理论与实践，2022（1）：1-6.

[4] 罗婷婷，李娇，鲜国建，等. 基于OWL+SKOS的期刊本体构建与应用 [J]. 数字图书馆论坛，2018（12）：49-54.

[5] NECHES R，FIKES R，FININ T W，et al. Enabling technology for knowledge sharing [J]. AI Magazing，1991，12（3）：36-56.

[6] GRUBER T. A translational approach to portable ontologies [J]. Knowledge Acquisition，1993，5：199-220.

[7] BORST W N. Construction of engineering ontologies for knowledge sharing and reuse [J/OL]. [2022-02-02]. https://www.researchgate.net/publication/41175785_Construction_of_Engineering_Ontologies_for_Knowledge_Sharing_and_Reuse.

[8] 冯惠玲，连志英，曲春梅，等. 回顾与前瞻：“十三五”档案学科发展调查和“十四五”档案学重点研究领域展望 [J]. 档案学通讯，2021（1）：4-15.

[9] 陆伟忠. 基于本体论的信息检索框架 [D]. 武汉：武汉大学，2005.

[10] 吴丽杰. 基于本体的特色数据库知识组织研究 [J]. 图书馆学刊，2012，34（3）：41-43.

[11] 梁恩平. 基于近代史研究者兴趣图谱的档案资源推送研究 [D]. 长春：吉林大学，2020.

[12] 陈玖瑜. 数字人文视阈下民国报纸知识图谱构建研究 [D]. 长春：吉林大学，2020.

[13] 孙辉，王颖，张智雄. 本体构建中的协同问题研究——以中华人民共和国史本体为例 [J]. 情报学报，2015，34（9）：958-969.

[14] 王颖，张智雄，孙辉，等. 基于本体的国史知识检索平台构建研究 [J]. 图书情报工作，2015，59（16）：119-128.

[15] 张云中，郭冬，王亚鸽，等. 基于知识图谱的红色历史人物知识问答服务框架研究 [J]. 图书情报工作，2021，65（16）：108-117.

[16] 王帅奇，陈晓美，邓君，等. 中国革命历史档案知识图谱构建 [J]. 兰台世界，2021（12）：39-44.

[17] 刘伟丽. 中共一大人物知识图谱构建研究 [D]. 保定：河北大学，2021.

[18] 葛勇文. 革命文物知识图谱构建研究 [D]. 保定：河北大学，2021.

[19] 肖仰光. 知识图谱概念与技术 [M]. 北京：电子工业出版社，2020.

[20] 贾君枝. LAM馆藏资源的元数据整合方法比较分析 [J]. 档案学研究，2022（1）：79-84.

[21] 何晗. 自然语言处理入门 [M]. 北京：人民邮电出版社，2019.

[22] 刘鹏. 知识表示与处理 [M]. 北京：电子工业出版社，2021.

[23] JIANG Z，CHI C，ZHAN Y. Research on Medical Question Answering System Based on Knowledge Graph [J]. IEEE Access，2021（99）：1.

作者简介

曾桢，男，1982年生，博士，教授，研究方向：信息资源管理、知识图谱。

赵浩宇，男，1996年生，硕士研究生，通信作者，研究方向：知识组织、知识图谱，E-mail：galaxyeric582@163.com。

The Knowledge Graph Construction and Empirical Research of Modern Chinese History Based on Literature

ZENG Zhen ZHAO HaoYu

(Information School, Financial and Economics of Guizhou University, Guiyang 550025, P. R. China)

Abstract: The content organization of historical document resources is usually recorded in the form of unstructured text, which has the disadvantage of insufficient systematization and semantics among the contents and to a certain extent hinders the deep utilization and development of historical document resources. Therefore, this paper proposes the ideas and methods of semantic description and knowledge organization of historical element resources related to modern Chinese history. Based on this idea, the historical ontology model of modern Chinese history is constructed. We develop a front-end platform through the Flask framework to realize front-end application layer services and complete the empirical study of the knowledge map of modern Chinese history. Based on the ontology model, the fine-grained correlation between historical knowledge elements is realized through the construction of the instance of the knowledge map of modern Chinese history, and the visualization display and knowledge query of the knowledge map are completed, which facilitates the development and utilization of related resources by scholars and users and provides reference for the subsequent in-depth research.

Keywords: Modern Chinese History; Literature Resources; Domain Ontology; Knowledge Graph Construction

（收稿日期：2022-03-28）