

基于内容的学位论文本体库构建研究^{*}

——以赛珍珠《大地三部曲》专题研究学位论文为例

□卢章平 邵澍赞

摘要 针对高校图书馆学位论文数据库多数只提供面向文献特征、不能深入文献内容层面揭示的问题,提出构建基于内容的学位论文本体库。以赛珍珠《大地三部曲》专题研究学位论文为例,设计包含学位论文基本信息、学位论文研究对象和学位论文研究方向的《大地三部曲》专题学位论文本体库,对该本体库进行建模,以此为例,尝试在满足用户学位论文文献基本特征信息查询的同时,提供具体内容特征的检索点和知识关联、知识发现信息。

关键词 文献检索 学位论文 本体 知识发现

分类号 G250.76

DOI 10.16603/j.issn1002-1027.2017.03.009

1 引言

本体作为语义网的关键技术,在对资源的语义关联、资源关系的定义表达、资源信息的聚合扩展等方面有很大优势。我国学位论文的本体建设,目前已有第二军医大学图书馆的陆敏等人提出了一种基于本体的医学学位论文检索系统^[1],中国科学技术信息研究所的王莉也提出利用本体工具建模过程中应该遵循的原则、步骤以及需要注意的问题等^[2],然而这些大都是以学位论文的外部文献特征作为构建的主要内容,深入内容层面的学位论文本体库构建尚未见报导。本文采用本体技术,以赛珍珠《大地三部曲》专题研究学位论文为例进行探索,深入到内容层面设计高校图书馆学位论文知识库,将文献外部特征与研究对象、研究方向等内容特征放在一个整体中描述,利用概念间的类、属性联系,实例的设置与关联,构建既包含作者、题名等传统文献著录信息,又覆盖文献研究具体对象、文献所属研究方向信息的学位论文本体库。在用户使用模糊检索词时,自动实现对检索内容的系统聚类,提供多元选择,可以实现对系统数据库中同一研究方向文献全覆盖式查询,同时掌握单一文章中未能全面涉及的研究对

象信息,便于用户自主完成深层次的知识拓展。希望能为学位论文本体建设的深入拓展提供一点借鉴与启发。

2 以《大地三部曲》专题学位论文为例的本体库设计

以江苏大学图书馆国际赛珍珠文献资源中心“赛珍珠研究学位论文库”作为文献来源,选择赛珍珠《大地三部曲》研究硕、博士学位论文文献,提出了一个具体的专题学位论文本体库。将《大地三部曲》专题学位论文本体库的构建分为三个部分:①学位论文基本信息库;②学位论文研究对象库;③学位论文研究方向库。基本信息库主要方便用户了解学位论文文献的外部特征,例如作者信息、学位信息等;研究对象库主要针对文献内容,揭示学位论文研究的主要对象《大地三部曲》书目信息;研究方向库则对学位论文的具体研究方向信息进行了组织。

2.1 从现有的学位论文元数据方案到学位论文基本信息本体

知识本体是元数据方案的立体化,有哪些元数据元素决定着需要设计哪些类和属性^[3]。学位论文

^{*} 国家社会科学基金项目“‘互联网+’思维下面向万众创新的图书馆服务创新研究”(项目编号:16BTQ004)和江苏省社会科学基金项目“转型环境下图书馆社会职能拓展研究”(项目编号:15TQB011)的研究成果之一。

通讯作者:卢章平,ORCID:0000-002-4432-271X,lzping@ujss.edu.cn。

本体的设计需要参考现有的学位论文数据项,并且掌握这些数据项之间的关系。

在我国,目前使用较广泛的是 CALIS 高校学位论文数据库元数据规范。CALIS 高校学位论文数据库元数据规范是由 CALIS 全国工程文献中心项目管理组依照国家标准《书目信息交换用磁带格式(GB/T2901-92)》及该标准推荐执行的《中国公共交换格式(CCFC)》制定的,对数据格式进行了完整详细的定义。共包含 14 个必备元素,包括题名(Title)、作者(Creator)、主题(Subject)、附注(Description)、导师(Contributor)、日期(Date)、资源类型(Type)、资源格式(Format)、资源标识(Identifier)、语种(Language)、相关文献(Relation)、权限管理(Rights Manage)12 个各类资源通用核心元素和学位(Degree)、馆藏信息(Location)2 个学位论文个别元素。

