

基于屈原作品的楚辞知识本体 构建思路及实现方法*

徐晨飞¹, 倪媛²

(1. 南通大学 管理学院, 南通 226019; 2. 东南大学-莫纳什苏州联合研究生院, 苏州 215123)

摘要: 以屈原作品为例, 对楚辞原文中各类专有名词进行概念化描述, 建立楚辞知识本体, 以期提高大众对先秦文学的赏析与理解能力。采用本体的工程方法, 针对楚辞作品知识特点, 提出以“人物名”、“植物名”、“动物名”、“地理名”、“物品名”、“乐曲名”、“星宿名”等七个核心概念类为基础, 构建楚辞知识本体。通过设置本体的相关属性和规则, 可实现楚辞显性与隐性知识的推理与深度挖掘。

关键词: 屈原; 楚辞; 本体构建; 知识推理

中图分类号: G254.2

DOI: 10.3772/j.issn.1673-2286.2015.05.003

1 引言

随着现代信息技术的迅猛发展, 知识组织的方法也在不断改进, 传统的分类法、主题法、元数据等知识组织模式已不能完全适应当今数字知识组织的需要。作为知识工程领域的新兴知识组织模式——本体, 已在农业、生物医学、机械工程等不同学科领域取得较为成功的应用。目前, 基于本体的知识组织体系构建研究大多侧重于自然科学领域, 而人文社会科学领域知识由于其较强的时空依赖性、争议性、主观性、模糊性等特征, 相关知识本体构建研究较少。

楚辞是先秦文学的重要组成部分, 是中国文学史上的“总集之祖”, 对整个中国文化系统具有不同寻常的意义。自古至今, 诸多学者对楚辞已展开了大量深入的研究, 包括对楚辞体例、编纂者、内容涵盖范围等信息考察辨析; 对作品中的字词句章的具体释义展开争鸣; 对前人研究成果进行评述; 对楚辞作品中涵盖的人物形象、文化风俗等背景知识进行深入研究等等。这些前人研究中蕴含了丰富的知识, 南通大学楚辞研究中心的钱智勇、周建忠等学者就曾撰文探讨如何基于本体构建楚辞研究文献的知识组织体系, 来帮助楚辞研究人

员提炼、组织知识^[1]。但另一方面, 楚辞作品本身也包含了大量可以挖掘的知识, 如楚语、楚声、楚地、楚物等, 将这类知识进行重新分类、整合、组织将有助于普通楚辞爱好者对楚辞原文乃至楚文化有更好地理解 and 掌握。本文则在此背景下以屈原所作楚辞原文为研究对象, 探讨楚辞知识本体的构建思路及实现方法。

2 相关概念概述

2.1 本体

本体 (Ontology) 的概念起源于哲学领域, 哲学中把本体定义为“对世界客观事物的系统描述”。在信息科学领域对本体的定义有很多, 最普遍和常用的定义由 Thomas R. Gruber 在 1993 年提出: “本体是概念模型的明确的规范说明”^[2]。一般来说, 本体由类、属性、关系、约束、公理和实例五大部分组成。“类”表示现有物体的一个基本单元, 无论是物理实体 (人物、动物、植物等) 还是抽象实体 (年代, 时间等); “属性”既可以描述一个类的内部固有本质或特征, 也可以用于连接不同类之间的关系; “约束”是关于概念的属性或关

* 本研究得到国家自然科学基金重大项目“东亚楚辞文献的发掘、整理与研究” (编号: 13&ZD112)、国家自然科学基金项目“楚辞文献语义化研究” (编号: 10BTQ031) 和南通大学人文社会科学研究项目“楚辞文献本体构建方法研究” (编号: 12W28) 资助。

系的一种规则或条例;“公理”是一种约束条件,其值始终为真,是语义推理的基础;“实例”(亦称为“个体”)是一个类的最特殊的实体^[3]。图1列举了本体的组成及其在楚辞知识组织中的一个示例:“芷”和“兰”在本体中都被定义为“香草香木类”的“植物”,“棘”被定义为“恶草恶木类”的植物,其中“兰”具有“描述过屈原”的性质,“芷”有属于“伞形花科”的固有本质;同时,“屈原”在该本体中定义的另一顶层概念“人物”中属于“忠能贤贞之臣”,官职为“三闾大夫”。

