

名人年谱资源的知识图谱构建^{*}

——以徐朔方《晚明曲家年谱》为例

徐彤阳 黄映思

(山西财经大学信息学院, 太原 030006)

摘要: 以数字人文的研究范式与知识图谱工具相结合, 实现名人团体文化的深度挖掘与展示, 将传统名人年谱资源转变为共享性的立体数字资源。以徐朔方《晚明曲家年谱》作为研究素材, 借助Protégé工具构建晚明曲家年谱本体, 结合Neo4j图数据库完成年谱知识图谱的构建, 以揭示当前传统年谱资源的数字研究价值与数字化再利用, 提出未来传统名人年谱资源走向数字化与智能化深度研究的相关展望。

关键词: 数字人文; 年谱; 本体; 曲家; 知识图谱

中图分类号: G250 **DOI:** 10.3772/j.issn.1673-2286.2022.12.004

引文格式: 徐彤阳, 黄映思. 名人年谱资源的知识图谱构建——以徐朔方《晚明曲家年谱》为例[J]. 数字图书馆论坛, 2022 (12) : 29-36.

目前数字文化建设广受重视, 推动中华文化“走出去”, 不仅需要文化工作者努力, 更需要推进文化的数字化开发和跨文化传播, 赋予中华文化“活的灵魂”。年谱可视化的方法重点展示了领域中人物之间的社会关系, 通过挖掘各个人物的历史时间、社会关系从而展示该人物在领域中的历史地位, 更直观地展现出人物为其所在的领域文化带来的影响。在传统文化数字化传承与发展的过程中, 数字化转型是关键的一步, 传统年谱资源的研究可以为传统文化数字化转型提供技术范式。

年谱是我国一种特殊的人物传记资源, 是按照时间节点考述一个人物或者群体人物的生平事迹, 其活动轨迹在特定的时间与空间上进行, 是一种以传记体裁为主的文献资源, 其本身可以为学术研究提供翔实的实证资料, 具有重要的学术价值和研究意义^[1]。我国的年谱发展自北宋以来已有上千年的历史, 《中国历代年谱总录》中共著录年谱3 015种, 加上族谱约有1万种^[2]。家谱和年谱都是一种经后人进行修缮而汇编的特殊文书。家谱是一种以血缘关系为主体记录的家族世系繁衍信息

和重要人物等信息, 具有家族内部的可追溯性。而年谱是按年月记载人物生平事迹的传记体裁, 通俗来说, 就是以时间为顺序编排谱主的生平活动和经历的一种特殊史书, 具有个人史和编年史的性质, 属于一种特殊的人物传记文书。目前年谱的数字化利用程度仍比较低, 数字化研究更少, 主要还是面向家谱文本资源比较多, 使得年谱本身所蕴含的价值未能充分体现。现阶段, 我国年谱的编撰随着数字人文研究范式的发展, 重新涌起了研究热潮^[3], 如何在创新年谱资源编撰方式的同时充分利用年谱资源进行领域文化的学术研究^[4], 是目前对于年谱资源数字化再利用且充分利用的关键点之一。此外, 年谱作为谱系文本资源的一种, 其发展经过了皇族帝王谱系到姓氏谱系、家族谱牒, 再到年谱成为一个独立的谱系文本^[5], 其所蕴含的文化、史料价值、学术价值及教育功能随着社会需求的不断变化而不断丰富内在含义, 对年谱资源的数字化再利用及深入研究, 也是学界通过学术研究的渠道传播年谱资源的文化、引导等作用的重要方式。

^{*} 本研究得到2022年度山西省回国留学人员科技活动择优资助项目“社会共建视角下山西省突发公共卫生事件数字记忆平台构建研究”(编号: 20220026) 资助。