在实际的论文资源组织工作中,各高校主要应用基于 CALIS 的学位论文提交系统来实现学位论文数据库的建设。目前我国高校采用的论文提交系统主要有北京拓尔思信息技术有限公司(TRS)、清华同方光盘股份有限公司(TPI)、杭州麦达电子有限公司(IDL)和北大方正电子有限公司等研发的 4 套系统。江苏大学图书馆使用的是方正 TASI 论文授权提交系统。该系统定义学位论文元数据信息项包括姓名、学号、培养单位、学科、专业、入学年份、答辩日期、授予学位、导师、论文题名、外文论文题名、论文关键字、外文论文关键字、论文摘要、外文论文摘要、论文总页码、参考文献总数、密级、中图分类号等。

在设计本体的过程中,首先需要考虑对已有本体的复用,在已有本体难以表达具体领域的数据及关系时,进行自定义,参考 DC 元数据标准及 CALIS 高校学位论文数据库元数据规范。结合方正 TASI 论文授权提交系统元数据信息项,本文定义了学位论文基本信息本体的类及数据属性,表 1 为学位论文元数据元素与学位论文基本信息本体中类和属性的对应。

定义题名、作者、学位、导师、主题、日期、关联文献、语种、资源标识、资源格式、馆藏信息类,根据《大地三部曲》学位论文文献个性特征省略资源描述、资源类型、权限管理项,同时去除冗余元素姓名汉语拼音、导师姓名汉语拼音、导师号、查询号,并将 Email

统一为中文表达方式电子邮件地址。在主题类下再划分子类关键词和摘要,日期类下设子类入学年份、提交日期、答辩日期和可获得日期,资源格式类下划分论文总页码、插入图表数子类,资源标识类下设中图分类号、UDC 分类号,馆藏信息类下划分馆藏号、馆藏地点。

本体的属性包括表示概念内在属性的数据属性和表示概念间关系的关系属性,本文定义关系属性写作(Write)与被写作(be Written by),指导(Direct)与被指导(be Directed by),引用(Quote),获得(Receive),时间(Time)。如题名与作者通过写作与被写作相关,作者与导师通过指导与被指导相关,题名与关联文献通过引用相关,作者与学位通过获得相关。其中写作与被写作,指导与被指导是互逆属性。

表 1 学位论文元数据与学位论文基本信息本体的对应关系

学位论文元数据		学位论文本体	
元数据项	元数据元素	类	属性
题名	中文题名	题名 Title	中文题名 外文题名
	外文题名		
创作者信息	姓名	作者 Creator	姓名 学号 专业 学校 电话 电子邮件地址
	姓名汉语拼音		
	学号		
	专业		
	培养单位		
	电话		
	E-mail		
学位信息	级别	学位 Degree	学位名称 学科名称 学籍
	学位名称		
	学科名称		
导师	导师姓名	导师 Contributor	导师姓名 导师单位
	导师姓名汉语拼音		
	导师单位		
	导师号		
主题	关键词	主题	主题
	摘要		
资源描述	目录	资源描述	资源描述
	成果目录		
	资助		
	附录		



日期	入学年份	主题 Subject	公元纪年
	答辩日期		
	提交日期	日期 Date	
	可获得日期		
资源类型	学位论文		
资源格式	论文总页码	关联文献 Relation	写作者 题目 来源 出版时间
	参考文献总数		
	插图数		
	插表数		
资源标识	中图分类号	语种 Language	语种名称 语种代码
	UDC 分类号		
	查询号		
语种	记录语种		
关联	组成部分	资源标识 Identifier	
	版本关联		
	参照		
	被参照		
权限管理	保密年限	资源格式 Format	
	访问权限		
	使用权限	馆藏信息	
馆藏信息	馆藏号	Location	

