

数字人文视域下古代灾荒文献 知识图谱构建研究*

周树斌 陈红丽 吴艳飞

(华中师范大学信息管理学院, 武汉 430079)

摘要: [目的/意义] 古代灾荒文献种类繁多、检索难度大的问题深刻制约了古代灾荒史的研究, 基于数字人文技术方法对其进行整理分析与研究利用有助于推动古代灾荒文献在新时代的创造性转化与创新性发展。[方法/过程] 本文以古代灾荒文献为研究对象, 一方面从本体层面依据古代灾荒文献的概念知识体系, 结合传统目录学视角构建面向古代灾荒文献的领域知识模型, 完成对古代灾荒文献的语义描述; 另一方面在本体构建的基础上从实证层面构建古代灾荒文献的领域知识图谱, 从场景应用层面完成对领域内文献、人物、时空、事件多维度知识的关联展示以及知识发现研究。[结果/结论] 通过案例化的实证研究验证了基于本体技术构建古代灾荒文献知识图谱的可行性, 为古代灾荒文献在数智时代的知识组织和开发利用提供了新的借鉴。

关键词: 灾荒文献 本体 知识图谱 知识组织 数字人文

分类号: G254

DOI: 10.31193/SSAP.J.ISSN.2096-6695.2024.02.12

0 引言

人类的历史也可以说是一部人类经受灾害困扰, 又坚持不懈与灾害相抗争的历史^[1]。中国地域广阔, 地理环境多变, 自古便灾荒频发。根据《中国救荒史》的统计数据, 从公元一世纪至二十世纪四十年代, 中国共计发生洪水、干旱、蝗灾、冰雹、风暴、疫情、地震、霜冻以及饥荒等多种自然灾害 5258 次^[2]。在这样的历史背景下, 中国形成了数量丰富的灾荒文献, 其中蕴含的海量灾荒知识具有弥足珍贵的史料价值, 而其中古人在历史上抗灾、防灾、救灾、治灾的宝贵经验教训对于依旧饱受灾荒威胁的当代社会大有裨益。

* 本文系国家自然科学基金青年项目“出土文献的多模态知识组织与融合研究”(项目编号: 23CTQ038)研究成果之一。

[作者简介] 周树斌 (ORCID: 0000-0002-9657-8178), 男, 博士生, 研究方向为数字人文, Email: zshubin001@163.com; 陈红丽 (ORCID: 0000-0002-5763-9986), 女, 博士生, 研究方向为信息行为、知识组织, Email: 1517898352@qq.com (通讯作者); 吴艳飞 (ORCID: 0009-0009-1235-4467), 硕士生, 研究方向为数字人文, Email: 323515425@qq.com。

古代灾荒文献作为古代灾荒史研究的一手研究资料,现存文献虽类型丰富,但其分布较广且琐碎,研究手段尚停留在传统目录学、文献学层面,且由于传统人文学者作为研究个体的独立性造就了较多“学术孤岛”的产生,研究成果难以被重复利用,其中蕴含的丰富史料知识尚未得到较好的开发利用。在数字人文浪潮的洗礼下,数字技术赋能传统人文研究逐渐受到业界广泛关注和认可,也为传统的目录学、文献学研究提供了新的方法和工具。从本体基于语义层面对领域内知识概念内涵的规范化表述和完整解读,到知识图谱对领域内知识概念的关联化展示和智能化查询,“本体+知识图谱”的知识组织模式悄然在数字人文领域生根发芽,形成了众多的理论及应用成果。这也为本文灾荒文献中灾荒知识的挖掘、组织和利用提供了方法学参考。

针对古代灾荒文献中概念不明、体系不清、利用困难的问题,本文立足数字人文视域下进行了古代灾荒文献知识组织模式的探索。一方面,利用本体技术对古代灾荒文献中的概念内涵进行明确,在充分结合传统目录学视角的基础上完成针对古代灾荒文献知识体系的准确设定,在“物”“人”“时空”的基础上融合“事件”元素完成古代灾荒文献领域本体的构建;另一方面,基于图数据库构建古代灾荒文献知识图谱,借助图模式的结构精准呈现古代灾荒文献知识,并通过图推理对古代灾荒文献隐藏知识做进一步关联展示以及多维知识发现研究,从而实现古代灾荒文献知识图谱的场景化应用。

1 研究现状

数字人文的兴起为古籍从整理加工到开发利用的整个生命周期提供了新的契机^[3]。当前,在数字人文浪潮影响下,数智技术驱动下的古籍智慧化保护及开发利用成为古籍领域在新时代背景下的新研究态势。其中,从知识组织层面来看,本体、关联数据、知识图谱、纳米出版物等语义网技术被逐步应用到古籍知识组织工作中,有的研究针对某一特定领域古籍进行切入,如程秀峰等^[4]以传统建筑古籍为载体,构建了“宋式大木作营造技艺”本体,并利用知识图谱展示和验证了其应用价值;张卫东等^[5]以《金匱要略》为例,通过构建中医古籍领域本体和知识图谱,对古籍中的非结构化知识进行结构化处理,实现了可视化展示;高劲松等^[6]以《沙湖志》为例,通过构建语义知识关联框架和知识图谱,实现了山水志史料资源的语义化组织和多维知识发现;谢玮等^[7]以《天工开物》版本图像为对象,通过构建本体架构和关联数据,实现了多版本资料的数字化转换和知识图谱应用。有的研究则面向宏观古籍领域,如欧阳剑等^[8]通过知识图谱技术和人机协作,实现了中国历代存世典籍大规模知识图谱的构建;陈涛等^[9]以语义技术为驱动,设计了古籍资源互联互通框架,并验证了该框架在碑帖资源中的合理性和可行性;王晓光等^[10]通过利用本体和纳米出版物等语义技术,对古籍注疏文本进行知识表示和语义化建模,揭示注疏文献中的语义关系,实现了注疏文献的语义化出版与再造。