1 研究综述

近年来,针对谱系文本的研究开始崭露头角,研究范畴较为主要集中在家谱文书资源,例如,在家谱本体构建方面,陈艳^[6]、夏翠娟等^[7]对中国家谱本体模型展开了研究,分别构建了中国家谱的领域知识本体及上海图书馆家谱本体模型,从领域知识单元组织关系的角度重新解读了中国家谱资源。在本体知识图谱构建方面,张政平等^[8]提出融合骨架法和七步法构建孔子世家谱本体的方法,并利用Protégé、Neo4j等工具构建本体并实现孔子世家谱知识图谱的构建;祝振媛^[9]则采用本体构建技术实现《吴氏西宅宗谱》的本体模型架构,并依据本体模型完成吴氏家谱知识谱系的实践;姜赢等^[10]基于Java EE框架B/S模式构建了家谱知识图谱检索系统,系统功能包括家谱检索、人物检索、字辈排行、追根溯源等;吴信东等^[11]基于知识图谱,开发融合了知识图谱构建技术的家谱数据中台——华谱系统(Huapu-CP)。

学界家谱本体的研究充分验证了本体技术、知识图谱与谱系文本融合研究的可行性,但是对年谱资源进行本体与知识图谱的构建依旧很少。笔者将尝试利用本体模型构建年谱知识图谱,在家谱本体构建模型的基础上,考虑现有模型复用的可能性,以徐朔方先生编纂的《晚明曲家年谱》作为原始材料构建年谱本体,借助中国历代人物传记资料库(CBDB)及其他历史数据库查询补充与纠正相关资料,结合知识图谱,深度挖掘晚明曲家核心人物的家族、作品、历史事迹等,探讨构建易读的、具备共享性的名人团体年谱知识图谱。知识图谱是利用图形之间的联系反映知识之间的关联关系,通过语义关系、关键词索引、知识实体构架联系等方式,强化知识之间的关系^[12]。知识图谱的构建方法^[13]主要包括自顶向下和自底向上两种体系,自顶向下通常应用于领域知识图谱,是一种通过定义完善的本体模型及其关系属性,依据原始数据进行实体抽取,并填充相应数据实例,以完成领域知识图谱的构建^[14];自底向上的体系多用于数据量较大的知识图谱构建,首先需要先从数据中抽取实体、关系、属性,再进行数据层的知识填充与处理,最终实现知识图谱的整体构建^[15]。笔者将依据自顶向下的架构方法论体系完成晚明曲家年谱知识图谱的构建工作,为后续的传统年谱的转型研究提供一定的借鉴参考,这也符合当前对于年谱资源数字再利用的需要,具有较强的现实意义。

2 晚明曲家年谱的本体建模设计思路

2.1 研究对象

本文将徐朔方先生所著的《晚明曲家年谱》(浙江古籍出版社,1993年版)作为主要研究素材。《晚明曲家年谱》共三大卷,分为苏州卷、浙江卷、皖赣卷,共收录起于徐霖终于金圣叹^[16],约二百年间的39位戏曲家年谱;囊括了明中后期至清初的重要曲家,几乎涵盖了晚明戏曲领域的重要人物,对全面了解明代戏曲的发展乃至名人年谱的编撰,甚至是对整个学术界都有非常大的价值,对促进年谱资源再利用有着实践意义。因此,对名人年谱资源展开知识图谱构建研究,是将年谱资源推向数字化发展的关键渠道。在原始数据预处理方面,首先利用OCR文字识别软件对扫描文件进行文字识别转录,然后人工校对古籍原文,最终形成研究所需的年谱文本。

2.2 年谱本体构建设计思路

《晚明曲家年谱》记载内容是按曲家籍贯划分,有年谱、行实系年、事实录存,所蕴含的知识本体类别单一又多重,如何准确提取分类点,设计一个充分融合年谱内容的知识本体类别框架,以便对知识本体进行因果关系的提取,这是晚明曲家年谱本体构建所面临的难点。

在本体模型的设计方面,为了能够实现对晚明曲家年谱中各个人物大事记的因果关系等方面的构建,笔者结合中国历代人物传记资料库、上海图书馆家谱知识服务平台等开放数据库查阅相关人物资料,以辅助本体模型中各个实体的分类与因果关系揭示。笔者还参考了上海图书馆家谱知识服务平台^[17]、中华寻根网^[18]以及现有的家谱本体模型的相关研究,提取适合晚明曲家年谱知识本体的分类与因果关系描述方法,其中Zeebaree等^[19]将TM模型的本体论进行了改进,从静态描述开始,发展出一个动态模型,以确定指定行为的事件,从而对概念模型中的因果关系进行分析,将一个概念转变成为可识别的模型内部结构的实体。Zeebaree等还将TM模型应用于分析实体的因果关系,因此本文将借助Zeebaree对因果关系的研究范式,对晚明曲家年谱中知识本体类别的因果关系进行提取和分析。

