

馆藏家谱文献知识化开发：价值、主体与过程^{*}

朱兰兰 霍婕 高玉婷
(郑州航空工业管理学院, 郑州 450046)

摘要: 家谱作为记载家族世系繁衍和重要事迹的珍贵历史文献,不仅可以满足人们日益增长的家谱文化需求,还为传承和弘扬中华优秀传统文化提供可行路径。基于馆藏家谱文献资源开发问题,构建家谱文献知识化开发的“价值-主体-过程”三维分析框架,以河南省档案馆馆藏明代正统本《河南如堂黄氏宗谱》为例,重点探讨家谱文献知识化开发过程中的数据获取、本体构建、知识抽取、知识融合和资源可视化等问题,以为家谱文献资源的深度开发提供新模式。

关键词: 家谱文献; 知识化开发; 知识图谱

中图分类号: G251 **DOI:** 10.3772/j.issn.1673-2286.2023.11.005

引文格式: 朱兰兰, 霍婕, 高玉婷. 馆藏家谱文献知识化开发: 价值、主体与过程[J]. 数字图书馆论坛, 2023 (11): 38-45.

家谱又称族谱、宗谱,是一种以表谱形式,记载一个家族世系繁衍及重要人物和事迹的珍贵历史文献,极具平民特色,对于深入研究历史学、民俗学等学科和保护中国传统家族文化均有不可替代的功能。近年来,随着“寻根热”的发展以及国家重点档案抢救与保护项目的实施,我国家谱文献的开发利用工作渐有成效,例如河南省档案馆开展了对家谱文献的抢救性保护工作和家谱文献数字化工作,开发了家谱档案管理系统平台,一定程度上集成了省内各级综合档案馆的数字家谱档案,家谱成为河南省数字档案馆的特色资源。目前在家谱文献资源开发领域相关研究主要集中在家谱价值研究^[1]、少数民族特色家谱研究^[2]、家谱目录整理^[3]、家谱资料汇编^[4]、家谱体例研究^[5],以及家谱收藏研究^[6]等方面。家谱资源知识化开发研究则更侧重于家谱本体的构建,如夏翠娟等^[7]设计的基于书目框架的上海图书馆家谱本体和陈艳^[8]构建的中国家谱知识本体等,但馆藏家谱文献智慧数据资源建设以及知识

化开发方面尚未形成操作性较强的理论方法体系,集全流程、多角度、宽视野于一体的家谱资源知识化开发研究仍较为缺乏。故而笔者在研究现有馆藏家谱文献资源开发情况基础上,尝试构建家谱文献资源知识化开发的理论模型并探索运行机理,以为家谱档案资源的开发提供新思路。

1 馆藏家谱文献知识化开发的提出

1.1 现存问题分析

(1) 价值失衡,资源效益发挥程度较低。近年来,信息技术的应用促使家谱文献资源建设取得了较大进展,但家谱文献资源的开发价值取向不明晰,与用户的多元需求不匹配,难以满足传承与发展家谱文化的需要。主要原因有三。①信息缺失。由于我国家谱文献年代久远、保存环境不当,家族信息的缺失已是常态,但

收稿日期: 2023-08-10

^{*}本研究得到2022年河南兴文化工程文化研究专项项目“河南家谱研究”(编号: 2022XWH047)、“黄河文化遗产数据资源的语义表示与组织研究”(编号: 20220214B)资助。

我国目前的家谱文献资源开发工作将重点放在了家谱数字化和新技术应用领域, 忽视了传统的修谱缮谱工作, 造成了家谱不完整、家族历史无法追溯的情况^[9], 无法满足用户的溯源需求。②易用性不足。我国馆藏家谱文献多为古籍, 多数家谱资源网站往往只提供数字化的家谱文本或简单的谱系图, 例如以“家谱”为关键词在我国各省数字图书馆中检索可以发现: 只有上海图书馆设立了专门的“特色文献-家谱书目”入口, 该系统具备姓氏浏览、时空查询、家谱文化查询、家规家训学习和在线修谱等功能^[10]; 在首都图书馆网站以“家谱+民国图书”和“家谱+古籍特藏”为检索词共得到84条检索结果^[11]; 其余省级图书馆家谱文献保存量极少且检索结果均以出版图书为主, 缺乏易于使用 and 理解的交互式功能, 无法为用户提供精准化的家谱信息服务。③特色资源开发滞后。中国是一个团结统一的多民族国家, 拥有许多独特珍贵的少数民族家谱资源, 但我国对少数民族家谱等特色资源的发掘明显不足。例如, 云南省作为我国少数民族最多的省份, 虽对少数民族家谱有所收录, 但少数民族特色家谱文献资源建设与开发较为滞后, 尤其是开发手段偏老化, 而特色家谱资源是历史的遗存, 文献中很多表达方式已经不被大众熟知, 这就造成了用户和资源之间的理解壁垒, 使其无法精准服务用户^[12]。

