

基于本体的楚辞书目隐性知识发现^{*}

周澍绮, 许富宏

(南通大学文学院, 南通 226019)

摘要: 在楚辞书目中, 存在着各种已知的未知的关系。本文通过本体规则构建楚辞书目的推理来挖掘楚辞书目之间蕴含的关系, 完善书目对象的关系描述, 并在此基础上完善楚辞书目未知信息。在探讨古籍书目的智能推理和隐含知识发现的同时探讨古籍知识提取的注意点。

关键字: 楚辞; 古籍版本; SWRL

中图分类号: G250

DOI: 10.3772/j.issn.1673-2286.2015.07.006

1 引言

在信息爆炸和知识蔓延的今天, 传统意义的文献学的研究无法满足日益多元化的需求, 人们对古典文献的需求不仅仅是要在一本古书, 一篇古文中, 更多的时候读者需要的是古典文献中符合其需求的知识。

楚辞作为先秦文学的重要组成部分, 数千年流传至今的版本多如繁星。楚辞不同版本书目在古典文献学意义上蕴藏着各种各样的信息。这些信息中存在着各种各样的知识关系。现有抽取的关系多是将句子中已存在的字词作为实体关系描述, 对本体的应用也仅仅停留在将其作为一个建模工具, 而忽视了它的推理功能^[1]。所以, 简单的楚辞书目关系抽取已不能满足楚辞研究爱好者的需求。

鉴此, 笔者尝试通过SWRL构建楚辞书目规则来挖掘实体未知关系, 以此完善本体知识库中楚辞书目的关系描述, 为楚辞书目的关联检索奠定坚实基础, 同时也为相关研究人员深入了解楚辞领域知识提供新的途径。

2 研究背景

2.1 语义规则语言

语义规则语言 (Semantic Web Rule Language) 即 SWRL, SWRL是由OWL子语言OWL-DL与OWL-Lite为基础的规则描述语言^[2]。SWRL语言在OWL本体组织的基础上加入具有很强的逻辑表达能力规则, 弥补OWL语言在推理中的不足, 实现了规则与知识深度结合。

SWRL的优势是能转化为规则系统中的规则, 从而实现OWL系统中规则与知识的关联和重用性。SWRL能够重构概念间的关系, 设定本体定义的属性, 推导本体逻辑中隐性的关系, 同时可将实例加入到相应关系中。所以, 使用具有强大的规则表达能力的规则语言SWRL来表达楚辞书目推理规则, 既能与楚辞书目本体有很好的交互, 又能获得较高的智能性楚辞书目信息检索。

2.2 JESS

JESS (Java Expert System Shell) 是一个使用Java

^{*} 本研究得到国家社会科学基金重大项目“东亚楚辞文献的发掘、整理与研究” (编号: 13&ZD112) 和国家社会科学基金一般项目“楚辞文献语义化研究” (编号: 10BTQ031) 资助。

编写的专家系统,由事实库(working memory)、规则库(Rule base)、推理机(inference engine)三大部分组成,它同时支持正向和逆向推理,并采用产生式规则作为基本的知识表达^[3]。JESS在人工智能的许多领域有着广泛的应用并且具有非常广阔的开发前景。

JESS中的推理机能够弥补SWRL不实现规则的推理的不足,它可以将知识和SWRL规则库连接起来。系统为JESS提供特有领域的规则和事实信息,JESS就可以执行该领域的推理任务。OWL本体知识和SWRL规则经过相应处理,转换为JESS支持的知识格式和规则格式。

2.3 楚辞书目关系

楚辞书目本体建设是楚辞知识库的重要部分,楚辞书目本体构建参考传统古籍版本体。本体构建所用属性以时间和排版为代表,前者包含万历本,五代本等类似信息;后者有卷数,鱼尾,象鼻等版本属性。本文的版本信息不包含广义版本中的稿本,写本,抄本等载体属性。

楚辞版本中各属性,主要分为以下四类:

呈现样式:依不同版本中鱼尾单双,象鼻有无,四周线色,黑白口等鉴定版本,其中字体,纸张,装帧等样式存在较大不确定性,此处暂未列入。

序跋内容:依序跋中所涉及的时间鉴定,同时序跋内容还可以鉴定版本的流传情况;

卷数变迁:卷数属性在同一系统的不同的版本大致相同。若卷数相近,分卷相似,就可以初步鉴定为同一系统;

藏书信息:藏书信息主要分为藏书印章,扉页藏书楼信息判断版本的收藏梗概,递藏关系。

自编目角度而言构建的楚辞书目,分为以下六类^[4]:

等同关系:内容和著作方式相同的作品,用于研究作品与复本、影印本、重印本以及其他复制品之间的关系;