2.3 晚明曲家年谱本体构建方法选取

目前本体模型构建的研究一般围绕“本体模型-关系型数据库”的研究范式而开展研究工作, 比较常用的骨架法(又称ENTERPRISE法)和斯坦福大学的七步法, 结合通用关系型数据库辅助本体模型数据的存储与提供信息检索服务, 以满足用户需求。骨架法^[20]主要包括确定本体应用的目的和范围、本体分析、本体表示、本体评价、本体建立; 七步法主要包括确定本体的专业领域和范畴, 考查复用现有知识本体的可能性, 列出本体中的重要术语, 定义类和类的等级体系, 定义类的属性, 定义属性的分面, 创建实例^[21]。前者缺乏本体模型的重用性, 易导致本体的可重构特性无法实现效用最大化^[22]; 后者步骤较为完整, 本体分析环节紧扣, 本文将考虑以七步法入手, 尝试融入其他本体方法步骤以完善整个晚明曲家年谱本体模型的构建。

笔者还考虑到晚明曲家年谱记载了晚明曲家较为详细的人物信息, 主要以时间线及人物历史大事记作为记载的内容, 包含人物的生卒年、人物评价、生平经历、宦官过往、著作作品等。相对于晚明曲家年谱数据来说, 现有的本体结构无法融合大规模的年谱数据, 容易导致数据冗余、查询体验感下降等问题。将选取Neo4j图数据库作为本文数据的处理工具, 一方面是Neo4j图数据库能够高效地处理大规模数据, 支持非常灵活和细粒度的数据模型, 且每一个节点都直接包含一个关系列表, 关系列表中能够存放该节点与其他节点的关

系^[23]; 另一方面是图数据库支持多种类型数据, 兼容性强, 能够以面向图的格式进行数据的存储和汇出, 提供了非常完善的数据库特性, 包括ACID事务的支持、集群支持^[24-25]等。

综上, 由于年谱文本为主的知识图谱构建存在一定的现实难度, 一方面是谱系领域本体构建与知识图谱方法并不全面, 现有的家谱本体与年谱本体又存在差异问题; 另一方面, 关系型数据库存储的效率较低且可视化需求无法满足用户的个性化检索需求。针对以上问题, 本文主要采用自顶向下的方式构建晚明曲家年谱知识图谱。首先在晚明曲家年谱原始数据的基础上, 融合七步法、骨架法及其他研究范式构建晚明曲家年谱本体结构; 其次根据年谱本体的结构, 对原始数据中的实体进行提取, 将实体数据整合为图数据库可以读取的CSV数据格式, 再汇入Neo4j图数据库, 完成初步的晚明曲家知识图谱构建; 最后通过Neo4j图数据库实现对数据的存储, 构建晚明曲家年谱知识图谱。

3 面向晚明曲家年谱的本体模型构建过程

结合晚明曲家年谱原始数据的特点以及人物关系的描述, 本文将结合具体的研究需要, 以七步法为基础, 借鉴已有的谱系本体构建框架并进行调整, 构建曲家年谱本体, 具体流程如图1所示。

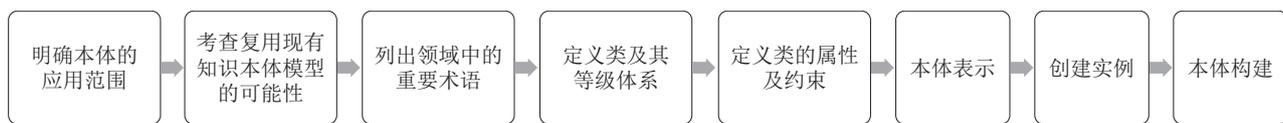


图1 晚明曲家年谱的本体模型构建流程

(1) 明确本体的应用范围。年谱本体应准确描述记载年谱人员的详细信息、历史事件信息, 由此才能厘清人物之间的关联关系, 以确保人物分类符合现存文献中的事实记录。

(2) 考查复用现有知识本体模型的可能性。由于目前以年谱作为本体构建的研究仍处于探索阶段, 而家谱与年谱之间具有一定的共同点和相似性, 通过调研已有且比较成熟的家谱领域本体模型, 对夏翠娟等^[7]构建的上海图书馆家谱本体模型、张政平等^[8]构建的孔子世家谱本体模型进行分析复用, 并参照本体的构建