(2) 结构松散, 机构工作协同能力不足。馆藏家谱文献资源的开发是一项跨学科的工作, 需要社会各界的广泛参与。在家谱文献资源的知识化开发中, 技术公司、馆藏机构和高校等科研部门互相支撑, 发挥着重要的作用。但纵观各省级馆藏机构的家谱文献资源开发情况, 我国馆藏家谱文献资源的开发存在进展不平衡以及技术割裂的问题。究其原因, 一方面是机构之间合作不够紧密, 高校等研究机构注重深挖文献背后的含义, 但缺少相应的技术手段, 技术公司缺乏与家谱相关的专业研究人员, 这限制了家谱开发的深度和广度; 另一方面是各参与主体过多专注自身领域的研究, 没有应用交叉学科的优势, 机构协同性不足, 致使各机构间缺乏信息资源共享和交流沟通, 浪费资源和时间。总之, 高校、技术公司和馆藏机构等需要加强合作, 充分发挥学科交叉优势, 增强协同性, 创造条件以推进家谱文献资源的知识化开发。

(3) 过程泛化, 资源开发粒度不细。数字化时代的到来和技术的进步使纸质文献数字化成为必然趋势, 数字资源库的出现打破了利用文献资源的时间和空间

壁垒。在家谱文献的数字化处理中, 许多机构只是将馆藏家谱扫描成数字图像, 却没有对文献中的内容进行全面的数字化处理。这样, 即使文献完成数字化, 资源的查找利用也需要人工辅助, 这大大降低了检索的效率。当前对馆藏家谱文献进行编研且建立了独立家谱数据库的收藏机构不多, 如上海图书馆和河南省档案馆等, 大多数收藏机构所开展的家谱文献数字化工作仅是对现有家谱文献进行扫描、整理、上传, 而包括虚拟家族树、家谱文献资源检索系统等在内的相关数字资源建设处于初始阶段。因此, 我国家谱文献资源知识化开发工作亟待提上日程。

1.2 基于“价值-主体-过程”的三维分析

由上述分析可知, 家谱文献资源的知识化开发存在着价值认识不足、主体意识欠缺和过程不完善的障碍, 这制约着整个体系的开发效果。因此, 在知识组织理论指导下, 提出基于“价值-主体-过程”的家谱文献知识化开发三维理论分析框架, 系统地分析知识图谱赋能家谱文献资源开发的内核与流程。

价值维度需结合家谱文献利用的情境, 研究馆藏家谱文献知识化开发的多元价值及其关联, 本文主要从基于需求的价值、基于技术的价值和基于文化的价值三方面展开讨论; 主体维度重在研究以馆藏机构为主导、高校等科研部门协同、社会机构参与的家谱文献知识化开发的主体构成, 开发主体之间的权责关系、运行机理等; 过程维度通过对非结构化家谱文献本体的语义解析来获取家谱文献知识要素, 并与已有的结构化家谱文献数据相融合, 研究家谱文献知识化开发流程、构建家谱文献的知识图谱以及家谱知识可视化展示的技术流程。

不同社会背景对应不同的隐藏信息, 家谱文献资源的知识化开发具有较强的背景性, 需要结合特定的时代环境和背景。所以在价值、主体、过程之外, 加入“社会背景”这一空间要素^[13], 构成家谱文献资源知识化开发的三维分析框架(见图1)。

2 馆藏家谱文献知识化开发的价值取向

2.1 满足用户的家谱文化需求

通常情况下, 家谱文献资源的用户可以分为专业

机构和社会大众两类。首先,从专业机构角度出发,家谱文献的知识化开发实际上是对家谱资源内容的深度挖掘和关联,为专业人员的调研学习工作夯实基础,扩大研究范围。其次,对个人而言,家谱文献的价值主要在于“寻根溯源”和“问祖归宗”。通过家谱文献的知识化开发能够不断更新家谱资源库,完善家谱文献服务系统,让社会大众能更为便利、直接地找到自己的祖先,满足用户的家谱文化需求,如美国国家档案与文件署(National Archives and Records Administration, NARA)网站不仅为用户提供文字记录的家谱信息,还能为用户提供3D虚拟族谱视频等。