单篇学位论文是一个集成的概念,用户在寻找学位论文信息时不仅希望获取文献实体信息,更大的需求是挖掘内容信息。由于原有的学位论文元数据方案对资源内容层面的挖掘还不够深入,因此笔者提出构建基于文献内容的“研究对象”和“研究方向”本体,希望实现文献外部信息与内容信息知识关联,通过推理完成知识发现。研究对象为《大地三部曲》书目内容,同时利用文献调查与内容分析,结合专家咨询法等对研究方向进行了定义。

2.2 学位论文研究对象《大地三部曲》本体设计

选取的学位论文文献的研究对象为《大地三部曲》书目,《大地三部曲》是美国著名旅华作家赛珍珠荣膺诺贝尔文学奖的代表作,被瑞典皇家学院誉为“对中国农村生活”具有“史诗般描述”的作品。小说塑造了一系列勤劳朴实的中国农民形象,生动地描绘了他们的家庭生活,向西方读者展示神秘而历史悠久的中国文化,内容翔实,同时蕴含了大量历史细节。

现有的开放性本体例如 DBpedia、freebase 等大多来源于国外,国内的大部分本体都参照国外相关

系统构建,可参考性不大,因此《大地三部曲》本体的构建只能在相关研究启发下,利用骨架法^[4]、评价法、七步法^[5]等本体构建方法进行重新开发。在构建过程中首先需要从学位论文文献中提取研究对象的核心知识框架,了解研究对象《大地三部曲》书目的基本组织模式,利用抽象的类概念表现核心知识对象,并通过数据属性和关系属性来描述其特征与关联,梳理书目知识脉络,实现书目内容间的语义关联,在此基础上揭示研究对象的整体内容。

《大地三部曲》相关学位论文文献中主要涉及对书目中人物、事件、时间、地理等内容研究。目前相关的知识本体可参考的有描述人与人之间关系的本体(Relationship Ontology)^[6],事件本体(Event Ontology)^[7]、时间本体(Time Ontology)^[8]、地理本体(GeoNames Ontology)^[9],还有描述个人信息和其相关社会网络的 FOAF(Friend-Of-A-Friend)^[10]本体等,同时由于研究对象《大地三部曲》为历史类长篇小说,可参考历史领域相关本体。在这些本体的基础上,结合学位论文文献中涉及的《大地三部曲》书目内容知识的具体个性特征进行建模,首先定义了人物、事件、时间、地点四个基本类。其中人物类为主体类。人物类包含主要人物和涉及人物,其中涉及人物子类作为主要人物类的补充类,指与主要人物相关的人物集合。其次根据事件、地点、时间与人物之间的关联与描述信息确定类的属性。以人物类为主体类,添加人物的数据属性,包括姓名、性别、身份、职业、籍贯等,为展示事件的概况,定义了主题、概要的属性;由于小说文本表达方式的特点,在时间的数据属性设置上包含公历民国纪年以及以人物经历标记的人物纪年。同时定义关系属性相关人物(Related Characters)、发生时间(at Time)、发生地(in Place)、参与事件(Participate in)等,如人物与事件通过参与事件相关,与时间通过出生年月、死亡年月相关,事件与地点通过发生地相关。图 1 为研究对象《大地三部曲》本体的概念属性设置。

根据人物、事件等涉及的内容信息划分层次关系,具体概念层次结构如图 2 所示。

《大地三部曲》系列包括《大地》《儿子》《分家》三部,研究文献涉及家族三代人物形象,因此定义三级子类第一代父辈、第二代子辈和第三代孙辈。根据人物生平经历概括将事件类分为婚姻、事业、学业,同时结合《大地三部曲》贯穿全系列的主题“土地”定