在中国古典文献中,灾荒文献作为一个重要的组成部分,广泛分布于各类文献之中,但其特点是缺乏系统性和整体性,表现为一种离散化的特征。这一领域的研究传统上属于历史文献

学和灾荒史的交叉学科。在这方面, 邓云特先生的《中国救荒史》是早期代表性专著之一, 它从通史的角度系统地回顾了中国几千年来救荒和荒政的发展, 揭示了中国救荒思想的全貌^[2]。自二十世纪八十年代以来, 出现了如《中国历代灾害性海潮史料》《中国历代天灾人祸表》《中国古代重大自然灾害和异常年表总集》《中国农业自然灾害史料集》《中国气象灾害大典》《民国赈灾史料初编》《中国荒政书集成》等一系列优秀的灾荒史料研究专著。尽管学者们已经在灾害史料的整理和研究方面做出了大量工作, 但相对于中华文明累积的历史典籍中的丰富史料, 这些研究仍然不足以全面覆盖。灾荒文献的论文研究成果从不同的视角出发, 有的专注于特定的文本或思想, 如马梦雪^[11]对《中国荒政书集成》中的救荒书文本进行研究, 李珍珍^[12]对明代中后期的荒政思想进行考察; 也有从宏观角度出发利用目录学视角整理灾荒文献的研究, 例如李芹^[13]对古典文献和目录的考察, 季旭^[14]对历史灾荒文献的系统分类及其发展和流传的变化进行论述。

综上, 现有古籍知识组织的相关研究中, 不同学者以不同的研究视角对古籍蕴含的内在历史文化知识进行了解构、重组、关联。然而现有研究或以“物”为主, 或以“人”为主, 抑或偏倚“时空”, 较少关注到古籍内的“事件”元素。本文以古代灾荒文献为研究对象, 其体例结构决定了其内容包含了丰富的灾荒“事件”元素, 同时该领域尚缺乏数字人文领域的介入, 对其进行知识组织研究从而促进其史料知识在新时代的开发利用有助于推动古代灾荒史学的发展, 而知识图谱的便利化亦可使其利用门槛降低, 由专业研究向大众科普转化。

2 古代灾荒文献知识本体模型设计与构建

本体作为“概念化的明确规范”^[15], 是语义知识组织应用最多的形式之一, 也是一种高效的语义知识建模工具^[16]。本研究基于“六韬法”^[17]从下述六个层面进行古代灾荒文献知识本体模型的设计与构建。

2.1 场景——古代灾荒文献概念分析与类目界定

场景分析是本体模型构建的基础, 将场景明确化是进行领域本体模型构建的基础和前提。本文主要通过对核心概念古代灾荒文献相关概念的分析 and 界定来完成明确本体应用场景的工作。传统研究中, 通常只有古典文献学领域和灾荒史研究领域会涉及古代灾荒文献的研究, 因此对其概念内涵的界定上相对驳杂, 尚未形成统一论断。广义上凡记载与灾荒相关资料的古代文献都可被称为古代灾荒文献, 即广义的古代灾荒文献以字句度量。而狭义的古代灾荒文献以篇目为基本单元。考虑到我国古籍数量众多, 在浩如烟海的典籍中实难完成对广义层面古代灾荒文献的梳理工作, 故本文从狭义视角出发, 将古代灾荒文献粒度最小界定为篇目级。在界定完全局概念的外延后, 需要根据不同文献的结构特点对其内涵类目做进一步划分, 本文主要参考季旭的“四系统说”, 将古代灾荒文献分为灾害志、救荒书、减灾书、集部灾荒文献四种^[14], 具体类目划分如图1所示。