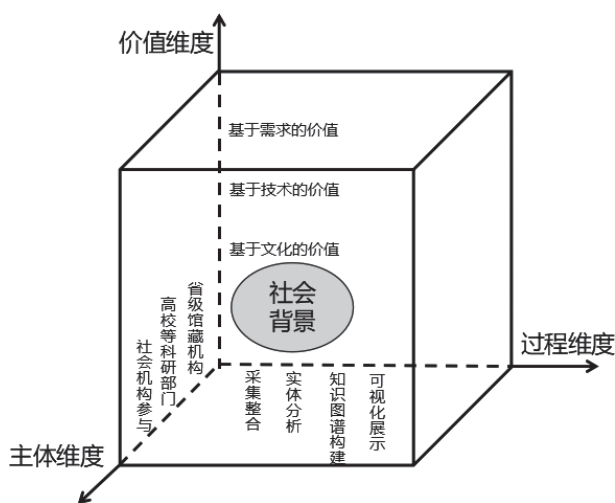


图1 家谱文献资源知识化开发的三维分析框架

2.2 强化智慧技术的创新应用

家谱文献资源知识化开发过程需应用多种技术手段:在信息采集和资源建设方面,采用征集、拍摄、3D打印、建立信息编码规则等管理和技术手段;在信息处理和存储上,采用文字和图像识别、图像处理、3D重建、数据信息智能关联、数据库建设等技术;在信息开发与服务上则综合运用虚拟现实、增强现实、基于专业词汇或自然语言的高效检索以及新媒体推广、App开发等技术。整个过程中多种技术相互交织,加快加深了衔接和渗透,强化了智慧技术的创新应用。

2.3 传承与弘扬优秀传统文化

家谱不仅是记录世系用以传世的工具,其中蕴含的文化价值还会深刻影响后代。家谱中记录有姓氏传承、始祖源流、优秀家训、史籍典故、精美刻印等内容。家族

的家谱反映的就是一个时代的习俗和一个民族的历史。因此,通过对家谱文献资源进行知识化开发一方面可以挖掘家谱中的优秀家训家规,并用其弥补现代教育体系中品德教育的缺失,助力传承中华优秀家风文化的精髓;另一方面,家谱中承载的历史典故、呈现的人物画像和印章具有很高历史文化价值和美学价值^[14],充盈了中华文化的内涵,为中华文化的传播奠定了坚实的基础。

3 馆藏家谱文献知识化开发的主体分析

3.1 家谱文献资源知识化开发的主体要素

家谱文献资源的知识化过程需要社会各方力量的广泛参与,它不仅需要历史工作人员对知识的挖掘,也需要科研部门和技术公司的加持和配合。家谱文献资源知识化开发工作应当顺应数字化发展潮流,利用数字技术共享数字化成果,打破各档案部门、高校科研部门和技术公司之间空间和时间上的壁垒,实现通力合作。数字人文背景下的家谱文献资源知识化开发主体要素呈现多元化趋势^[15],各主体可以实现线上随时交流,知识发现和图谱构建工作同步进行,可大大提高家谱文献资源知识化开发的效率,开拓家谱文献资源合作开发的新渠道。

3.2 开发主体合作关系分析

作为一个复杂开放的体系,家谱文献资源知识化开发合作体系自身具有特定结构,在加入知识开发技术等要素后转变为面向公众的服务体系。从合作结构层面出发,家谱文献资源的流动、家谱知识的抽取和家谱知识可视化平台的构建连接了家谱文献资源知识化开发中的主体合作关系(见图2)。

省级馆藏机构在家谱文献资源知识化开发过程中处于主导地位,既是家谱档案资源的主要提供者,也是高校和社会机构的协调者。在纵向结构上,各省级馆藏单位应积极协调和组织各级机构,发挥跨地域和跨机构的资源组织优势,实现从中央到地方、从上级至下级的统一行动。在横向结构上,其又需要与各社会机构和技术公司建立长效的合作关系。

高校等科研部门是家谱文献知识化开发的核心力量,也是主要的技术构建方。他们可以针对复杂的知识图谱需求,使用本体定义模块,与馆藏机构通过交互式

配置完成概念层级以及实体和关系属性的继承定义,并且可以进行后续的可视化展示,使得知识结构更为清晰明确。此外,科研部门也可以利用知识抽取模型,实现事件、实体、关系、属性的自动化抽取,并以“概念-实体-关系”图谱形式进行知识推理和相关结果呈现。

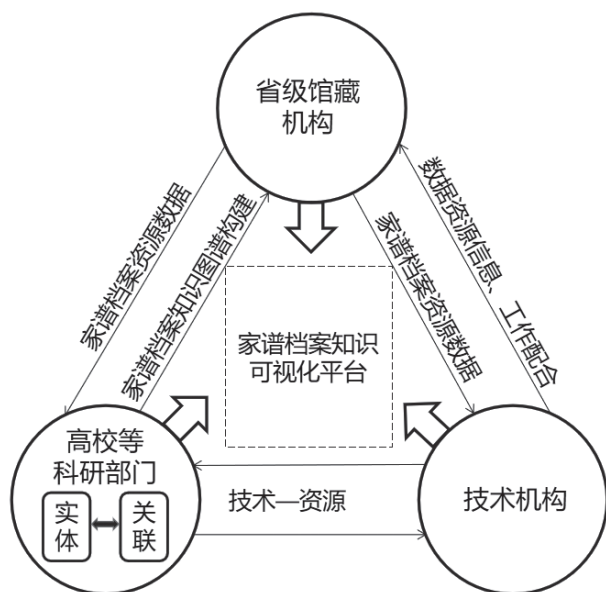


图2 家谱文献资源知识化开发主体合作关系分析

社会机构是家谱文献知识化开发过程中不可缺少的后备力量。无论是开展家谱资源的数字化和内容消歧,还是协助高校科研部门进行实体抽取和关系核验,社会机构都发挥了承上启下、配合调动的全能辅助作用,是家谱文献资源知识化开发的强劲后援。