方法论, 以保障如期能够实现晚明曲家年谱本体的初步构建。

(3) 列出领域中的重要术语。本文从晚明曲家年谱中提取出相应的重要术语, 包括卷次(volume)、引论(introduction)、谱序(number of the chronicle)、人物(person)、出生日期(date of birth)、死亡日期(date of death)、籍贯(native place)、性别(gender)、名(given name)、字(courtesy name)、号(pseudonym)、曲腔派系(song cavity factions)、人物评价(character comments)、著作名称(title of book)、著作年份(year of

work)、著作评价(book comments)、事记年份(year of the event)、事记(events)、历史事件(historical events)、历史活动地(historically active)、居住地(residence)。

(4) 定义类及其等级体系。将本文的术语归纳出六大核心类别,分别为卷次、谱序、人物、著作、事记数、曲腔派系,其他术语归纳核心类别的子类别。

(5) 定义类的属性及约束。将文中的其他类别进行归纳整理,将其作为属性划分给对应的类,并确立类别之间的关联。类的属性包括对象属性与数据属性。晚明曲家年谱本体的五大核心类涵盖11个对象属性和20个数据属性。对象属性包括:属于卷次、属于曲腔派系、属于谱序、有人物、有卷次、有历史事件、有引论、有派系、有社会关系、有著作、同一人物;数据属性包括:事记ID、事记年份、人物ID、人物评价、出生日期、卷次ID、历史事件、历史活动地、名字、号名字、居住地、性别、曲腔派系ID、死亡日期、籍贯、著作ID、著作名称、著作年份、著作评价、谱序ID。

(6) 本体表示。晚明曲家年谱本体OWL文件部分内容表示如下。

①类的定义

```
<owl:Class rdf:about="http://www.chronologyLateMingDynastydramatiston.com#曲腔派系"/>
```

②对象属性的定义

```
<owl:ObjectProperty rdf:about="http://www.chro
```

```
nologyLateMingDynastydramatiston.com#有派系">
<rdfs:subPropertyOf rdf:resource="http://www.w3.org/2002/07/owl#topObjectProperty"/>
```

```
<rdf:type rdf:resource="http://www.w3.org/2002/07/owl#TransitiveProperty"/>
```

```
<rdfs:domain rdf:resource="http://www.chronologyLateMingDynastydramatiston.com#卷次"/>
```

```
<rdfs:range rdf:resource="http://www.chronologyLateMingDynastydramatiston.com#曲腔派系"/>
```

```
</owl:ObjectProperty>
```

③数据属性的定义

```
<owl:DatatypeProperty rdf:about="http://www.chronologyLateMingDynastydramatiston.com#名">
```

```
<rdfs:subPropertyOf rdf:resource="http://www.chronologyLateMingDynastydramatiston.com#名字"/>
```

```
</owl:DatatypeProperty>
```

(7) 创建实例。在Protégé中填充实例以便判断类与类之间的关系是否明确,本体结构是否符合应用需求。图2是Protégé中的一位戏曲家的年谱本体实例化的成功运用,说明整个本体模型能够使用,且可以准确地表达晚明曲家年谱中主要术语、对象属性和数据属性之间的组织关系。