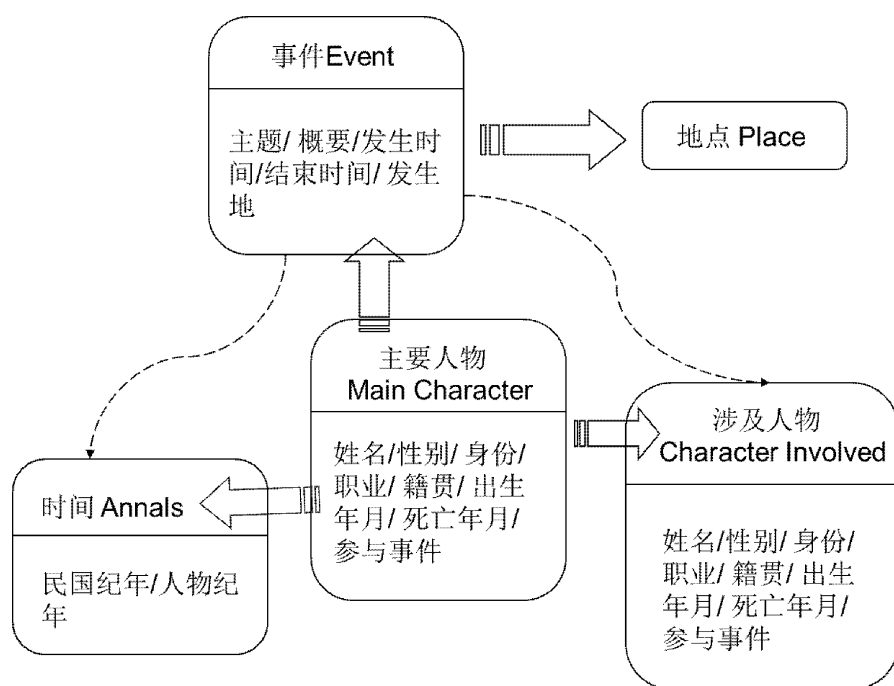


图1 概念的属性

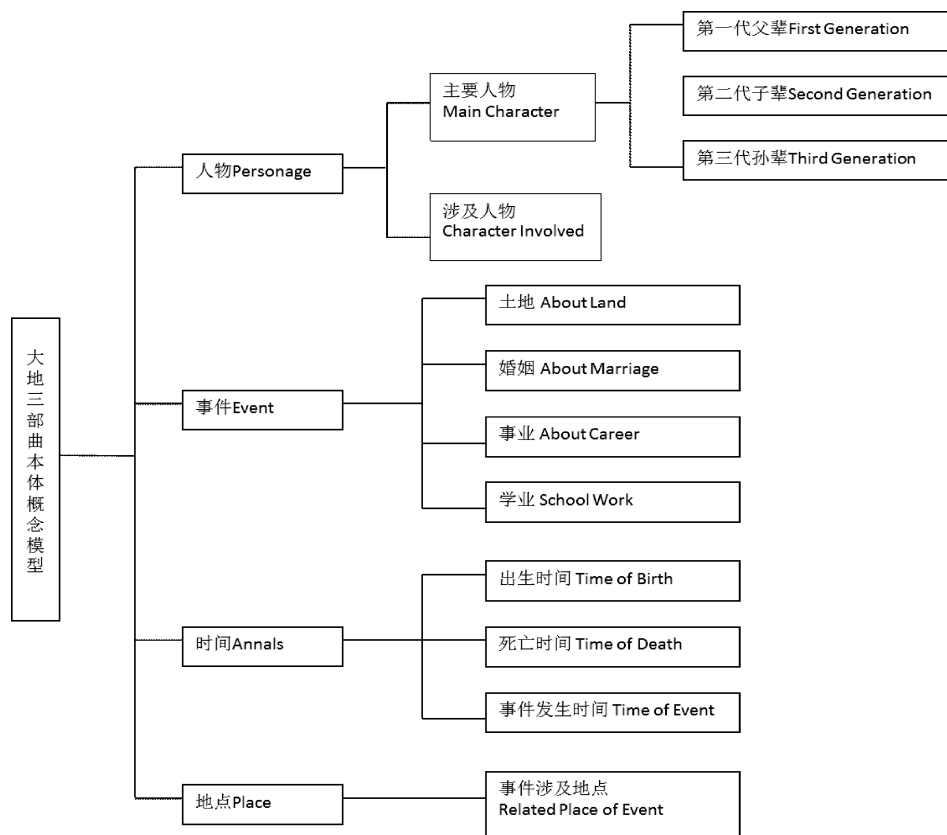


图2 《大地三部曲》本体概念模型

义该本体特有类土地;将时间类分为出生时间、死亡时间、事件发生时间;文学作品的特性决定了地点与事件之间的关联较为密切,设置子类事件涉及地点

便于实现与事件相关的地点信息的整合。

2.3 学位论文研究方向本体设计

依据文献调查及内容分析的方法,对选择的学