2.2 基于本体的知识组织体系

知识组织体系构建的终极目标是将知识资源有序化和规范化后按需求提供给用户,其间需要用户行为模式与知识资源之间的匹配与互动,是一种知识“固化-活化”的相互转换升级,旨在提供可控性的高效知识服务^[4]。

知识组织体系的基本特性就是结构性。其本质就是通过一定的技术手段,如本体、主题图等,将内在于知识体系中反映知识结构的概念体系组织成一个能够有效获取知识的知识网络^[5]。主题图比较直观易懂,而且能将概念及其相关概念联系起来。但是由于主题图特定的语法形式,其概念体系与本体相比,推理功能不足^[6];而本体不仅能准确规范地描述相关概念及其之间的关系,构造丰富的语义关系网络,而且还具有知识推理功能,可以深度挖掘概念之间隐含的知识关系。因此,本体亦成为目前各学科领域构建知识组织体系的首选方法。

3 楚辞知识本体构建思路

3.1 楚辞作品知识特点分析

楚辞与上古歌谣和神话、诗经、先秦散文共同构成了先秦文学体系。在构建楚辞知识组织体系时首先要充分考虑其作品的知识特点。首先从体制与内容两个角度来进行分析。

从体制上来看,楚辞在漫长的传播过程中,它逐渐固定为三种含义:第一,诗歌的体裁,即出现在战国时代,楚国地区,以屈原为代表的诗人在楚国民歌基础上开创的一种新诗体;第二是作品,是指战国时代的楚人及以后的一些汉人创作的一批作品,其收入了屈原、宋玉的作品以及汉代贾谊、王褒、刘向等人的仿骚作品

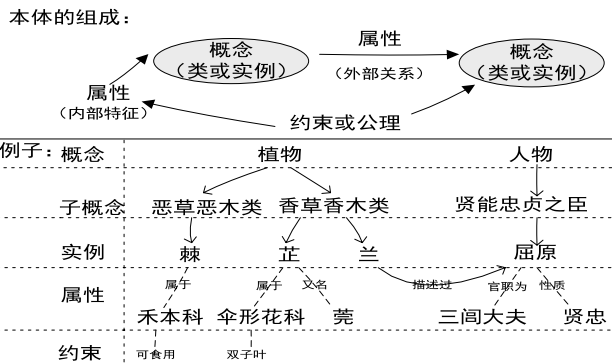


图1 本体的组成和一个示例

等;第三是书名,指汉人对楚国人以及汉人缩写的诗歌辑选而成之书^[7]。

从内容上来看,楚辞既非出于一个时代,也非出自一人之手,它是由不同的时代和作者逐渐纂辑增补而成。楚辞一书的篇目内容与纂辑过程透露了“楚辞”二字的意义:即屈原辞赋以及宋玉以下的汉人伤悼屈原、以事名篇的拟骚辞赋或骚体赋、以事名篇的拟骚辞赋或骚体赋^[8]。

通过体制与内容上的界定,我们选择楚辞的精华所在,即屈原本人所作楚辞部分作为知识本体构建的参考基础。

楚辞中屈原的作品大致可分为三类:《离骚》类,《天问》类和《九歌》类。第一类的主体为政治抒情诗,具有“自传性”。如《离骚》、《九章》是带有其生活经历和思想的诗歌,在比较现实的描写之中也带有神奇的非现实成分,常以幻想境界表达自己的情感。第二类是以《天问》为代表,展现了屈原学术思想和哲理智慧的抒情诗。《天问》中大多是对宇宙天体,神话传说和历史兴亡的提问,通过该部分可以得到一些古代传说的一些梗概,甚至可以据此考订神话传说的时代性真伪;第三类以《九歌》为代表,该部分是对南楚民间文学的继承与创造。该部分祭祀了大量的神灵,有东皇太一、河伯、湘君、湘夫人等。