4 馆藏家谱文献知识化开发的过程——以《河南如堂黄氏宗谱》为例

基于知识图谱技术的家谱文献资源开发可以实现家谱文献从实体管理到内容管理的转变。以河南省档案馆馆藏明代正统本《河南如堂黄氏宗谱》为例,从家谱文献资源数据获取、本体构建、知识抽取、知识融合和知识图谱可视化5个部分展示馆藏家谱文献知识化开发的过程。

4.1 家谱文献资源数据获取

家谱资源数据获取的第一步就是征集家谱资源并进行数字化。在数据采集的具体过程中,应该遵循现有资源的数据化、文本化和网页数据同步获取、多源融合

的原则。一方面,针对馆藏家谱资源进行数据获取。依托河南省档案馆家谱档案管理平台的建设,《河南如堂黄氏宗谱》已取得初步的数字化成果,但仅停留在扫描图像的展示层面,单独的文本抽取工作还未开展^[16]。因此,需要将家谱全文文本化,进行文本的校对及预处理,为后续的三元组关系抽取和基于深度学习的图谱构建提供基础。为补充获取家谱文献资源数据,具体可采取以下几种举措:采用“众包”的方式,向社会大众开放平台,获取开源的《河南如堂黄氏宗谱》新知识,尝试补全家谱残缺内容;利用应用程序接口、中间数据库等方式获取相关家谱文献数字平台的数据;借助网络爬虫获取网络动态家谱信息,数据录入和获取同时进行,实现对资源数据的全面采集。

《河南如堂黄氏宗谱》流传年代久远,故而在数据采集的过程中可能会遇到以下问题:首先是由于纸质家谱流传时间久远,家谱本身可能存在信息的缺失或者繁体简体混用的问题;其次是在文献识别过程中某些字符被错误识别成异体字;最后是在人工校对过程中,因校对人、责任人理解偏差而产生标点位置错误。针对这些问题,一方面可以采取人工阅读的方法及时调整更正;另一方面可以在校对时使用Python软件编写文本处理代码,进行基于规则的抽取,在抽取结果中发现不符合实际的情况,再进行反馈校对和数据清洗。

4.2 家谱文献资源本体构建

本体是对家谱文献资源的共同理解,是对家谱中核心概念的归纳,也是知识图谱模式层构建的基础,为家谱领域内知识的复用和推理提供了可能。

七步法和骨架法是构建领域本体的常用方法。骨架法分五步进行,虽建立了完整的反馈机制,但缺乏对现有本体复用的思想。七步法虽考虑了对现有本体的复用,但是没有对本体进行有效性检测。借鉴相关研究^[17],综合七步法和骨架法的核心内容,采用兼顾本体复用和评价机制的本体构建方法。

首先,明确本体应用范围为《河南如堂黄氏宗谱》,其中包含的知识种类繁多,内容兼具广度和深度。①家谱表现形式多样:《河南如堂黄氏宗谱》中除了单一的家族谱系文字记录外,还有“冬字一支”“月字一支”“衮字一支”的世系图、御史中丞赠银青光禄大夫左散骑常侍遗像等,图文结合,表意鲜明。②家谱内涵丰富:《河南如堂黄氏宗谱》记录了黄氏的优良家

风家训和名人事迹,例如“不忘本”“敬祖宗”“堂正诚实”“廉洁清贫”“孝敬父母”“善待他人”等家训和“黄香温席”的典故。

其次,现有家谱领域本体主要有两大类,即面向文献的本体和面向人物关系的本体。面向家谱文献的本体主要包含资源、世系、文献、地区、民族、版本及载体。面向人物关系的本体更加注重人物的基本信息及其关系,例如字、号和配偶、父子等。上海图书馆家谱本体同时考虑上述内容,复用了书目框架的部分类和属性,形成了上海图书馆家谱本体模型,本研究借鉴了这一方法。