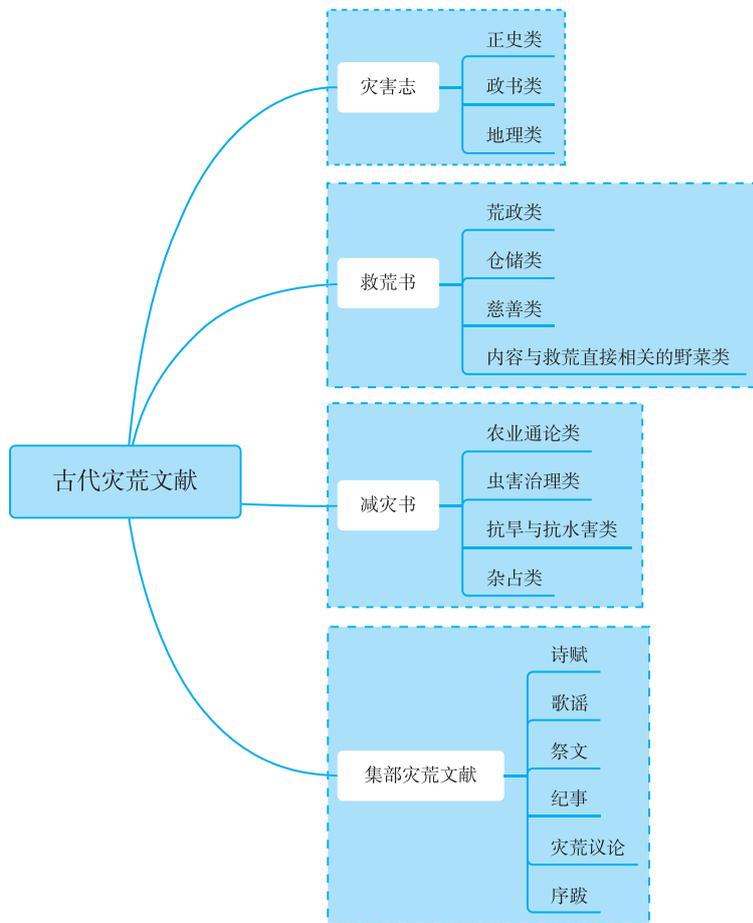


图1 古代灾荒文献类目结构划分

2.2 复用——已有本体复用

在本体的复用维度，主要涉及对现有概念框架的属性进行系统性的审查，旨在识别和采纳那些与特定领域内已定义的类及其属性相吻合的元素。本体的复用有助于降低构建本体的整体成本，并减少工作过程中的重复性任务。此外，当研究人员按照统一的标准规范来定义通用实体时，可以显著提升本体的可靠性与跨系统的可移植性。具体到复用的操作层面，通用本体上主要对 Dc、Foaf、Time、Geo 以及上海图书馆本体模型（SHL）等进行了复用。为了满足灾荒文献领域的特殊需求，自建灾荒文献词表 Disaster Famine Literature（缩写为 DFL）作为补充。

2.3 事物——类及其层级结构设定

事物在本体中表现为实体的具体概念要素，在本体中被称为类。类及其层级结构构成了本体模型的基本概念框架。在明确基本场景的基础上对类进行设定，除基本概念外，还考虑到社会主体、时空、史料来源、事件等诸多要素，共设定七类一级实体，即灾荒文献（DFL: DisasterLiterature）、灾荒类型（DFL: DisasterType）、灾荒事件（DFL: DisasterEvent）、人物（FOAF: Person）、空间（GEO: SpatialThing）、时间（TIME: TemporalEntity）、来源文献（DFL: Sources）七类，通过子类对各类实体做出进一步细分，如灾荒文献包含减灾书（DFL:

DisasterMitigationBook)、救荒书 (DFL: RescueBook)、灾害志 (DFL: DisasterChronicle)、集部灾荒文献 (DFL: CollectionDisasterLiterature) 四个子类, 具体灾荒文献本体类及其层级关系如图 2 所示。

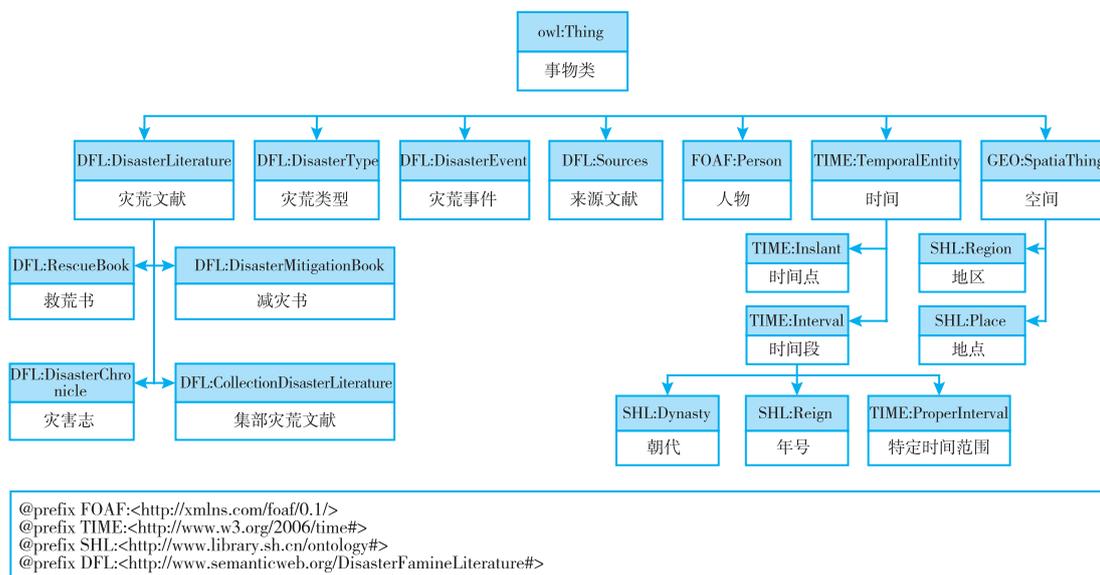


图 2 灾荒文献本体类及其层级结构

从灾荒文献本体自建专有类的维度来看, 灾荒文献是本领域本体的核心客体要素, 其他实体要素均以灾荒文献为核心并与其具有相关性。灾荒类型主要用于表征具体灾荒文献中所记载的灾荒类型, 其主要与灾荒文献和灾荒事件相关联, 考虑到知识图谱的图结构特点, 故将其设为类实体而非属性以便于方便可视化展示。灾荒事件主要用于表示灾荒文献中所记录的具体灾荒事件, 其通常涉及记载于何处、何时、何地、何类, 故与灾荒文献、时间、地点、灾荒类型相关联, 对灾荒事件的具体定义和描述也是本研究区别于其他古典文献知识组织的重要特征所在。由于灾荒文献最小粒度为篇目级, 故很多灾荒文献来源于更大的文献范畴, 而灾荒文献之更高层次的外延并非属于灾荒文献, 故通过来源文献用于表征灾荒文献的母体文献以保证史料资源的完整性与顺承性。