此外,楚辞作品中有大量的实词、虚词、专有名词以及社会风俗、典章制度等。综合屈原所作楚辞作品的知识特点,笔者选择其中的专有名词为特定研究对象并加以分类,目的是厘清各类专有名词的概念及它们之间的联系,并以此构建楚辞知识本体。

3.2 楚辞知识本体构建方法

在本体的众多构建方法中,七步法、IDEF5法、

METHONTOLOGY法、TOVE法、骨架法等具有一定的先进性,但其使用的领域不同^[9]。例如METHONTOLOGY法专用于构建化学领域本体,IDEF5法主要用于获取企业领域本体等。相比较而言,由斯坦福大学创建的七步法较为成熟,且适用于领域本体的构建。基于楚辞知识本体的构建需求,本文主要以七步法为基础,结合其他本体构建方法的优点,最终确定楚辞知识本体构建方法,其流程如图2所示。相关步骤包括:

(1) 识别楚辞知识本体的专业领域和受众范畴。该部分包括楚辞知识本体的所属领域,目标用户的设定以及系统功能的界定等。

(2) 考察复用现有本体的可行性。由于楚辞研究长期存在争议性、模糊性以及主观性,目前亦无成型或可使用的本体系统,因此复用现有本体的可能性极小。

(3) 楚辞本体库的设计。具体包括:在需求分析基础上,以楚辞原文和相关研究文献为基础,明确楚辞知识本体的核心概念集;对核心概念集依照自顶向下的原则进行扩展,建立等级层次关系;定义本体的属性,包括对象类型属性和数值类型属性;定义属性的分面,主要是进一步定义属性值的类型,属性的定义域和值域以及属性之间限制关系如逆关系(inverse)、传递关系(transitive)等;创建实例。

(4) 逻辑检测和推理。在最后阶段可用推理机对该本体中的概念及实例进行一致性和冲突检测,以确保逻辑上的正确性。此外,为挖掘楚辞知识体系中的隐含知识,可利用本体系统中的推理功能,将隐含其中的知识显性化,实现知识的推理功能。

3.3 楚辞知识本体构建工具

通过对国内外几种较经典的本体开发工具进行比较与试用,笔者最终采用Protégé^[10]作为楚辞知识本

体的开发工具。Protégé使用简便,同时它支持数据存储、图形化本体编辑模式、一致性检测和语法检测等功能。此外,Protégé还支持插件扩展,如可视化组件、推理引擎等等。

4 楚辞知识本体构建步骤

4.1 创建类及其层次等级结构

构建楚辞知识本体,其核心工作是定义本体的类、属性及关系,并明确其适用范围。首先,我们对楚辞中的专有名词概念进行区分与定义,确定每个概念的边界,再根据前人研究内容以及目标用户需求设定本体的概念层次等级结构。

(1) 确定核心概念集

通过对楚辞原文的梳理并借鉴赵逵夫先生的《楚辞语言词典》^[11],本文对屈原所作楚辞部分的专有名词进行了概念划分,最终确定将人物名、植物名、动物名、地理名、物品名、星宿名、乐曲名等七个大类作为一级核心概念集,每个概念皆可划分各种不同的属性。

(2) 构建层次等级结构

由于屈原创作的内容中既包含了对现实世界的描写,还包含了对精神世界的追求,因此作品中存在大量写实与神话的概念。故如人物名、动物名、植物名、地理名等类别还可以向下细分为“神话类”和“现实类”两个类别。相关类的层次等级划分方法如下:

“人物名”的划分

在楚辞作品中,所描述的非现实人物不仅有神话人物,还有一部分是作者臆想出来的人物对象(例如“女巫”)。考虑到二者都属于非现实人物,因此笔者将之归为“神话臆想类”,并细分为两个子类:“神话人物”和“臆想虚拟人物”。再者,屈原所作楚辞部分,有相当一部分作品蕴含了其丰富的政治情感。例如在《离骚》中,屈原既用了大量的篇幅从反面揭露群小丑态以明其节,又有直抒胸臆,正面赞扬前贤美德以申其志,以希冀君主明辨等,因此我们可将“现实人物”进一步划分为“君主”、“大臣”两类。“君主”可细分为“贤君”和“昏君”两类,“臣子”类又可细分为“贤忠之士”和“昏庸奸佞之辈”两类。此外,“现实人物”类中还应包括一些“后宫女子”(如褒姒、闵妃等)以及接舆、渔夫和伯乐之类的“普通人”,故再添加“后妃”和“百姓”两类。人物类的概念层次图如图3所示。