再次,基于《河南如堂黄氏宗谱》内容,将基本实体定义为人物、时间、地点、事件4个部分。其中的人物的属性包括名、字、姓名、性别、排行、职位。《河南如堂黄

氏宗谱》中心人物以男性为主,同篇会涉及中心人物的父亲、母亲、妻子、儿子、孙子,所以人物对应的关系为夫妻、父亲、儿子、孙子、曾孙5种。时间主要指某一特定时间点,其属性为时代。人物和时间之间的关系包括出生时间、死亡时间、乔迁时间。地点的属性有名称、相对位置和环境。人物和地点之间的对应关系有出生地点、死亡地点、居住地点、乔迁地点。事件是指中心人物所受的奖惩和中心人物引发的有记录意义的事件,如黄香温席、黄霸判案。主要采用Protégé工具来构建《河南如堂黄氏宗谱》本体中的类、关系和属性(见图3)。

最后,将与黄香有关的事件、人物、地点、时间进行编辑输入,得到基于中心人物黄香的本体实例,如图4所示。

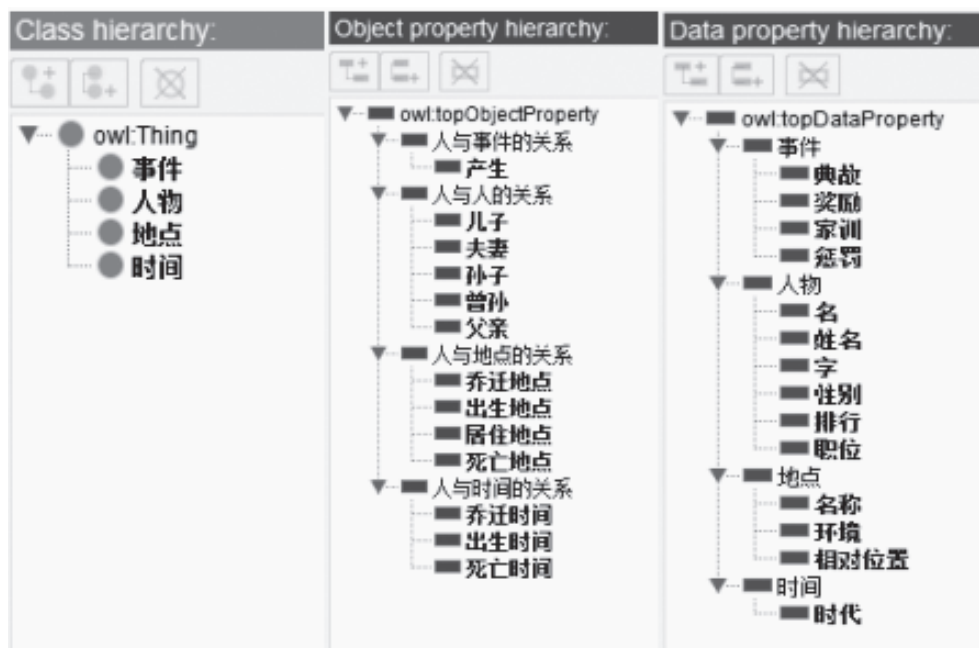


图3 《河南如堂黄氏宗谱》本体中的类、关系、属性

4.3 家谱文献资源知识抽取

知识抽取指的是从收集到的家谱文献资源中抽出实体、关系、属性数据,是家谱文献知识图谱构建的基础。《河南如堂黄氏宗谱》中包含世系图、文本、图画和印章等,具有多模态特性。因此,应根据已建立的连续词嵌入空间向量的办法,结合对抗训练、强化学习和深度学习等方法进行实体、关系、事件的抽取。

《河南如堂黄氏宗谱》的知识抽取重点在于有效地实现对古汉语知识的抽取。考虑到家谱文献知识密度大的特点,借鉴使用了序列标注法,该方法在运行过

程中利用二元标签组的优势,可同步开展实体抽取和关系抽取^[18],提高了知识抽取的速度和准确度。

序列标注法具体步骤如下。首先将文本中“字”“号”“生于”等无用字符剔除,然后将一段字符用二元标签组(a, b)标记。 $a \in \{B, I, E, S\}$,代表了该字符在字符串中的相对位置,其中: B代表beginning,指一个字符串的开端; I代表inside,通常表示一个字符串的中间部分; E代表end,指字符串的结束; S为single,表示单独的字符。b为之前定义过的任意实体类型。举例来说,标记“黄香,字文强,居江夏安陆”这段字符,首先剔除无用字符“字”和“居”,则文