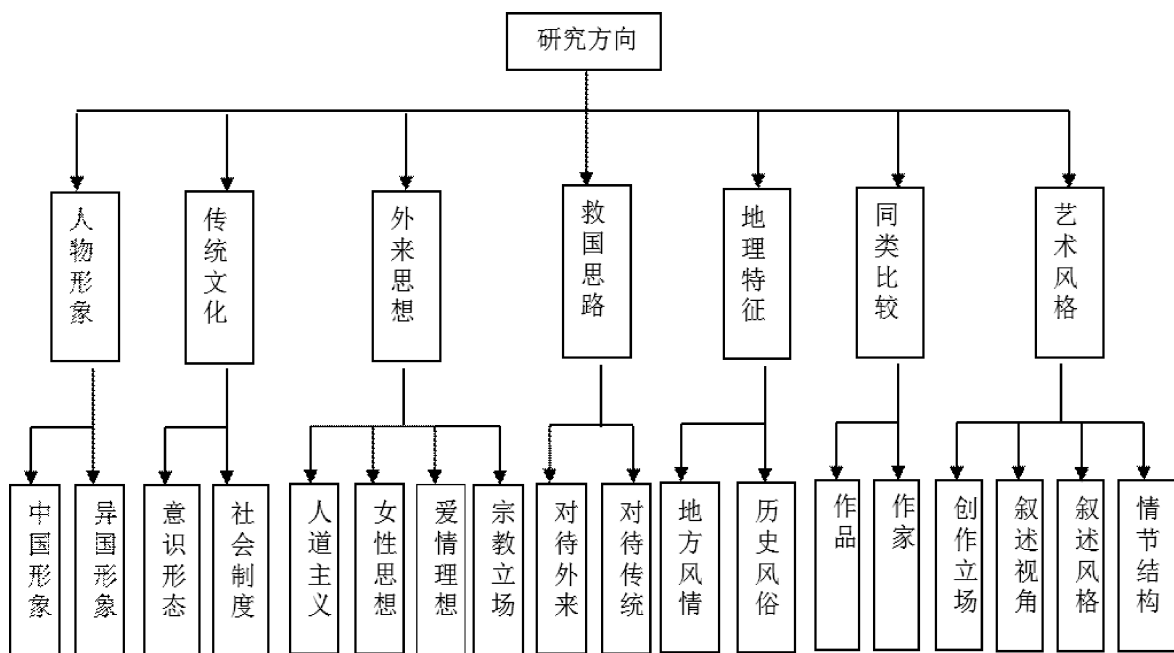


图3 研究方向本体概念层次

位论文文献进行知识分析和抽象,在相关领域专家的帮助下进行了《大地三部曲》相关学位论文研究方向本体的设计,对知识节点的确定是对学位论文文献内容提取的过程,随着研究文献的不断增多,研究内容不断地更新和增加,可结合实际需要对概念层次的划分进行修正与补充。目前初步确定人物形象、传统文化、外来思想、救国思路、地理特征、同类比较、艺术风格七大类。下划18个子类,具体见图3研究方向本体概念层次结构。