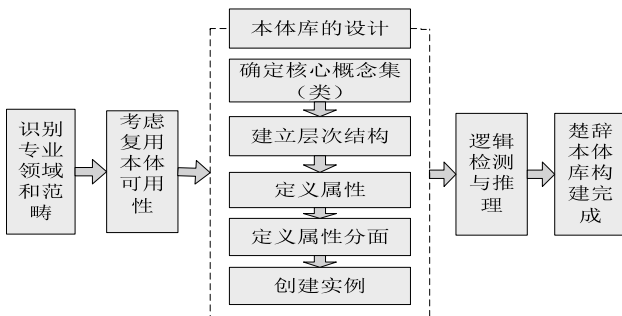


图2 《楚辞》知识本体构建流程图

“植物名”的划分

古人表达感情比较含蓄, 善用比拟, 经常用山川草木来隐喻爱恨情仇。对于“植物名”的细分, 屈原在楚辞中提及的植物有上百种, 从寄寓言志的角度可将约略分为香草(木)、恶草(木)两大类。香草(木)象征贤良、忠贞之士, 恶草(木)象征昏庸奸佞小人。至于“香”与“恶”的区分, 则大多与植物的特性有关^[12]。大部分的植物整体或部分器官带有香气的, 如花、果、叶等部分带有香气, 都是楚辞中的“香草香木”, 例如蒹、芷、兰、留夷、揭车等; “恶”就是指那些植株气味难闻或身上带刺的植株、以及一些杂草和蔓藤类, 如艾、薺、菘、蕲等。还有一些并非是“恶草恶木”或“香草香木”, 但仍可以划分为上述两类之中, 这与作者的写作意图有关。例如藿(豆)并非是恶草, 但《楚辞》中作者用其反衬美好正直的事物, 具有负面意义, 因此也将其划分到“恶草恶木”一类中。此外, 按释文内容判断其代表的用途, 将剩余的其他植物划分为“写景写物”类和“经济植物”类。“写景写物”指的是一些写景寓情的植物, 如宿莽、菅、青莎等, “经济植物类”是指那些用材的树木或食用的植物, 如柏、栗、黄粱、蒲等。

“动物名”的划分

“动物名”这一概念类的划分标准有很多, 比如划分为有翼动物和无翼动物, 或者脊索动物和无脊索动物, 再如爬行类和两栖类等等。但由此划分出来的子类太多, 且从诗人情感表达以及文意的理解角度来看也不具太多现实意义, 于是本文参照周秉高的《楚辞原物》中对动物的划分原则并通过多种分类方法的比较分析, 最终确定将“动物名”划分为“禽”、“兽”、“虫”、“鱼”四类^[13]。《尔雅》有云: “二足而羽谓之禽, 四足而毛谓之兽”。因此, 这里的“禽”主要泛指鸟类, “兽”指那些陆地行走的动物。

“物品名”的划分

楚辞中出现的物品可反映当时的社会风俗和文化背景。楚辞中的“物品名”根据用途可划分为“器物类”和“服饰类”。“器物类”又具体包括“乐器”、“兵器”、“工具”(包括舆、轩、鞅等交通工具)、“器皿”和“珠宝玩物”等5个子类。“服饰类”则可再划分为“衣服类”、“饰物类”和“质料类”。其中“质料类”是指作品中描述到的一些衣料, 如罗、縞、绮、纡等。