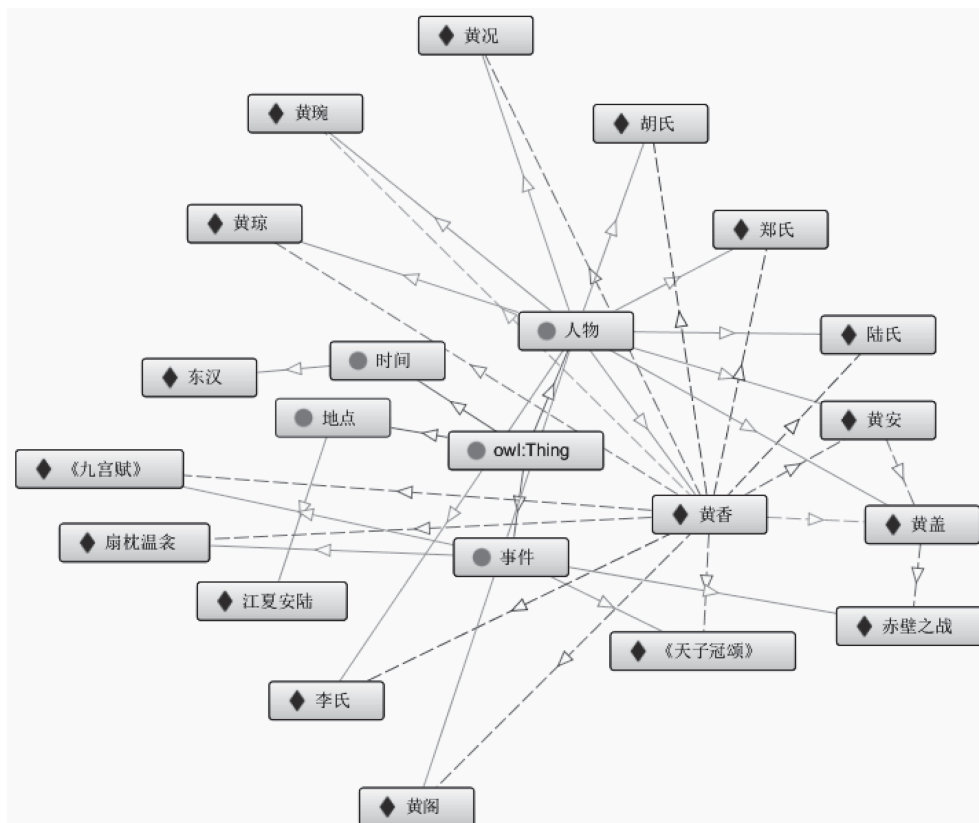


图4 基于中心人物黄香的本体实例

本内容变为“黄香,文强,江夏安陆”,接着将文本序列标记为“(B,姓名),(E,姓名),(B,字),(E,字),(B,居住地),(I,居住地),(I,居住地),(E,居住地)”。在该示例中可以抽取实体为“黄香”“文强”“江夏安陆”,进而可以抽取关系“中心人物-字:黄香-文强”“中心人物-居住地:黄香-江夏安陆”。

此外,对于《河南如堂黄氏宗谱》中的图像数据,可以对其进行格式转换和标引,将其链接到知识图谱的相应位置。

4.4 家谱文献资源知识融合

在家谱文献资源知识融合的过程中,数据来源仅是《河南如堂黄氏宗谱》及现代黄姓人员的公开信息,故而不涉及模式层融合,只进行数据融合即可。

在家谱文献资源知识融合的过程中会出现很多数据层面的问题。首先,鉴于家谱文献的自身特点,家族中难免会存在同名同姓的现象。其次,家谱文献年代久远,其字体和现代汉字也稍有出入,可能会出现识别不出的问题。最后,数字化技术并不完全可靠,在知识抽取过程中不可避免地会产生一些错误。因此,实体消歧

和冲突检验是十分必要的两项工作。

实体消歧就是对同一名字的不同个体按数据类型进行消歧,消除实体指向不明的问题。在综合考虑古人命名方法和家谱文献特点的基础上,可以采用字辈标注法,必要时可以引入其父亲的字辈加以参考,通过为每个实体加入标识符来区分不同实体,实现实体消歧。

冲突检验的主要作用是检查是否存在内部知识冲突。知识图谱功能的强大之处就在于可以进行知识的推理,但是完全依赖机器自主工作,推理的准确性不能得到保障。因此,冲突检验主要从检验实体的合理性、检验实体间关系是否与事实不符、检验推理部分是否与原有事实相悖三方面进行,从而保证推理结果的正确性。

4.5 家谱文献资源的可视化

家谱文献资源的可视化有两层含义:一是《河南如堂黄氏宗谱》知识图谱的可视化,二是家谱知识的可视化。可视化过程兼顾内部逻辑理论构建与外部知识服务平台搭建,实现了从理论到实践的全程可视化。