对于类的设置,人物形象特指《大地三部曲》中涉及的一系列人物设定,具体涉及到中国形象和异国形象。传统文化主要指涉及的中华传统民族特质和风貌,各种思想文化、观念形态的总体表征,具体涉及到意识形态和社会制度。由于目标书目中设定的时代背景是中国农业社会受到西方外来思想冲击的时期,部分学位论文文献对一些新兴的外来思潮进行了重点探究,因此设置外来思想类,具体涉及人道主义、女性思想、爱情理想和宗教立场。救国思路指分析先进的知识分子在混乱的大变革时代下对救国救民道路的探索,分为对待外来及对待传统两部分。部分文献对《大地三部曲》书目中涉及的中国地域特色和传统民俗进行了介绍,因此设置地理特征类,具体涉及到地方风情和历史风俗。同类比较指将《大地三部曲》书目与同类型的文学作品或将美籍旅华作家赛珍珠与具有中西文化背景的作家进行比

较研究。具体分为作品和作家两部分。艺术风格指研究艺术创作呈现出的一种独特个性,具体涉及到创作立场、叙述视角、叙述风格、情节结构。

其次,定义多个关系属性:描述(Describe),被描述(be Described by),展示(Show),被展示(be Showed by),冲突(Conflict),影响(Affect),被影响(be Affected by),比较(Compare),关联(Related)。

(1)描述,逆属性为被描述,用于艺术风格与人物形象、传统文化、救国思路、地理特征、外来思想之间。

(2)展示,逆属性为被展示,用于人物形象与传统文化、救国思路、外来思想之间。

(3)冲突,用于传统文化与外来思想之间。

(4)影响,逆属性为被影响,用于传统文化与救国思路、地理特征,外来思想与救国思路,人物形象与地理特征之间。

(5)比较,用于同类比较与艺术风格、人物形象之间。

(6)关联,表示两者间具有相关关系。

2.4 《大地三部曲》专题学位论文本体库建模

选择利用Protégé本体构建工具尝试建模,将学位论文基本信息本体、学位论文研究对象《大地三部曲》本体、学位论文研究方向本体三者进行知识关联,得到基于内容的赛珍珠《大地三部曲》研究学位论文本体库模型(图4)。定义“研究对象(Research