“地理名”的划分

楚辞作品中出现了大量的地理名称, 具体可划分为“国名”、“州名”、“地名”和“山川名”四大类。其中

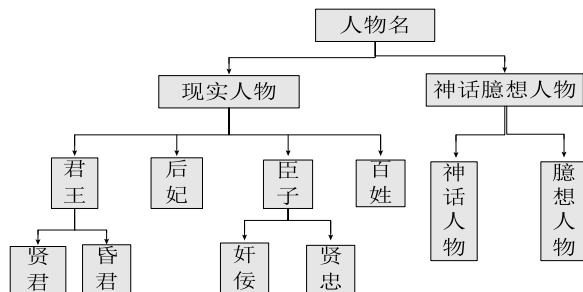


图3 “人物名”概念层次结构

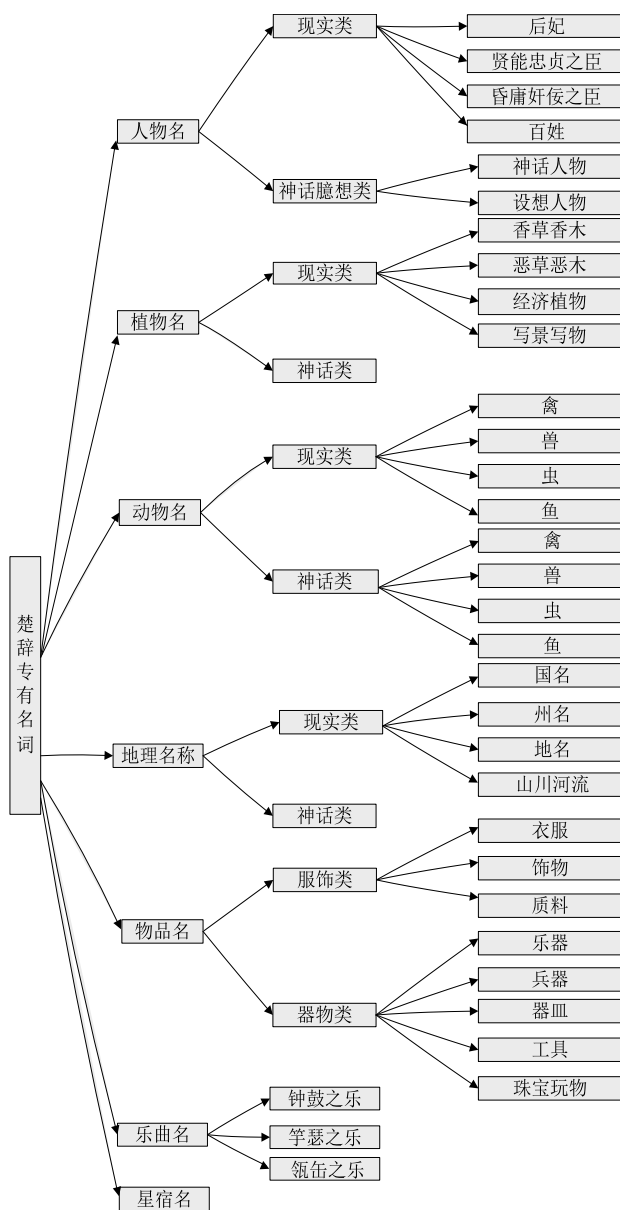


图4 楚辞知识本体的概念层次图

“地名”和“山川名”一部分是属于神话想象, 一部分是属于写实, 因此有必要将其划分为“现实类”和“神话想象类”。

“乐曲名”的划分

楚辞中还记录了许多举办宫廷宴享、民间祭祀的乐舞场面，这不仅给我们留下了宝贵的南楚音乐，也对我们研究楚文化有重要的意义。《墨子·三辩》中把战国时期乐舞在不同表演场合配用不同乐器的乐舞分为三类：“钟鼓之乐”、“竽瑟之乐”、“瓠缶之乐”，楚辞之记载的乐曲也大致可以这样分^[14]。“钟鼓之乐”多用于宫廷宴享，“竽瑟之乐”多用于祭祀仪式，而“瓠缶之乐”则是指古代劳动人民随手拿器物伴奏歌唱的乐曲。例如《渔父》中描述到渔父“鼓枻而去，乃歌曰……”即属于该类。