(1) 知识图谱可视化。在三元组构建的基础上将家谱数据注入本体模型,运用rdf2rdf软件实现OWL文

件向RDF文件的转换并保存分析得到的家谱文本及对应的图像索引,选择合适的可视化方式生成《河南如堂黄氏宗谱》文献知识网。常见的家谱知识图谱实现形式有树状结构图、关系网络图、时间轴图、地图图表等。针对黄氏家族成员之间的复杂关系,关系网络图是最适合的可视化实现形式。关系网络图可以将家族成员之间的关系以节点和连线的方式展示,每个节点代表一个家族成员,连线代表家族成员之间的关系。这种可视化形式可以更全面地展示家族成员之间的复杂关系,包括亲属关系、婚姻关系、血缘关系等。另外,关系网络图可以根据家族成员之间的关系密切程度进行布局,将关系密切的家族成员放在相邻的位置,从而更好地展示家族成员之间的关系。同时,关系网络图还可以通过调整节点和连线的颜色、粗细等属性来表达不同类型的关系和重要程度,使得信息更加丰富和清晰。当然,当数据量较大、关系复杂度较高时,关系网络图也可能存在展示效果不佳的问题,此时可以考虑使用其他可视化实现形式进行补充或者调整。这些可视化实现形式可以根据具体需求和数据特点进行选择 and 组合,以实现更好的可视化效果。

(2) 家谱可视化服务平台开发。需要建立家谱文献资源服务平台,将开发的家谱文献知识以不同形式上传到平台上,实现“资源-用户”的连接。

数据库是家谱可视化平台服务的内核,家谱可视化平台的数据库应该根据用户需求和数据特点进行设计和建设。一个功能集成的平台应该建设有以下三类数据库。①关系型数据库。关系型数据库是常见的数据库类型,它使用表格来存储数据,每个表格包含多个字段,每个字段代表一个数据项。家谱可视化平台可以使用关系型数据库来存储家族成员的信息,例如MySQL、PostgreSQL等。②图形数据库。图形数据库是用于存储图形数据的数据库,它可以更好地存储和管理具有复杂关系的数据,如家族成员之间的关系。可以使用Neo4j等图形数据库。③NoSQL数据库。NoSQL数据库是一种非关系型数据库,它可以存储半结构化或非结构化数据,具有高可扩展性和灵活性。家谱可视化平台可以使用NoSQL数据库来存储家族成员的信息,如MongoDB等。此外,在建设家谱服务平台的数据库时,还需要考虑家谱统计数据表的设计、家族数据模型的构建、数据库性能优化和数据库安全保障等方面。

平台的服务功能是连接资源与用户的枢纽,家谱可视化服务平台的功能应该根据用户需求和平台定位

进行设计。一般的家谱资源服务平台应该包括一些基础功能版块:家族成员管理版块,用户可以通过平台添加、编辑、删除家族成员的信息,包括姓名、出生日期、死亡日期、婚姻状况、子女信息等;家族关系管理版块,用户可以通过平台管理家族成员之间的关系,包括亲属关系、婚姻关系、血缘关系等,以便更好地展示家族成员之间的关系;可视化展示版块,平台可以提供多种可视化展示形式,如树状结构图、关系网络图、时间轴图、地图图表等,让用户可以更好地了解家族成员之间的关系和分布情况;数据分析版块,平台可以对家族成员的信息进行数据分析,如家族成员年龄分布、家族成员出生地分布、家族成员婚姻状况分布等,以便更深入地了解家族成员的特点和趋势;数据导入和导出版块,平台可以支持数据的导入和导出,方便用户将已有的家族成员信息导入平台,或将平台中的家族成员信息导出到其他平台或工具中进行使用;用户管理版块,平台可以设置用户管理功能,包括用户注册、登录、修改密码等,以及用户权限管理,以便管理员管理用户访问和操作的权限;数据安全和隐私保护版块,平台需要采取相应的措施来保障数据安全,包括数据加密、访问控制、备份和恢复等,以防止数据泄露和丢失。此外,功能创新是家谱资源服务平台的核心竞争力,各家谱服务平台可以通过增加“DNA寻亲”等模块,将先进技术与家族感情相连。一方面提供更可信的寻亲结果;另一方面增加用户的使用兴趣,吸引潜在用户,扩大用户群^[19]。

由于家谱数据具有私密性,家谱可视化服务平台需要采取多种措施来保障平台的数据安全和保护用户隐私。①数据加密,使用加密技术对数据进行加密,包括传输加密、存储加密、数据脱敏和数据匿名化等,以防止数据在传输和存储过程中被窃取;②访问控制,采用用户身份验证、权限控制等确保只有授权用户才能访问和操作平台;③网络防护,使用防火墙和安全组等技术来限制非法访问、阻止网络攻击,以保护平台;④数据备份和恢复,定期进行数据备份,并采取相应措施来保障数据的完整性和可恢复性,以防止数据丢失和损坏;⑤安全审计,进行安全审计和日志监控,以便及时发现和处理安全漏洞和异常情况。家谱可视化服务平台需要综合考虑平台的特点和风险,采取相应的措施来保障平台的安全和保护用户隐私。在平台开发和运营过程中,需要及时地进行安全评估并修复漏洞,以确保平台的安全和稳定性。