(8) 本体构建。上述步骤完成之后,最终构建完成晚明曲家年谱本体模型(见图3)。

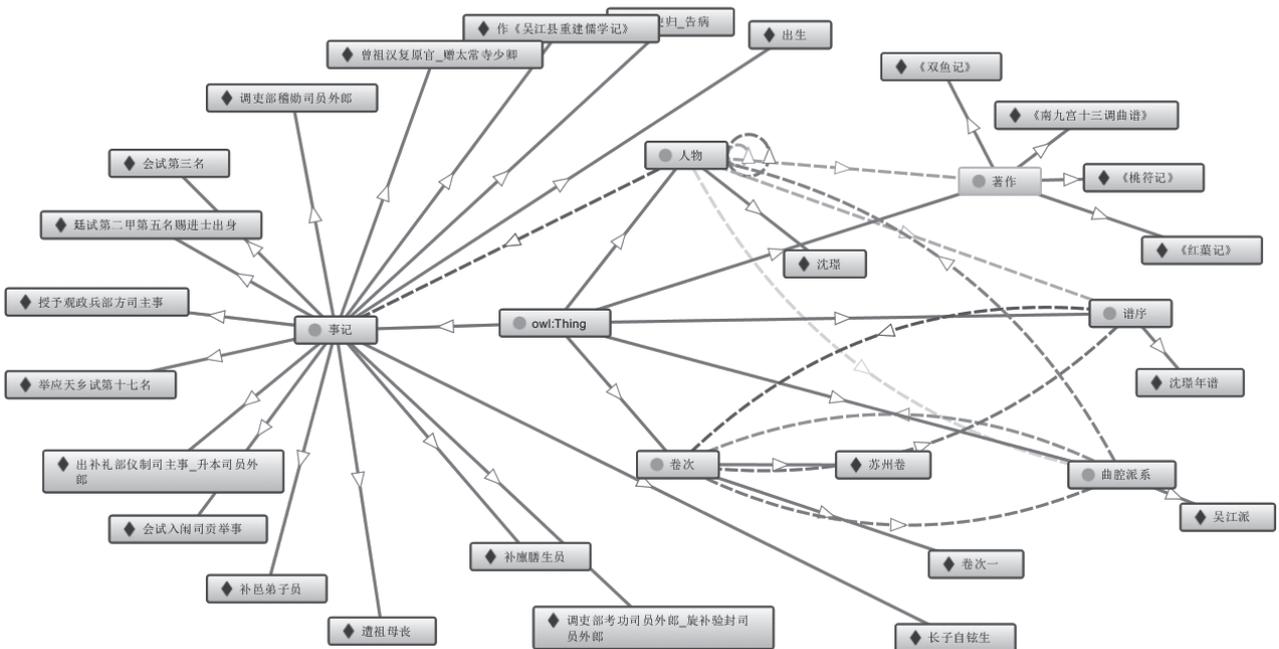


图2 晚明曲家年谱本体沈璟实例化示意

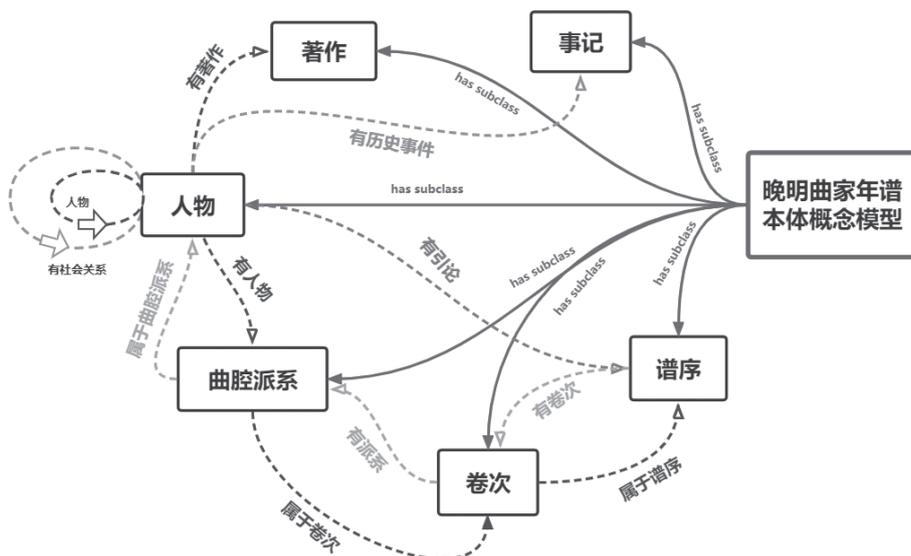


图3 晚明曲家年谱本体概念模型

4 面向晚明曲家年谱的知识图谱构建流程

本文利用图数据库Neo4j来完成CSV数据的存储和可视化呈现,构建晚明曲家年谱的知识图谱,基于晚明曲家年谱本体模型中节点关系之间的设定,对节点的CSV文件进行整合,确保每个节点有唯一的ID值及清晰的关系属性,利用Neo4j-admin import工具将数据导入图数据库Neo4j,最终实现晚明曲家年谱知识图谱的可视化。具体流程包括实体及关系抽取、节点CSV文件构建、数据导入以及可视化4个步骤。