从通用类来看, 人物是灾荒文献中的主体, 一方面, 其作为灾荒文献的撰写者, 灾荒文献的形成受其社会身份所影响和制约; 另一方面, 其作为参与者可构成灾荒文献的内容组成部分。考虑到人物作为社会活动主体, 必然受到时空所制约, 故其与灾荒文献、空间、时间、来源文献相关联。空间和时间实体实例的生成依存于灾荒文献、灾荒事件、人物、来源文献等实体并与这些实体相关联, 构成了这些实体的时空要素, 其中空间包含地区 (SHL: Region) 和地点 (SHL: Place) 两个子类, 用于描述物理空间存在的“面”和“点”的特征。时间包含时间点 (TIME: Instant) 和时间段 (TIME: Interval) 两个子类, 用于描述时间存在的“点”特征和“线”特征, 考虑到中国古代时间描述的文化特色, 将时间段进一步细分为年号 (SHL: Reign)、朝代 (SHL: Dynasty) 和特定时间范围 (TIME: ProperInterval) 三个子类。

2.4 联系和约束——属性词表设定

联系在本体中具体表现为对象属性和数据属性。约束的概念在本体中通过属性的定义域与值域来具体化,用以刻画联系的内涵与外延。在对象属性方面,其定义域与值域均为类别。而数据属性的定义域为类别,其值域则为特定的数据类型。本研究重点关注灾荒文献本体构建中实体类别间的关联。对象属性在此中扮演关键角色,其具备矢量性,能够表达对象间一对多或多对一的关系。例如,在灾荒文献本体中,灾荒文献与灾荒事件之间存在的收录(DFL: recordDisasterEvent)关系,灾荒文献与来源文献之间的来源(DFL: subordinate)关系,以及灾荒文献与人物间的创作(SHL: creatorOf)、被创作(DC: creator)和相关(DFL: releventPerson)关系,均是对象属性的具体表现。研究共设定了21种对象属性,如表1所示。实体自身固有的特性则构成数据属性,通过设定数据属性,本研究充分描述了古代灾荒文献概念的内在属性。具体来看,设计了21种数据属性,如表2所示,其中包括名称(DC: title)和描述(DC: description)等通用属性。为了便于后续在图数据库中构建知识图谱并查找相关实体,这些数据属性的定义域被设定为顶层实体(OWL: Thing),即本体内所有类别共有的属性。

表1 古代灾荒文献对象属性

属性 (Property)	标签 (Note)	定义域 (Domain)	值域 (Range)
DC: creator	创作者 / 责任者	DFL: DisasterLiterature DFL: Sources	FOAF: Person
DFL: releventPerson	相关人物	DFL: DisasterLiterature	FOAF: Person
DFL: subordinate	来源	DFL: DisasterLiterature	DFL: Sources
SHL: temporal	所属朝代	DFL: DisasterLiterature DFL: Sources FOAF: Person	SHL: Dynasty
DFL: releventDisaster	相关灾荒	DFL: DisasterLiterature DFL: DisasterEvent	DFL: DisasterType
DFL: recordTime	记载时间	DFL: DisasterLiterature	TIME: TemporalEntity
DFL: recordPlace	记载空间	DFL: DisasterLiterature	GEO: SpatialThing
DFL: recordDisasterEvent	记载灾荒事件	DFL: DisasterLiterature	DFL: DisasterEvent
DFL: createdTime	创作、创建时间	DFL: DisasterLiterature DFL: Sources	TIME: TemporalEntity
DFL: happenedTime	发生时间	DFL: DisasterEvent	TIME: TemporalEntity
DFL: happenedPlace	发生地	DFL: DisasterEvent	GEO: SpatialThing
TIME: intervalDuring	持续时间段	DFL: DisasterEvent	TIME: Interval
SHL: beginYear	起始于	TIME: Interval	TIME: Instant
SHL: endYear	结束于	TIME: Interval	TIME: Instant
DFL: belongsToRegion	所属地区	SHL: Place	SHL: Region
SHL: creatorOf	创作	FOAF: Person	DFL: DisasterLiterature DFL: Sources
DFL: activeReign	活跃时代	FOAF: Person	SHL: Reign

续表

属性 (Property)	标签 (Note)	定义域 (Domain)	值域 (Range)
SHL: birthDay	生年	FOAF: Person	TIME: Instant
SHL: deathday	卒年	FOAF: Person	TIME: Instant
SHL: nativePlace	籍贯地	FOAF: Person	SHL: Place
DFL: entryTime	入仕年	FOAF: Person	TIME: Instant

表 2 古代灾荒文献数据属性

属性 (Property)	标签 (Note)	定义域 (Domain)	值域 (Range)
DC: title	名称	OWL: Thing	RDFS: Literal
DC: description	描述	OWL: Thing	RDFS: Literal
DC: identifier	资源标识符	DFL: DisasterLiterature	RDFS: Literal
DC: type	文献类型	DFL: DisasterLiterature	RDFS: Literal
DFL: content	内容	DFL: DisasterLiterature	RDFS: Literal
DFL: version	版本	DFL: DisasterLiterature DFL: Sources	RDFS: Literal
FOAF: familyName	姓氏	FOAF: Person	RDFS: Literal
FOAF: gender	性别	FOAF: Person	RDFS: Literal
SHL: courtesyName	字	FOAF: Person	RDFS: Literal
SHL: pseudonym	号	FOAF: Person	RDFS: Literal
SHL: officialExperience	任职经历	FOAF: Person	RDFS: Literal
DFL: writingIdentity	著书时身份	FOAF: Person	RDFS: Literal
SHL: briefBiography	小传	FOAF: Person	RDFS: Literal
SHL: createdWork	著述情况	FOAF: Person	RDFS: Literal
TIME: inXSDDateTime	日期	TIME: TemporalEntity	RDFS: Literal
GEO: lat	纬度	GEO: SpatialThing	RDFS: Literal
GEO: long	经度	GEO: SpatialThing	RDFS: Literal
SHL: city	市	SHL: Region	RDFS: Literal
SHL: country	国家	SHL: Region	RDFS: Literal
SHL: county	县	SHL: Region	RDFS: Literal
SHL: province	省	SHL: Region	RDFS: Literal