次生关系:原作和根据原作修改后得到的作品;用于研究作品的版本源流、改编出处等;

描述关系:原作品和对原作评论、注释、翻译的作品,用来发现注释本、评论作品和原作的关系;

附属关系:书目与其附件之间关系,如续编作品;

连续关系:书目呈现的时间性关系,刊刻时间,翻刻时间等;

共通关系:书目有相同著者、编者、出版信息、馆藏信息等,共通关系涉及到的人物、出版、地名。

书目中的关系大多可以归类到这六类中,但书目关系中有很多并未直接指出的两者间的相互关系,关联其与书目实体才能推出楚辞书目间隐含关联。

3 楚辞书目语义推理实现

3.1 楚辞书目本体

本文以学界权威姜亮夫著《楚辞书目五种》对楚辞书目版本的划分为基础,分析楚辞书目相关条目,结合本体模型五元组(类、关系、函数、公理和实例),完成楚辞书目本体类及其层次关系。

书目本体分两方面展开,一方面从普通用户信息查询入手,主要关注楚辞书目的书名、作者,主要内容,馆藏地等信息。另一方面从领域专家研究展开,主要有书目的版本,卷数,装帧,序跋摘要等。据此,如表1建立楚辞相关类属性。

表1 书目相关类与属性

名称	类/属性	语义描述	所属类别
Book	类	楚辞书目根节点	Thing
Bname	类	楚辞书名类	Book
Edition	类	书目的版本	Book
Writer	类	作者类	Book
Wname	类	字号,曾用名,别名	Writer
Address	类	书目馆藏地类	Edition
Date	类	刊刻时间	Edition
Preface	类	序	Edition
Postscript	类	跋	Edition
LetterColor	属性	四周线色	Edition
paper	属性	页数	Edition
Centre	属性	黑白口	Edition
bookbinding	属性	装帧	Edition
volume	属性	卷数	Edition
content	属性	内容	Edition
Paper	属性	纸张	Edition
Script	属性	字体	Edition
Carver	属性	刻工	Edition
Petime	属性	作序时间	Preface
Pttime	属性	作跋时间	Preface

Sameas(x, y)用来定义相等关系的规则。推理能够在诸类之间建立联系,并据此归类隐性关系。楚辞书目推理所需使用的部分定义如表4。

表4 楚辞书目推理中部分定义

规则	备注
Livingstone(?x)	人名实例
Placename(?x)	地名实例
Edition(?x)	版本实例
Date(?x)	日期(年代)实例
Productionname(?x)	作品实例
Oldusedname(?x, ?y)	曾用名关系
Birthin(?x, ?y)	出生地关系
Hasareference(?x, ?y)	参考文献
Iscreatorof(?x, ?y)	著作关系
editin(?x, ?y)	刊刻时代
Samedate(?x, ?y)	刊刻年代相同
iseditorof(?x, ?y)	版本编辑者
.....	

规则推理结果分为归类和断言。就书目推理结果,归类是规则根据相同属性,将两个或多个版本赋予同一版本属性,如当卷数相同,序跋相同,成书时期与作者相同,即确定其源流基本相同。这类结果通常需要多个元素共同作用以提高推理的准确性。断言类推理主要通过单一版本所具有的属性判断其是否某一已存在类。这两类结果都能够在楚辞书目不同的实例与实例,实例与类,类与类之间建立一种未曾设定的隐性的关系。

楚辞书目中版本关系、作品名等关系涉及的部分推理规则。

两个未知书目,具有相同的卷数和书目,时代,辑著者皆不同,判断其是否为同一源流,前后的时间关系;

$$Bname(?x) \wedge bname(?y) \wedge Volume(?x) \wedge volume(?y) \wedge$$

$$Writer(?x) \wedge Writer(?y) \wedge Edition(?x) \wedge edition(?y) \wedge$$

$$Editionname(?x, ?m) \wedge editionname(?y, ?n) \wedge$$

$$Iseditedof(?x, ?p) \wedge iseditedof(?y, ?q) \wedge$$

$$Sameas(?m, ?n) \wedge sameas(?p, ?q) \rightarrow$$

Samedition(x, y)