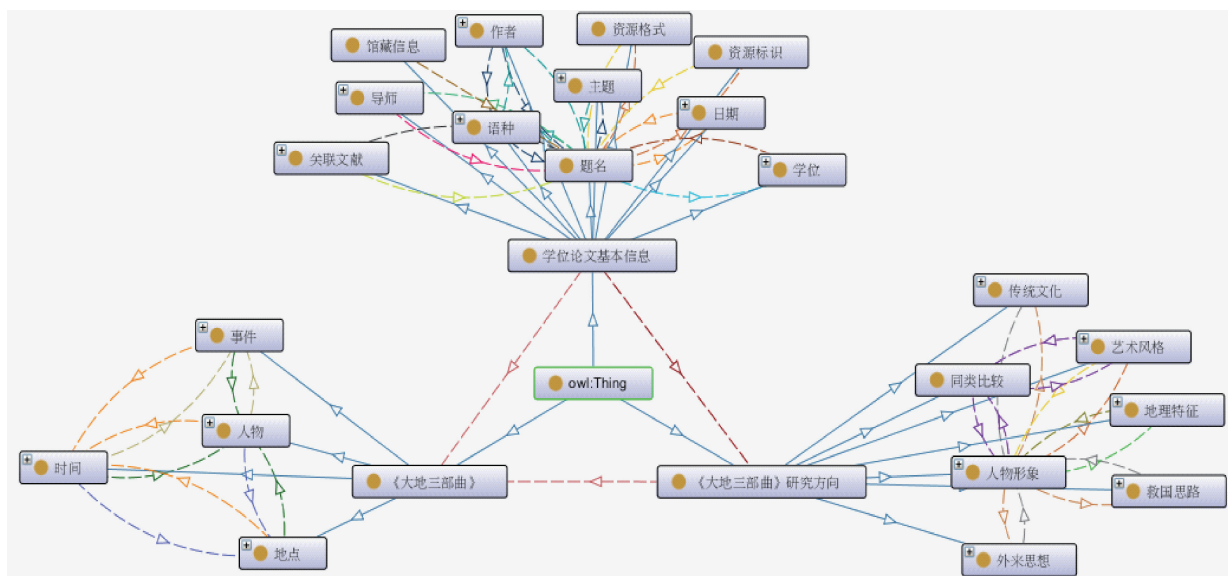


图4 《大地三部曲》专题学位论文本体库模型

Object)”“研究方向 (Research Direction)”属性,学位论文基本信息与《大地三部曲》通过研究对象相关联,与《大地三部曲》研究方向通过研究方向相关联,《大地三部曲》与《大地三部曲》研究方向通过研究对象相关联。通过这种模型可以将学位论文基本信息及内容资源整合在一起,实现一体化的组织与管理,便于有效的知识发现。

3 本体实例化

实例是类的具体化,完整的本体是由类、属性和实例共同构成的。每一个类都对应着很多不同种类的实例,对实例的设计是构建本体库的一项重要内容,实例化的质量直接关系到知识检索的效率。

文章从江苏大学“赛珍珠研究学位论文库”《大地三部曲》研究硕、博士学位论文文献中提取基础实例元数据,将文本挖掘技术与人工修订相结合,围绕文本的核心知识内容确立对应本体的实例属性及关系。整理文本表达别称词表,使用自动标注技术完成语义标注,实现对多篇文献中知识对象的提取,定义赛珍珠《大地三部曲》研究文献的特定识别规则,自动抽取学位论文作者、学位、书目相关人物、事件、时间、地点等信息。结合关系抽取技术提取相关语句,发现关联事实,同时根据《大地三部曲》专题学位论文本体库中类、数据及关系属性的设置制定关系抽取规则“n-v-n (名-动-名)”三元组,如“作者-写作-题名”“人物-参与-事件”“事件-发生-地点”等,并结合文本语句分析文本中显性和隐性

的语义关系,将其作为构建实例的基础。

利用机器自动处理可以减少人工构建的工作量,节省了构建时间,同时帮助发现一些隐含知识。但自然语言的复杂性限制了文本挖掘的准确度,因此在本体库的构建过程中需要与人工修订相结合。相关领域专家与研究人员在进行审核、筛选与知识补充的基础上,最终确定了102个题名与作者,23个人物,496个事件,8011个关联文献等规范化的实例数据,添加至赛珍珠《大地三部曲》研究学位论文本体库。

由于涉及实例较多,篇幅有限,在此仅以其中的《大地三部曲》专题硕士学位论文《〈大地三部曲〉人物组群与作品旨归研究》为例。设计学位论文基本信息的实例:添加题名实例“《〈大地三部曲〉人物组群与作品旨归研究》”,作者实例“陈磊”,学位实例“硕士学位”,导师实例“阮航”,关键词实例“赛珍珠;《大地三部曲》;人物组群;启蒙”,入学年份实例“2004年”,提交日期实例“2007年”等。设计学位论文研究对象《大地三部曲》的实例:添加内容相关人物实例“王龙”“王虎”“王源”等,事件实例“美国留学”“学校教书”等,地点实例“海滨城市”“老家祖屋”等,时间实例“少年时期”等。设计学位论文研究方向的实例:添加中国形象实例“家族世代”“知识分子”“男性形象”“女性形象”“他者形象”“普通民众”等,异国形象“宗教人员”“外国教师”“外国友朋”等。并对具体实例的属性进行赋值,如学位论文基本信息中作者“陈磊”实例:姓名“陈磊”,专业“比较文学