4.1 晚明曲家年谱资源实体及关系抽取

利用ORC扫描工具将晚明曲家年谱纸质文本转换为数字文本,总结出以年谱谱主为核心实体单元,并以其为切入点,再以年谱文本内容的原始框架及概念模型,选择围绕人物相关的知识单元作为晚明曲家年谱本体模型中的实体,保障实例数据汇入的准确性,实现本体的实例化及相关可视化。

由于笔者在自然语言处理及其实体模型训练方面的能力有限,本研究的实体抽取选择传统的人工抽取方式,抽取完成后按照本体知识构建的需要进行CSV文件的保存。一是实体,核心实体以晚明曲家年谱文体结构为准,整个谱主论述中包含人物、卷次、谱序、事记(历史事件)、著作、社会关系六大类内容。由于晚明曲

家年谱是以谱主为核心的人物类传记资料,所以围绕六大类主体,并着重以人物这一概念实体为准,人工抽取符合年谱文本结构的相关实体数据,并存储为CSV文件格式。二是构建实体之间的关系,即实体关系数据。确定实体数据之后,结合其他历史资料,分析人物本身具有的社会属性与社会身份,如生平经历、交友关系、社会评价等。再参照复用本体的概念实体关系,复用或依据年谱资源特点构建整个实体的关系网络,形成一个以人物为中心的关系网络,即以人物实体为映射的实体关系数据,最终融合概念本体模型,设计出知识图谱所需要本体模型框架。

4.2 节点CSV文件构建

使用Neo4j-admin import工具存储知识图谱,需要建立节点知识图谱的节点CSV文件,一是进行数据文件与头文件(header file)分开建立,二是头文件与数据文件合并建立,其中每个节点都有唯一的ID值,以便进行关系分类时能够直接引用。依据本体模型的六大核心类别分别建立CSV文件,将数据文件与头文件合并建立并存储。

设置的节点文件包括6个。①卷次节点文件:juan_ci_nodes.csv。②谱序节点文件:pu_xu_nodes.csv;③曲腔派系节点文件:qu_qiang_nodes.csv。④人物节点文件:ren_wu_nodes.csv。⑤事记节点文件:shi_ji_nodes.csv;⑥著作节点文件:zhu_zuo_nodes.csv。

数字人文技术和研究体系,将年谱文本作为一种信息资源,并以年谱知识图谱模型研究为切入点,融入数字人文研究对象范畴,实现从技术的角度突破传统年谱资源的研究,让年谱知识图谱研究向数字人文领域靠拢,以达到理论与技术上的借力;并依据自身的资源特色,从资源内化出发,激发创新的年谱资源知识检索与挖掘等智能系统的研究,不断深化和丰富年谱知识图谱的内在资源,打造一个具有领域代表性的数字年谱知识图谱检索系统,真正将年谱资源转换成大众所需要的知识,让更多的用户能够参与到年谱的知识“活化”中。

5.2 产学研融合打造年谱知识图谱数字空间

数字人文的内涵之一便是人文与科技的深度融合,学界在数字人文的跨领域研究方面已经有了相对成熟的技术体系,年谱作为特殊的历史资源,本质上也是一种文化资源,但是面对数字经济新业态背景下,传统的年谱资源已经不能满足大众的数字文化需求,大众更倾向于有个性化、互动性及可视化的数据需求,而年谱知识图谱所具有的交互性、知识组织性恰好能够满足大众的个性化需求,从共享交互的角度为用户提供年谱数字资源平台。此外,提高企业、高校及科研机构对年谱资源化、数字化的认可,强调年谱资源的学术价值、历史和社会价值,助推年谱数字空间的开发,从经济、技术、人文的角度丰富年谱知识图谱的内涵。年谱知识图谱数字空间的构建,是年谱资源从传统文本资源转变为数字资源的重要升级,也是年谱资源未来走向深度智能化的关键一步。