由书目间的次生关系,得到书目的同一作品多个版本或复制品之间的相互推理:

$$Productionname(?x) \wedge placename(?n) \wedge$$

$$Iscollectedat(?x, ?n) \wedge oldusedname(?x, ?y) \rightarrow$$

$$Iscollectedat(?y, ?n)$$

作品名关系推理规则:

$$Productionname(?x) \wedge Productionname(?y) \wedge$$

$$Productionname(?z) \wedge$$

$$Contain(?x, ?y) \wedge Contain(?y, ?z) \rightarrow$$

$$Contain(?x, ?z);$$

$$Productionname(?x) \wedge Productionname(?y) \wedge$$

$$Hasareference(?x, ?y) \wedge containcreatplace(?y, ?x) \rightarrow$$

$$placename(?x);$$

$$Productionname(?x) \wedge placename(?n) \wedge$$

$$Iscollectedat(?x, ?n) \wedge oldusedname(?x, ?y) \rightarrow$$

$$Iscollectedat(?y, ?n)$$

SWRL Editor^[6]编写面向楚辞版本、作品名、人名、地名、参考文献类的规则时可以直接引用OWL知识库中的类、属性与实例。楚辞书目推理规则编辑如图2。

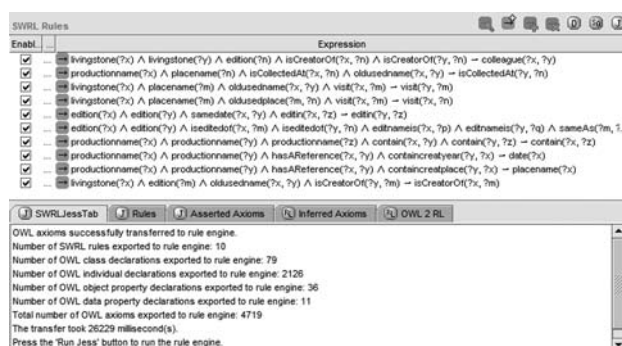


图2 楚辞书目推理规则及编辑界面

3.3 楚辞书目语义推理

JESS作为推理引擎是在转换过程中对规则本身语法并没有做改动,而只是将规则应用于存在的类、属性及实例。利用引擎将SWRL本体规则转换为JESS规则。此次完成转化的规则共有10条,类79个,实例2129个,公理4719个,对象属性36个,数据属性11个,如图3。运行JESS,推理得5319个结果,如图4。

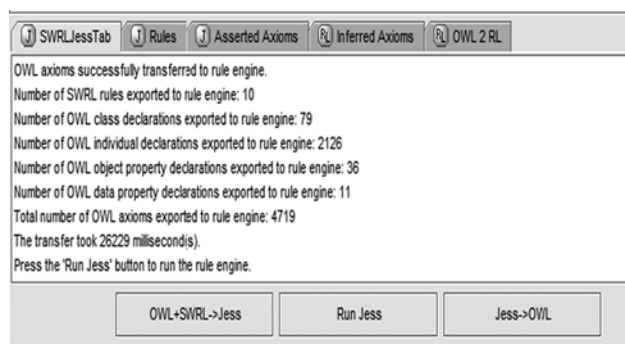


图3 楚辞书目规则由SWRL至JESS

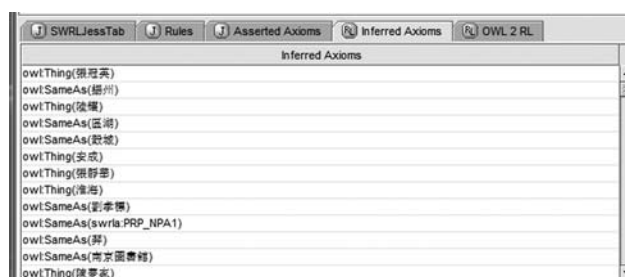


图4 楚辞书目推理结果

为实现对楚辞书目关系的完善,推理得到属性将会被添加到类中,同时将所新得关系和原本体添加至OWL库中。

3.4 推理结果验证

执行规则后,完成推理所产生新的属性会被添加到初始类中,通过与本体知识库中实例间的相互关联得到的隐含关系将被添加到本体知识库以完善楚辞书目本体中的描述关系。但是因为正确的逻辑并不意味着正确的结果,条件的改变会直接影响结果,所以需要验证推理一致性。

验证的结果通常分为两部分,一部分是指可以直接用于建构楚辞知识本体实例的数据,如楚辞书目的版本规则中,将相同卷数,相同书名,不同作者和不同序跋的书目作为同一源流进行归类,同时依照序跋的先后顺序进行排序。取南通大学楚辞研究中心所藏《楚辞章句》系列版本为例,中心现有《楚辞章句》夫容馆仿宋刻本。自日本所收集的宽延三年(1750年)江都书林前川六左卫门刊本,及文渊阁四库全书丁氏补抄本,名皆为《楚辞章句》,卷数同为十七卷,但宽延本为四册,前有王世贞等人所作序三篇。而四库全书本并无序,亦无册数信息。在执行已设定有条件,计算机标示此三个