2.5 评价——模型评估与展示

评价主要基于本体模型评估及其具体建构展示来实现。本研究采用 Protégé 工具对本体进行具体工程化建模。本研究利用 Protégé 内置推理机进行古代灾荒文献知识本体模型的评估和检验^[18]。检验结果显示, 本体中各要素间不存在互斥关系, 符合本体构建规范。最终, 古代灾荒文献知识本体模型如图 3 所示, 其从模式层完成了对古代灾荒文献的知识描述, 明确了古代灾荒文献概念体系, 也为古代灾荒文献知识图谱构建提供了可靠的知识表示规范。

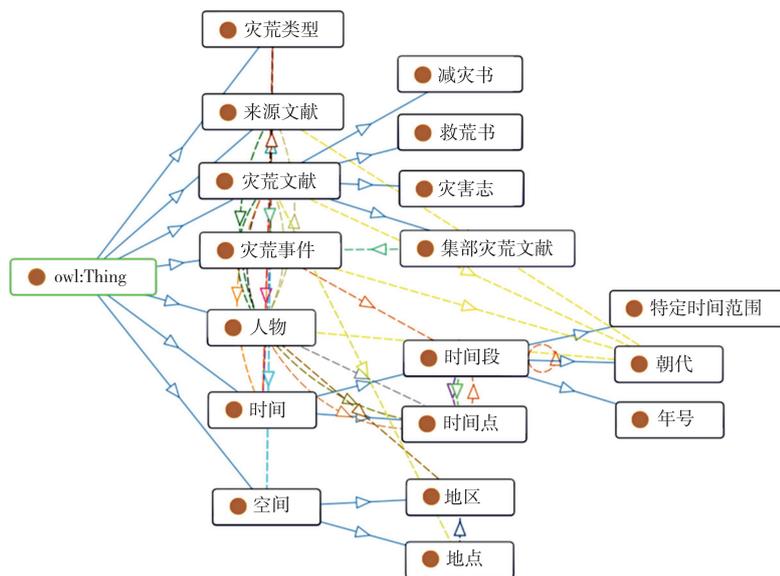


图3 古代灾荒文献知识本体模型

3 古代灾荒文献知识图谱构建

3.1 案例选择及数据采集

本案例选择参考了季旭《中国古代灾荒文献研究》^[14]对古代灾荒文献编目著录情况的整理数据，以其为基础进行了数据外延的扩展，汇集正史、通志、会要、地方志、农书及集部文献等众多权威史料，以确保古代灾荒文献数据的准确性与完整性，共计收集到汉至清代灾荒文献数据 582 条。此外，人物详细数据主要来自国家图书馆历史人物数据库和上海图书馆人名规范库等官方数据库，依据《中国历史人物生卒年表》等文献进行校对，共计收集到人物数据 589 条。空间数据方面，根据文献中的空间信息，参照《中国历史地图集》等文献进行了校对核实。时间数据主要参考上海图书馆编纂的中国历史纪年表。经过人工方式的数据采集和校对，最终将所得数据以结构化形式保存，存储为 CSV 格式文件，为本案例研究提供了可靠的数据基础。

3.2 知识标注与融合

在对古代灾荒文献进行深入的采集和分析后，灾荒事件、灾荒类型及其相关的时空信息等仍以非结构化文本的形式存在于灾荒文献中。为了对这些非结构化文本进行语义标注，采用基于本体模型的方法，构建了语义标签，并根据本体类和属性关系来构建抽象概念和概念实例。本研究主要采用 MARKUS^[19]古籍标注平台来完成知识标注工作。标注完成后，将实体标注数据导出为 CSV 格式的结构化数据。值得注意的是，由于语境和表达方式的不同，标注过程中可能会出现实体和关系的歧义。为了处理这些歧义，通过构建同义词词典实现了各类实体的对齐，并保留歧义实体指向的主要称谓。此外，为每个实体赋予唯一的 ID 标识，以解决同名实体的问题。

3.3 知识映射与存储

知识融合后，通过在 Protégé 中制定规则将处理好的结构化数据导入本体，从而完成 CSV 格

式数据到 RDF 格式数据的知识映射。知识存储方面, 以 Neo4j 图数据库为存储介质, 其具有图模式的知识存储方式, 通过节点和边的关联性形成图式知识结构^[20]。本文主要通过 Neosemantics (n10s) 插件实现 Neo4j 中 RDF 数据的导入, 该插件保证了本体模型及其实例数据到 Neo4j 中的完整迁移。从而实现了基于本体到图数据库的结构映射的古代灾荒文献知识的图数据存储。这一过程不仅为灾荒文献知识的存储提供了新的方式, 也进一步推动了本体在灾荒文献领域的应用和发展。

3.4 知识图谱可视化

在知识图谱中, “节点—边—节点”的三元组构成了知识结构的核​​心, 为本研究提供了一种全面把握古代灾荒文献领域内知识脉络的方法。本研究中, 通过古代灾荒文献知识图谱的可视化展示, 可以将与灾荒文献、灾荒事件、灾荒类型、人物、时空、来源文献等实体数据形成知识网络从而挖掘提取灾荒文献中蕴含的隐性知识并赋能知识发现, 将静态化、扁平化的知识转化成立体化、动态化的知识关联网络, 最终从供给侧促进古代灾荒文献的高效利用。如以集部灾荒文献“《苏学士集》文集卷二《吴越大旱》”为核心节点进行的可视化展示, 如图 4 所示, 清晰地展示了与之相关联的灾荒事件“康定元年吴越大旱”, 记录的灾荒类型为“旱灾”, 作者为北宋时期的苏舜钦。通过这种直观的知识图谱展示方式, 本研究有效地呈现了相关知识的丰富性和深度。