5.3 基于年谱知识图谱和深度学习技术开发年谱数字产品

目前数字人文已逐渐趋向于研究对象的知识完备性,年谱文本资源的特殊性决定了自身资源的原始完备性,如何将资源的完备性通过数字技术转化为能够让计算机解读的数据对象,是运用数字人文技术研究年谱资源所面临的挑战。年谱知识图谱能够在数字化的过程中,最大化地保存原始数据的完备性,无疑破除了数据超大规模分析中数据原始完备性的损耗。此外,深度学习技术在数字人文领域已有较多的成熟应用,如循环神经网络诗歌数据分析、卷积神经网络古籍量化分

析以及文本词句解剖等,在技术层面上能够最大限度地融合到历史文化资源的研究中。因此,基于年谱知识图谱和深度学习技术,解决年谱文本资源的原始完备性、语义关系等相关问题,研发年谱知识图谱的衍生数字产品,使得年谱资源更趋向于数字智能化,调动年谱资源的转型升级。

5.4 以数字人文理念研发年谱知识图谱智慧数据库

智慧技术赋能年谱本文资源的数字转型,年谱知识图谱的构建一方面是对传统年谱资源的新尝试,从技术层面将年谱资源拉到知识重构研究对象领域,是打开年谱资源走向智慧化的技术基础;另一方面,年谱知识图谱的构建本质上是依靠数据库进行的数据分析,离不开用户需求。因此,在年谱资源复杂多样的情况下,对年谱知识图谱构建和深化要借助数字人文理念开展相关研究,用可视化、智慧化的手段去呈现年谱名人群体的精神文化;以群体需求为主,搭建年谱知识图谱智慧数据库,实现年谱资源的深度挖掘和深度分析,这也是数字人文领域研究的核心。因此,以年谱知识图谱为基础,扩充年谱知识图谱的技术路线,研制年谱知识图谱智慧数据库,是促进年谱资源走向数据人文领域的渠道,同时也从技术对象上丰富了数字人文的研究范畴,两者是互相成就的良性融合。

6 结语

本文以徐朔方《晚明曲家年谱》作为研究素材,通过数字人文研究范式、本体构建工具及知识图谱结合的方法,转变传统的纸质年谱资源共享性的立体数字资源,多维度地呈现年谱中人物、历史事件、社会关系等内容;探究本体技术与知识图谱技术在年谱资源中可视化研究的可能性,为我国传统年谱资源,如名人年谱、名人团体年谱、传记年谱及其他领域年谱资源的研究提供了技术框架的参考与借鉴,符合目前对于传统谱系文书资源转型研究的时代要求。但是仅以年谱知识图谱为基础实现多样化数字化是远远不够的,未来还需要对年谱知识图谱进行更深层次的研究,才能真正实现年谱资源的“活化、活用”。