“星宿名”的划分

古代天文学家们按照人间的某些政治格局来想象天上的星宿格局，即有尊卑之分。例如《淮南子·天文训》中将浩淼天空化为五块：东方其帝为太皞，其兽为苍龙；西方其帝为少昊，其兽为白虎；南方其帝为炎帝，其兽为朱鸟；北方其帝为颛顼，其兽为玄武；中央其帝为黄帝，其兽为黄龙。屈原的作品中共出现了十八种星宿，但其笔下的星宿无尊卑，充满着人性，是文学形象的具体表现^[15]。并且，对于楚辞中出现的十八种星宿，历代注家认知亦不尽相同。因此，笔者对“星宿名”这一核心概念将不予细分子类。

按上述划分方式处理后，最终得图4所示的本体概念层次结构图。再通过Protégé本体构建工具，遵循自顶向下的本体建立原则，构造楚辞知识本体的类及类层次结构，如图5所示。

4.2 定义类属性及其约束

构建了本体的类及其层次等级结构之后，形成的只是基础框架，还不足以描述整个领域内的知识，因此要

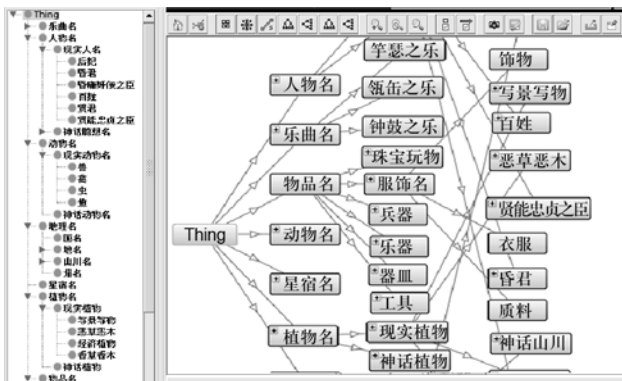


图5 楚辞知识本体类层次结构

继续向其中添加属性并为这些属性添加约束和限制。属性分为数据类型属性 (Datatype Property) 和对象类型属性 (Object Property)。数据属性主要用于描述概念自身固有的特性，所以又被称为概念的内在属性，其值域只能是某一数据类型。例如人物名的数据属性有别称 (字、号、谥)、性别、官爵、品阶、参与的事件等；植物的数据属性有古名、今名、所属科种等。对象属性主要用于描述同类概念之间以及不同概念之间的联系，所以又被称为概念的外在属性。对象属性将类与类之间，实例与实例之间相互关联，对基于本体的推理提供了重要的语义基础^[16]。如人名对象属性“是……的父亲”、“是……的臣子”描述了两个人名类实例之间的父子关系和君臣关系，植物的对象属性“指代过……”描述了植物名类与人物名类之间的关系，通过这些关系能将不同的概念联系起来。

人物关系错综复杂，不仅包含自身的一些性质 (别称、性别、品阶等)，还包含一定血缘关系 (父母与子女、兄弟与姐妹、亲属等)、社会关系 (朋友、敌人、情人等)，因此构建人物属性时较其他类别更为繁琐。笔者尝试以“人物名”为中心，通过相关属性与其他类概念联系起来，如图6所示。

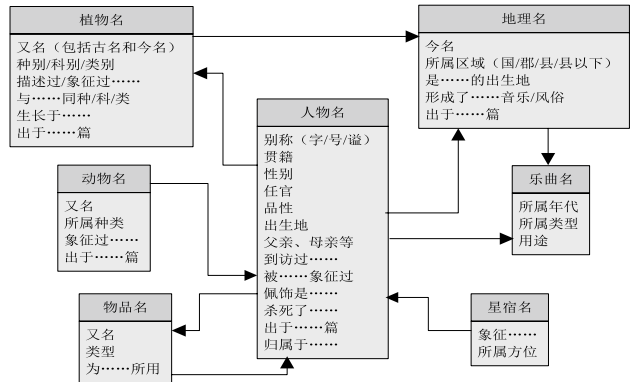


图6 类概念属性联系

4.3 添加实例

添加实例是本体开发过程中工作量较为繁重的一部分，需将楚辞作品中专有名词按照类进行标引并提取出来。通过软件工具与人工评估取舍的方法，最终确定了共451个实例，其中人名类实例129个、动物名67个、植物名62个、地理名称72个、古乐名15个，物品名88个、星宿名18个。表1列举了《离骚》中提取的实例：