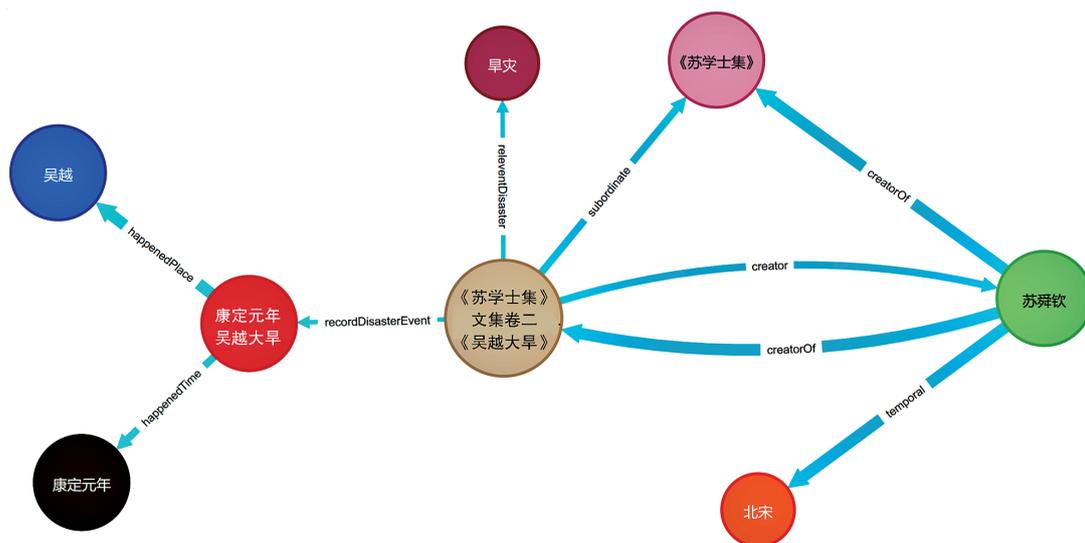


图 4 古代灾荒文献知识图谱可视化呈现 (样例)

4 基于知识图谱的古代灾荒文献研究场景应用分析

4.1 场景一：古代灾荒文献显性知识查询

显性知识查询是一种基于已有知识的明确的信息检索方法。在显性知识查询中, 可以通过明确指定查询条件和参数, 从已有的知识库中获取与查询条件相匹配的信息。在古代灾荒文献研究中, 显性知识查询可以用于快速准确地获取所需的明确信息, 相比于隐性知识发现, 显性知识查

数, 而是通过对数据的全面分析和挖掘, 发现其中的隐藏知识。在古代灾荒文献研究中, 隐性知识发现可以帮助揭示出古代社会对灾害的应对和救助措施中隐藏的知识。

4.2.1 宏观外部知识关联分析

在古代灾荒文献的隐性知识发现中, 宏观外部知识关联分析通过对文献中的知识进行统计和分析, 从而实现对古代灾荒救助整体背景和相关性的揭示。以救荒书为例, 在宏观外部知识关联分析中, 可以通过分析古代灾荒文献中的地域信息, 发现不同地区在救荒工作中的贡献和重要性, 并进一步探究地域因素对古代灾荒救助的影响。通过分析文献中作者的身份背景和官职经历, 可以揭示出地方官员在古代灾荒救助中的作用和影响力, 以及他们与救荒书的关联程度。由于大部分救荒书作者所处年代在明清时期, 故以明清两代救荒书为基础进行知识展开, 形成“朝代—救荒书—人物—地点”的知识发现链, 如图6所示。基于对知识图谱中作者实体的籍贯数据属性的统计, 明清时期在有明确记载籍贯信息的40位救荒书作者中, 有31位作者来自浙江、江苏、上海、江西、山东、福建等华东地区。明清时期, 华东地区作为国家经济发展的核心区域, 承担了大量国家财政税收。与此同时, 面对灾害时经济发达地区往往会遭受更严重的损失, 因此华东地区在防灾救荒方面显得尤为重要。此外, 通过对作者个人经历的数据属性分析, 可以发现救荒书籍作者多数具有官职经历, 部分作者曾任地方官员或具有地方任职背景。地方官员在灾害发生时, 通常处于抗灾前线。如刘世教在《荒箬略》中记录了万历三十七年的一次救荒行动, 其中包括地方官员和士大夫的共同努力。从中也可以看出当时官员在救灾工作中的重要作用。