5 结语

家谱是寻亲问祖的重要依据,也是中华文化的重要载体,对理清家族脉络、加强家族凝聚力、了解家国历史和传承中华文化都具有重要的意义。各级馆藏机构应加强家谱文献资源的开发工作,深度发掘利用其潜在价值。当前我国家谱文献资源知识化开发程度总体较低。在明确家谱文献资源开发价值取向的基础上,家谱收藏者、研究者、社会机构和高校等多元主体应协同发挥主体优势,利用知识图谱技术推进家谱文献资源细粒度知识开发。本文仅以河南省档案馆馆藏《河南如堂黄氏宗谱》为例,探讨了家谱文献资源知识化开发过程中的数据收集、图谱构建和资源可视化等问题,提出了基本的方法,但是知识图谱具有较强的灵活性,在具体构建过程中需要依据不同年代文本的不同语义信息采取合适的构建方式。下一步将继续探寻家谱领域通用连续词嵌入空间向量的训练方法,力求使得研究成果更加可信、可用。

参考文献

- [1] 邓绍兴. 简谈家谱档案及其收集[J]. 历史档案, 1983 (4): 111-113.
- [2] 肖敏, 华林. 西南少数民族谱牒档案及其利用研究[J]. 山西档案, 2016 (6): 9-11.
- [3] 王鹤鸣. 中国家谱总目[M]. 上海: 上海古籍出版社, 2008.
- [4] 常建华. 中国族谱资料的整理、研究和数字化建设[J]. 安徽大学学报(哲学社会科学版), 2014, 38 (1): 95-105.
- [5] 吕春阳. 明代徽州家谱内容与体例研究[D]. 芜湖: 安徽师范大
- 学, 2016.
- [6] 徐建华. 家谱的地方性特色及价值[J]. 福建论坛(人文社会科学版), 2005 (9): 56-59.
- [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, 34 (11): 5-19.
- [8] 陈艳. 中国家谱的知识本体构建[D]. 上海: 上海交通大学, 2007: 36-54.
- [9] 王晖. 乡镇村志编纂存在的问题与对策[J]. 中国地方志, 2018 (2): 15-20, 124.
- [10] 上海图书馆. 上海图书馆[EB/OL]. [2023-05-06]. <https://jiapu.library.sh.cn>.
- [11] 首都图书馆. 首都图书馆[EB/OL]. [2023-05-06]. https://primo.clcn.net.cn/primo_library/libweb/action/search.do.
- [12] 夏妍. 档案馆馆藏家谱档案开发利用研究: 以我国省级综合档案馆为调研对象[D]. 上海: 上海师范大学, 2021.
- [13] 高淞, 王向女. 数字人文视域下口述历史档案资源开发利用研究[J]. 山西档案, 2021 (3): 61-70.
- [14] 李宝玲, 关京红. 黄氏宗谱的研究价值[J]. 档案管理, 2016 (3): 35-36.
- [15] 朱兰兰, 薄田雅. 数字人文视域下家谱档案资源多元化开发[J]. 浙江档案, 2019 (10): 31-33.
- [16] 河南省档案馆家谱档案管理系统. 家谱谱影[EB/OL]. [2023-05-06]. <http://222.143.34.175:8080/jpda>.
- [17] 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.
- [18] 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 (3): 582-600.
- [19] 姜丰. 乡村振兴话家谱一家谱数字化服务平台构建与视觉设计研究[D]. 武汉: 华中师范大学, 2021.

作者简介

朱兰兰, 女, 博士, 教授, 研究方向: 档案学理论与实践, E-mail: 1325056142@qq.com。

霍婕, 女, 硕士研究生, 研究方向: 档案学理论与实践。

高玉婷, 女, 硕士研究生, 研究方向: 档案学理论与实践。

Intellectual Development of Genealogical Literature Collection: Value, Subject, and Process

ZHU LanLan HUO Jie GAO YuTing
(Zhengzhou University of Aeronautics, Zhengzhou 450046, P. R. China)

Abstract: As a precious historical document to record the reproduction of family lines and important deeds, genealogy can not only meet people's growing needs for genealogy culture, but also provide a feasible way to inherit and carry forward the excellent traditional Chinese culture. In this paper, based on the development of genealogical literature resources, the "value-subject-process" three-dimensional model of knowledge-based development of genealogical literature is constructed. Taking the authentic genealogy of the Huang Family Archives in Rutang, Henan in Ming dynasty as an example, the problems of data acquisition, ontology construction, knowledge extraction, knowledge fusion, and resource visualization in the knowledge-based development of genealogical resources are explored, so as to provide a new way for the in-depth development of genealogical literature.

Keywords: Genealogical Literature; Knowledge-Based Development; Knowledge Map

(责任编辑: 王玮)