表1《离骚》中的专有名词实例

人物名	高阳(颛顼)、伯庸、屈原、尧、舜、禹、汤、周文王、夏桀、商纣(辛)、彭咸、鲧、夏启、五子(五观)、羿、浞、浇、少康、皋陶、傅说、武丁、吕望、宁戚、齐桓公、羲和、望舒、飞廉、丰隆、宓妃
动物名	骐骥、鸷鸟、鸱、鸱鸺(伯劳)、鸾皇(凤凰)、鹭、虬
植物名	江离、茝(芷)、兰、木兰、宿莽、椒、桂、木根、薜荔、蕙、荃、留夷、揭车、杜衡、薺、葳、蕤、芙蓉、艾蒿、茅、萧艾、檉、扶桑、若木
地理名	阝、苍梧、閼风、涓盘、白水、流沙、赤水、西皇、羽山、有娥、有虞、傅岩、昆仑、县圃、崦嵫、咸池、穷石、不周、西海
物品名	鞿、纆、规、矩、绳墨、杓、筭、箝、瓊佩、佩帉、瑶象
星宿名	摄提、孟陬
乐曲名	《九辩》、《九歌》、《韶》

我们使用Protégé将所有实例添加到每一个类及其子类中,并在实例编辑界面直接添加每个实例的属性关系。例如添加人物的父子、母子、夫妻、君臣等关系;植物与人物的象征与被象征关系;人物名与地名的出生地、拜访地等关系等等,通过设置这些关系为后续的知识推理做好铺垫。此外,还可在编辑界面中对先秦古汉语的实例进行现代汉语解释和说明,以方便其他本体使用者进行阅读和理解。

4.4 可视化展现

利用Protégé进行本体设计时,父类与子类间的关系以及实例之间的关联均可通过Ontograf插件以网状结构图的形式形象化展现出来。Ontograf中提供了检索界面,在检索框中输入关键词即可查询实例之间,或实例与类之间的关系。例如,“人物名”类的实例中有“屈原”和“伯庸”等人,他们都归属于“楚国”,而“楚”是国名类的实例,从而产生了“人名—国名”关联,如图7所示。

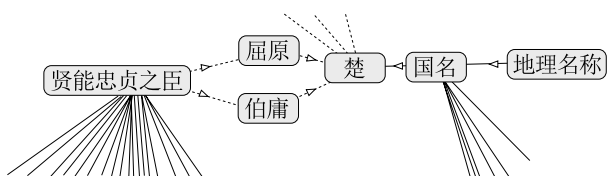


图7 归属关联实例图

此外,“人物名”类与其他类之间还有其他更多的关系可以挖掘展示:如“屈原”与“兰”之间的联系体现的是“人物名”类与“植物名”类之间的象征与被象征的关系;“屈原”与“沅”、“江”、“汉北”、“洞庭”等体现的是“人物名”与“地理名”之间的到访关系、归属关系和出生地关系;而“屈原”与“楚怀王”、“顷襄王”之间则是“人物名”类之间的君臣关系,与“伯庸”是祖孙关系……这些属性关系皆可在本体中通过可视化视图进行展现。如图8所示。

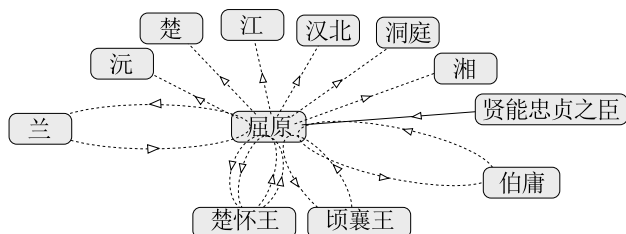


图8 以人名类为中心的关联实例图

5 基于本体的楚辞知识推理

楚辞知识本体构建完成后,自身即包含着多种多样的关系,有些关系是显性的,有些则是蕴含在其中未被发现的。比如,楚辞中涉及到的人物中有同乡、同龄、同朝为官等关系;植物中有同类(科)等关系等。这些关系可以通过本体的推理功能获得。