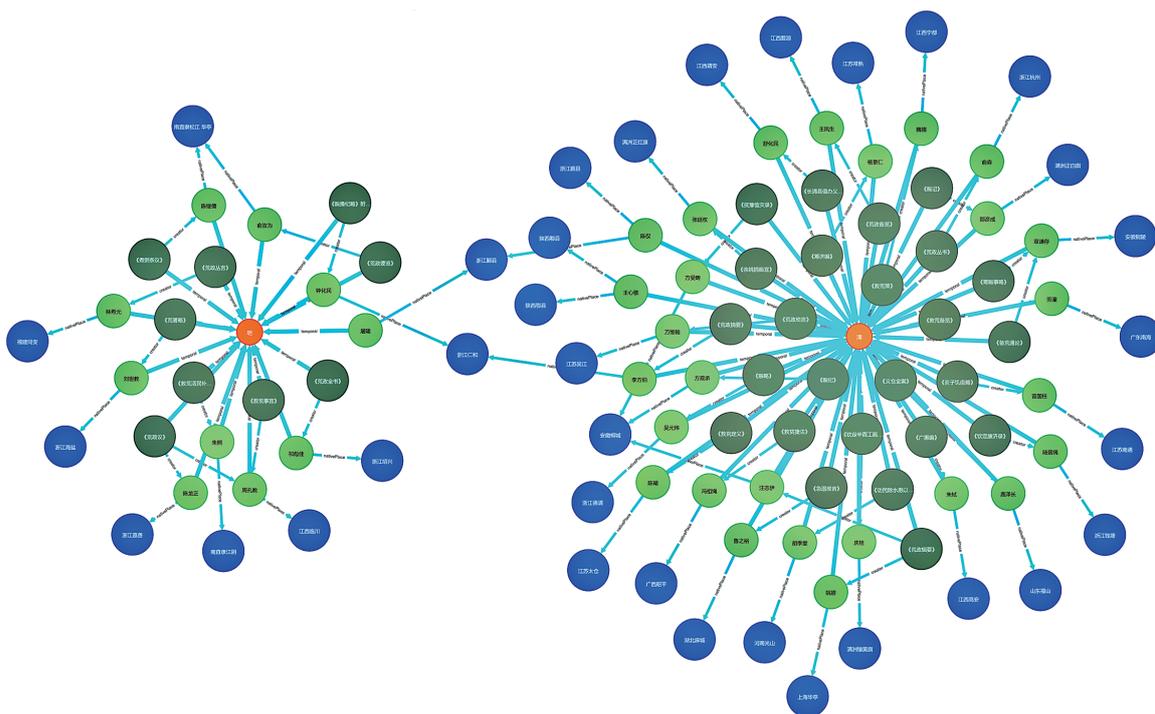


图6 明代救荒书知识关联分析

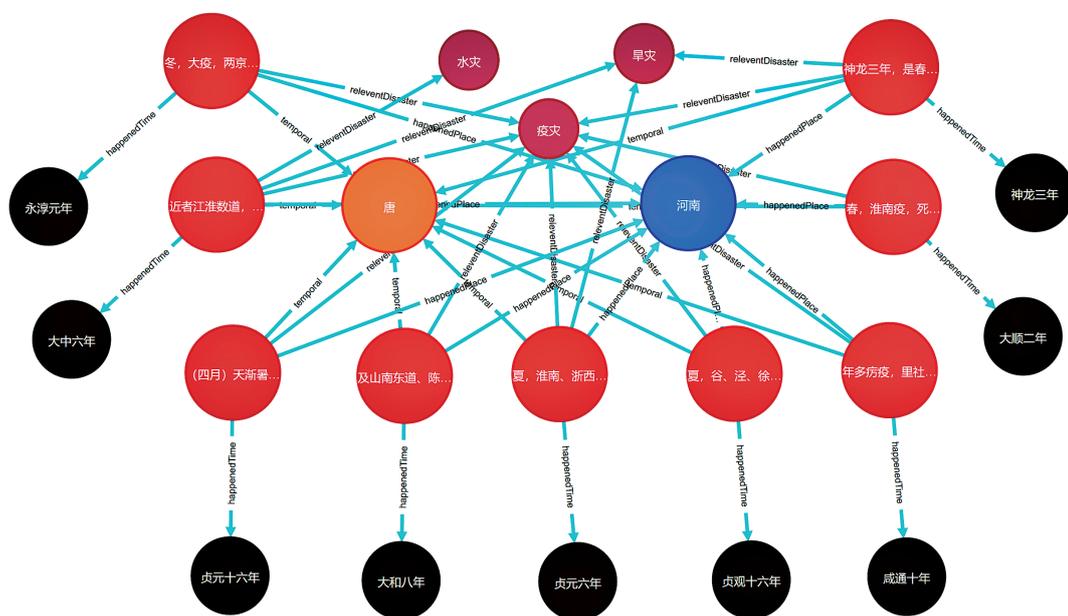


图 8 唐代河南疫灾事件知识发现

5 结语

本研究以古代灾荒文献为研究对象, 立足于数字人文背景, 结合目录学视角基于本体技术对古代灾荒文献领域知识进行了标准化建模, 在此基础上从实证层面进行了知识图谱的构建, 进而在场景应用层面完成了对古代灾荒文献领域内显性知识查询以及隐性知识发现研究。数字人文方法的注入在一定层面上为传统的古代灾荒文献研究提供了在数智时代的知识组织思路。从古代灾荒文献角度来看, 本研究一方面是为了明确和建立古代灾荒文献知识的概念体系。古代灾荒文献历来有种类多样、体例不一的特点, 概念体系众说纷纭, 难成一家之言, 而本体通过对领域内知识规则、体系的制定, 有助于推动其概念体系的标准化发展。另一方面是为推动古代灾荒文献的利用。通过本文的研究可为古代灾荒史研究的学者和爱好者提供更为便利的应用产物, 知识图谱查询的高效性有助于帮助灾荒史研究人员更快、更全、更准地获取所需知识, 而无须再次翻阅海量古籍。从数字人文角度来看, 本研究突破了现有古籍知识组织研究中以“物”“人”“时空”为主要研究对象, 较少关注到“事件”元素的局限, 充分考虑到了古代灾荒文献中包含的“事件”元素, 以及多类型知识元素之间的因果关系。同时, 研究领域的拓展延伸势必既要加强数字技术与人文方法融合的深度, 也要加深研究对象的广度, 本研究以古代灾荒文献为对象在一定层面上也为古典文献学、目录学与数字人文的“破壁”交流提供了一定的案例化参考。