版本同源。经相关专家核较,这三个版本确属同一系统。同时,据崔富章先生所著《楚辞书录解题》^[7]上册所述,此三版本确为同一版本。类似结论有如某楚辞刊本,缺书名,扉页有楚辞评林,吴郡八咏楼板等信息,推理后得此刊本为朱熹所撰《楚辞集注》。楚辞书目推理的实施能够辅助非专家对楚辞中信息控制。例如屈平与曾用名屈原之间有的等同关系,在屈原的属性中添加出生地等信息,推理规则实现后实例屈原中自动添加了与屈平相同的属性信息。

通过验证的结果使非专业领域人士亦可参与,在一定程度上降低了楚辞本体构建参与者的门槛,减少了手工添加次数,但是推理结果中亦存在不合理,如在楚辞书目本体的人名类的实例中,屈原和景差皆增辑过楚辞释文篇《大招》,由构建的推理规则在逻辑上可以推出是同僚关系。在已完善的本体知识库中可以查询到与人名类中实例屈原相关的隐含关系:和景差之间是同事关系、出生地是秭归,是《屈原赋二十五篇》的创作者等信息,这显然与学界常识有出入,也无法为知识库所接收,此类信息在删选时需要格外谨慎。

综上,虽然楚辞书目推理规则并不能完美的推演出楚辞本体知识库中的隐性信息。但随着构建的不断完善,楚辞书目规则能够方便完善楚辞本体中的书目关系,为楚辞关联检索的实现夯实基础同时,也为楚辞相关信息的挖掘提供新的途径。

4 结语

笔者对楚辞书目关系和本体知识库的分析研究,采用SWRL规则语言构建了楚辞书目推理规则并用JESS引擎实现了楚辞书目的语义推理。这有助于楚辞书目隐含关系的发现,使非领域专家在检索某一模糊概念的楚辞书目时,能查询到更多与关键字相关的书目,发现更多感兴趣的楚辞书目内容。

本文只选取了楚辞书目的版本类、作品名类、书目中涉及到的人名类做了主要的分析和说明。在事实的楚辞书目本体构建中,还涉及字体,纸张等主观性较强的信息,这些信息在一定程度上具有共性,如古典文献学中常会据此判断版本的年代,如明初继元版式多黑口赵体,嘉靖至万历年间皆白口方字仿宋,崇祯朝白口长字有讳;又如不同地域刊本的特点,如四川尊颜(体),两浙崇欧(体),福建学柳(体),江西兼有之。这类共性在实际转换为规则时适用性较差,而且存在很大程

度上的主观性。另,在古籍序跋中,存在序与刊刻时间不一致的情况,学界对序跋与刊刻的问题历来就存在争议,情况也较复杂,所以不能一概而论的就序跋信息去判断。所以在遇到这方面的问题时,笔者选择信而惑之,而非定而论之,仅简单罗列以供参考,未将它们转换为实际使用的规则。

针对楚辞专业领域中的书目规则构建,在一定程度上实现了关联搜索和精准定位,这为实现古籍书目内容的推理规则及隐含知识发现提供良好的开端。但是这仅仅是个开始,后期工作将在充分利用楚辞知识本体的基础上,细化领域和完善推理规则,寻找更为有效的隐性知识验证方法。在丰富楚辞爱好者或相关研究者的领域知识的同时,更好地为楚辞语义工作提供服务。

参考文献

- [1] 丁晟春, 江超男. 基于SWRL规则推理的隐含关系挖掘[J]. 现代图书情报技术, 2011(3): 68-72.
- [2] 黄勇奇. 基于地理本体和SWRL规则的农业地理信息时空查询研究[M]. 广州: 世界图书出版广东有限公司, 2012: 110.
- [3] Erner J. Friddman-Hill: Jess 6.1 manual[M]. Sandia national Laboratories, 2003.
- [4] 高红. 书目关系的综合研究[J]. 图书情报工作, 2006, 50(9): 108-112.
- [5] 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, 2009.
- [6] O' Connor M, Knublauch H, Tu S, et al. Supporting Rule System Interoperability on the Semantic Web with SWRL[C]// Lecture Notes in Computer Science. 2005: 974-986.
- [7] 崔富章. 楚辞书录解题[M]. 北京: 高等教育出版社, 2010.

作者简介

周澍琦, 1988年生, 南通大学古典文献学研究生, 研究方向: 先秦文献的语义化。
许富宏, 男, 南通大学文学院教授。

Research of the ChuCi Unconspicuous Knowledge based on Ontology

ZHOU ShuQi, XU FuHong

(College of Chinese Language and Literature, Nantong University, Nantong 226019, China)

Abstract: In Chuci knowledge, there are all kinds of relations, some relationships did not clarify directly in the literal but they are embedded, those need to add a certain amount of logical thinking to launch. The method provides reference for studying the intelligent reasoning of ancient books and the implicit knowledge discovery.

Keywords: Chu ci; Ancient Books; SWRL

(收稿日期: 2015-05-14; 编辑: 王立学)