参考文献

- [1] 王挺, 刘水云. 浙江省戏曲传承人口述史及数字档案创建的探讨[J]. 浙江传媒学院学报, 2012, 19(5): 96-100.
- [2] 方进. 《中国历代年谱总录》简介[J]. 文献, 1979(1): 320-323.
- [3] 简贵灯, 陈俊佳. 电影先驱顾肯夫影戏活动及创作年谱[J]. 电影文学, 2021(7): 60-64.
- [4] 闫海田. 当代文学历史化的有益尝试——论《东吴学术》“年谱研究”的意义[J]. 当代文坛, 2014(4): 130-133.
- [5] 陈鹏. 从名籍到谱牒: 中古中国的皇族档案[J]. 档案学通讯, 2018(1): 39-43.
- [6] 陈艳. 中国家谱的知识本体构建[D]. 上海: 上海交通大学, 2007.
- [7] 夏翠娟, 刘炜, 张磊, 等. 基于书目框架(BIBFRAME)的家谱本体设计[J]. 图书馆论坛, 2014(11): 5-19.
- [8] 张政平, 倪建成. 基于本体的孔子世家谱知识图谱[J]. 通信技术, 2021, 54(6): 1370-1377.
- [9] 祝振媛. 我国传统家谱档案的可视化构建——以浙江仙居高迁《吴氏西宅宗谱》为例[J]. 档案学通讯, 2020(2): 72-79.
- [10] 姜赢, 张婧, 朱玲萱. 基于本体的家谱知识图谱模型及检索系统[J]. 电子设计工程, 2017, 25(12): 161-165.
- [11] 吴信东, 盛绍静, 蒋婷婷, 等. 从知识图谱到数据中台: 华谱系统[J]. 自动化学报, 2020, 46(10): 2045-2059.
- [12] 张吉祥, 张祥森, 武长旭, 等. 知识图谱构建技术综述[J]. 计算机工程, 2022, 48(3): 23-37.
- [13] 田玲, 张谨川, 张晋豪, 等. 知识图谱综述——表示、构建、推理与知识超图理论[J]. 计算机应用, 2021, 41(8): 2161-2186.
- [14] 陈雅茜, 邢雪枫. 基于本体建模的动态知识图谱构建技术研究[J]. 西南民族大学学报(自然科学版), 2021, 47(3): 310-316.
- [15] 姚萍, 李坤伟, 张一帆. 知识图谱构建技术综述[J]. 信息系统工程, 2020(5): 121-123.
- [16] 陆林. 《晚明曲家年谱》金圣叹史实研究献疑[J]. 文学遗产, 2002(1): 68-78.
- [17] 董行. 面向跨语言家谱服务的多源关联数据匹配研究——上海图书馆开放数据应用比赛作品Learn Chinese Surnames[J]. 大学图书馆学报, 2018, 36(4): 50-57.
- [18] 张捷. 论中华寻根网的资源构架[J]. 图书情报导刊, 2021, 1(6): 132-134.
- [19] ZEEBAREE S R M, AL-ZEBARI A A, JACKSI K, et al. Designing an ontology of e-learning system for Duhok Polytechnic University using Protégé OWL tool[J]. Journal of Advanced Research in Dynamical and Control Systems, 2019, 11(5): 5.
- [20] 段宇锋, 黄思思. 本体构建方法研究[J]. 情报杂志, 2015, 34(11): 139-144.
- [21] 傅柱, 王日芬, 孙铭丽. 本体存储技术研究综述[J]. 情报理论与实践, 2013, 36(9): 118-123.
- [22] 王余蓝. 图数据库Neo4j与关系数据库的比较研究[J]. 现代电子技术, 2012(20): 77-79.
- [23] 李雪. 一种基于Neo4j图数据库的模糊查询研究与实现[J]. 计算机技术与发展, 2018, 28(11): 16-21.
- [24] 刘爽, 杨辉, 李佳宜, 等. 面向多数据源的少数民族文化知识图谱构建[J]. 计算机技术与发展, 2021, 31(8): 191-197, 203.
- [25] 王仁武, 袁毅, 袁旭萍. 基于深度学习与图数据库构建中文商业知识图谱的探索研究[J]. 图书与情报, 2016(1): 110-117.

作者简介

徐彤阳, 男, 1976年生, 博士, 副教授, 研究方向: 数字人文、信息检索, E-mail: tyxu2004@163.com。
黄映思, 女, 1998年生, 硕士研究生, 研究方向: 数字人文。

Ontology-based Knowledge Graph of Celebrity Chronicle Resources:
Taking the Xu Shuofang's *Chronicles of Dramaturgists in the Late Ming Dynasty* as an Example

XU TongYang HUANG YingSi
(College of Information, Shanxi University of Finance and Economics, Taiyuan 030006, P. R. China)

Abstract: By combining the research paradigm of digital humanities with the tools of knowledge mapping, it enables the deep excavation and presentation of the culture of celebrity groups, transforming traditional celebrity chronicle resources into a shared three-dimensional digital resource. Using Xu Shuofang's *Chronicles of Dramaturgists in the Late Ming Dynasty* as the research material, the ontology of the late Ming Qu family chronicle is constructed with the help of Protégé tools, and the construction of the knowledge map of the chronicle is completed by combining the Neo4j graph database, in order to reveal the digital research value and digital reuse of the current traditional chronicle resources, and to put forward the outlook of the future traditional celebrity chronicle resources towards digital and intelligent in-depth research.

Keywords: Digital Humanities; Chronicle; Ontology; Dramaturgists; Knowledge Graph

(收稿日期: 2022-10-13)