本文的研究不足在于, 本文只是以案例化的方法对古代灾荒文献进行了数据采集和知识图谱构建工作, 因此只能在局部“小而精”地对古代灾荒文献进行知识组织和发现研究, 后续考虑引入自动化的知识抽取方法“大而全”地完成宏观层面的知识组织和应用研究。此外, 将古

代灾荒文献最小粒度设定为篇目级,易导致先秦以前等年代久远文献史料难以被收录到研究范畴中,考虑到其数目较少的特点后续可通过字句级粒度针对典籍语料进行单独本体类目设定和数据采集。

【参考文献】

- [1] 张建民, 宋俭. 灾害历史学 [M]. 长沙: 湖南出版社, 1998: 1.
- [2] 邓云特. 中国救荒史 [M]. 上海: 商务印书馆, 2011: 417.
- [3] 贾楠, 龚蛟腾. 数字人文视域下古籍研究进展及趋势探析 [J]. 图书馆研究与工作, 2023, 226(4): 23-29.
- [4] 程秀峰, 田园, 舒科慧, 等. 面向技艺传承的传统建筑古籍的知识组织——以《营造法式》大木作制度为例 [J/OL]. 图书馆论坛: 1-12 [2024-05-10]. <http://kns.cnki.net/kcms/detail/44.1306.G2.20231102.1741.002.html>.
- [5] 张卫东, 张晓晓. 中医古籍数字资源知识组织与可视化研究——以《金匱要略》为例 [J]. 情报科学, 2022, 40(8): 107-117.
- [6] 高劲松, 周树斌, 高颖, 等. 山水志史料资源语义知识关联与多维知识发现研究 [J]. 情报资料工作, 2023, 44(5): 82-92.
- [7] 谢玮, 衡雨, 邱菊芯. 面向《天工开物》版本图像资源的知识图谱应用研究 [J]. 包装工程, 2023, 44(S1): 480-492, 535.
- [8] 欧阳剑, 梁珠芳, 任树怀. 大规模中国历代存世典籍知识图谱构建研究 [J]. 图书情报工作, 2021, 65(5): 126-135.
- [9] 陈涛, 夏焱, 杨鑫, 等. 语义技术驱动的古籍资源互联互通框架设计与实现 [J/OL]. 图书馆论坛: 1-11 [2024-05-10]. <http://kns.cnki.net/kcms/detail/44.1306.G2.20230601.1316.002.html>.
- [10] 王晓光, 翁梦娟, 侯西龙, 等. 古籍注疏的知识表示与语义化建模研究 [J]. 中国图书馆学报, 2023, 49(3): 75-91.
- [11] 马梦雪. 救荒书中的灾害书写研究 [D]. 西安: 陕西师范大学, 2021.
- [12] 李珍珍. 明代中后期荒政思想研究 [D]. 武汉: 华中师范大学, 2018.
- [13] 李芹. 中国古代灾荒文献编目与著录研究 [D]. 合肥: 安徽大学, 2007.
- [14] 季旭. 中国古代灾荒文献研究 [D]. 西安: 陕西师范大学, 2018.
- [15] Gruber T R. A translation approach to portable ontology specifications [J]. Knowledge Acquisition, 1993, 5(2): 199-220.
- [16] 周树斌, 高劲松, 张强, 等. 文化基因视域下诗词资源多维知识重组与可视化研究——以茶文化为例 [J]. 图书情报工作, 2023, 67(16): 111-123.
- [17] 王文广. 知识图谱: 认知智能理论与实战 [M]. 北京: 电子工业出版社, 2022: 56-71.
- [18] 周育伟, 杨朝红, 王宏宇. 军事领域本体构建 [J]. 计算机时代, 2022(9): 96-99.
- [19] Ho H, De Weerd H. Markus-text analysis and reading platform [EB/OL]. [2024-01-20]. <https://dh.chinese-empires.eu/markus/beta/>.
- [20] 高劲松, 付家炜. 面向书画著录的文献循证与时空关联构建研究 [J]. 大学图书馆学报, 2022, 40(5): 26-36.
- [21] 徐光启. 农政全书 [M]. 长沙: 岳麓书社, 2002: 749.
- [22] 高云波. 唐代疫疾流行与社会主要应对机制研究 [D]. 昆明: 云南师范大学, 2020.

Construction of Knowledge Graph for Ancient Disaster Famine Literature from the Perspective of Digital Humanities

Zhou Shubin Chen Hongli Wu Yanfei

(School of Information Management, Central China Normal University, Wuhan 430079, China)

Abstract: [**Purpose/Significance**] The complexity and difficulty in retrieving ancient disaster famine literature significantly constrains the study of ancient disaster history. Utilizing digital humanities techniques for the organization, analysis, and research of these documents can promote their creative transformation and innovative development in the new era. [**Method/Process**] This study focuses on ancient disaster famine literature. On the one hand, based on the ontological framework derived from the conceptual knowledge system of ancient disaster famine literature and incorporating perspectives from traditional bibliography, a domain knowledge model is constructed to semantically describe the literature. On the other hand, building on this ontology, knowledge graphs for ancient disaster famine literature are developed from an empirical standpoint. From the perspective of practical application, the study showcases the multi-dimensional knowledge association and discovery research of documents, persons, spatiotemporal data, and events within the domain. [**Result/Conclusion**] The empirical case studies validate the feasibility of constructing knowledge graphs for ancient disaster famine literature using ontology technology, providing new insights into the knowledge organization and utilization of such literature in the digital and intelligent era.

Keywords: Disaster famine literature; Ontology; Knowledge graph; Knowledge organization; Digital humanities

(本文责编: 孔青青)