5.1 推理实例一

楚辞中没有直接说明“屏风(莽)”与“柘(甘蔗)”的关系,且两者一个属于“写景写物类”,一个属于“经济植物类”,但由于二者具有共同的属性,即皆属于“禾本科”,于是可以推理得出两者是同一科的植物。

具体方法为在Protégé中附加Jess推理引擎,通过构建SWRL推理规则,实现楚辞知识推理。在本实例中,判定两植物之间是同科关系的SWRL规则代码如下:

```
现实植物(?x) ∧ 现实植物(?y) ∧ family(?x, ?z) ∧ family(?y, ?a) ∧ sameAs(?z, ?a) → same_family(?x, ?y)
```


- [4] 马文峰, 杜小勇. 关于知识组织体系的若干理论问题[J]. 中国图书馆学报, 2007, 33(2): 13-17, 46.
- [5] 滕广青, 田依林, 董立丽, 等. 知识组织体系的解构与重构[J]. 理论与探索, 2011, 34(9): 15-18.
- [6] 张莉萍. 基于本体论的知识组织与知识检索[J]. 情报资料工作, 2008 (06): 27-30.
- [7] 金开诚. 楚辞选注[M]. 北京: 北京出版社, 1980.
- [8] 李大明. 汉楚辞学史[M]. 北京: 华龄出版社, 2004.
- [9] 廖作芳. 《三国志》历史领域本体的构建与推理研究[D]. 武汉: 华中师范大学, 2011.
- [10] The Protégé Ontology Editor and Knowledge Acquisition System [EB/OL]. [2011-04-20]. <http://protege.stanford.edu/>.
- [11] 赵逵夫. 楚辞语言词典[M]. 上海: 上海辞书出版社, 2013.
- [12] 潘富俊. 草木零落, 美人迟暮: 楚辞植物图鉴[M]. 北京: 九州出版社, 2014: 7.
- [13] 周秉高. 楚辞原物[M]. 内蒙古: 内蒙古大学出版社, 2008: 165-169.
- [14] 徐徐. 楚辞中的楚乐舞考辨[J]. 荆州师专学报(哲学社会科学版), 1987(4): 42-47.
- [15] 周秉高. 楚辞星宿考[C]//中国屈原学会. 2007年楚辞学国际学术会议论文集. 北京: 学苑出版社, 2007: 55-66.
- [16] 闫宇. 领域本体中规则推理研究与实现[D]. 上海: 华东师范大学, 2010.

作者简介

徐晨飞, 1981年生, 南通大学管理学院讲师, E-mail: zidane5@ntu.edu.cn。
倪媛, 东南大学-莫纳什苏州联合研究生院硕士生。

Knowledge Ontology Construction and Realization of Chu Ci: A Literature Work from Qu Yuan

XU ChenFei¹, NI Yuan²

(1. School of Management, Nantong University, Nantong 226019, China;

2. Southeast University-Monash University Joint Graduate School, Suzhou 215123, China)

Abstract: Example is works from Qu Yuan - Chu Ci, which has conceptualize the proper nouns and build its knowledge ontology, with expectation to improve the public's ability in appreciating and understanding Pre-Qin literature. Referring to the characteristics of Chu Ci, using ontology engineering method, the paper sets Character Name, Plant Name, Animal Name, Geography Name, Item Name, Music Name and Stars Name as core concepts which are basis of the construction of Chu Ci knowledge ontology. By setting relevant properties and rules of the ontology, it can realize Chu Ci dominant and implied knowledge reasoning and deep mining.

Keywords: Qu Yuan; Chu Ci; Ontology construction; Knowledge reasoning

(收稿日期: 2015-04-29; 编辑: 王立学)