与世界文学”,学校“西南交通大学”。学位论文研究对象《大地三部曲》人物“王源”实例:姓名“王源”,性别“男”,身份“王家第三代代表人物”,职业“教师”。

4 基于 Protégé 的《大地三部曲》专题学位论文本体库实现

Protégé 软件是斯坦福大学基于 Java 语言开发的本体编辑和知识获取软件,属于开放源代码软件,是语义网中本体构建的核心开发工具。利用 Protégé 能有效完成对《大地三部曲》专题学位论文本体库的形式化表达,同时实现知识检索的可视化,建立知识导航图谱,通过例间关系网络实现知识的推理与发现。

4.1 OWL 描述语言

构建本体的语言是计算机可读的形式化语言,现在可使用的本体语言已经有很多,如 RDF、DAML、KIF、Ontolingua 等等,伴随语义网的发展,OWL 描述语言被广泛应用。OWL 的编写形式是 XML,在各种类型的操作系统间兼容性较强。在 Protégé 中实现了《大地三部曲》专题学位论文本体库,可以获得本体库的 OWL 文件,以下是 OWL 文件的片段:

```
<Prefix name="owl" IRI="http://www.w3.org/2002/07/owl#" />
<Prefix name="rdf" IRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#" />
<Prefix name="xml" IRI="http://www.w3.org/XML/1998/namespace" />
<Prefix name="xsd" IRI="http://www.w3.org/2001/XMLSchema#" />
<Prefix name="rdfs" IRI="http://www.w3.org/2000/01/rdf-schema#" />
<Declaration>
  <ObjectProperty IRI="#被描述"/>
</Declaration>
<Declaration>
  <NamedIndividual IRI="#《论赛珍珠笔下中西合璧的男性形象》"/>
</Declaration>
<Declaration>
  <NamedIndividual IRI="#学校教书"/>
</Declaration>
<Declaration>
  <NamedIndividual IRI="#知识分子"/>
```

```
</Declaration>
<Declaration>
  <NamedIndividual IRI="#《大地上的寻根者》"/>
</Declaration>
<Declaration>
  <ObjectProperty IRI="#被影响"/>
.....
```

4.2 可视化知识图谱

目前,Protégé 软件常用的有 OWLViz 和 OntoGraf 两种可视化插件。与 OWLViz 相比,OntoGraf 呈现效果更加的丰富,能够展示类之间的层次关系,同时还能显示类的实例及实例的属性。两者之间的关系用连线表示,不同颜色的连线代表的关系不同,通过连线的颜色和数量可以判断例间关系的复杂度。

OntoGraf 插件可实现检索功能,本文使用 Protégé 5.0 版本,检索条件包括:Contains(包含)、Start With(前词匹配)、End With(后词匹配)、Exact Match(精确匹配)和 Regexp(正则表达式)^[11]。根据需求限制检索条件,在检索框里输入检索词,点击“Search”进行检索,可以获取可视化知识图谱。以下做部分举例说明。

(1)学位论文题名检索。检索学位论文题名可以查阅到该学位论文的著录信息,同时可以了解该文献的研究对象及研究方向等内容信息。

图 5 为学位论文《〈大地三部曲〉人物组群与作品旨归》实例检索 OntoGraf 可视化图谱。由图可见,通过对《大地三部曲》专题学位论文本体库的检索,既可以查阅到该学位论文的著录信息,如作者陈磊,学位文学硕士等,又可以了解内容信息,如论文研究对象“王源”“王龙”、论文研究的具体方向“男性形象”“知识分子”等。图中《〈大地三部曲〉人物组群与作品旨归》与事件“学校教书”“与爱琳恋爱”,时间“少年时期”,地点“海滨城市”等是研究对象的关系,与“他者形象”“家族世代”是研究方向的关系。而研究方向“男性形象”的研究对象又与人物“王源”相关,人物“王源”与人物“王虎”则存在父子关系。

(2)研究对象检索。检索单一研究对象可以查询到与该对象相关的文献及针对该对象已有的研究方向。

图 6 为研究对象《大地三部曲》人物“王源”实例检索 OntoGraf 可视化图谱。可以看到,人物“王源”

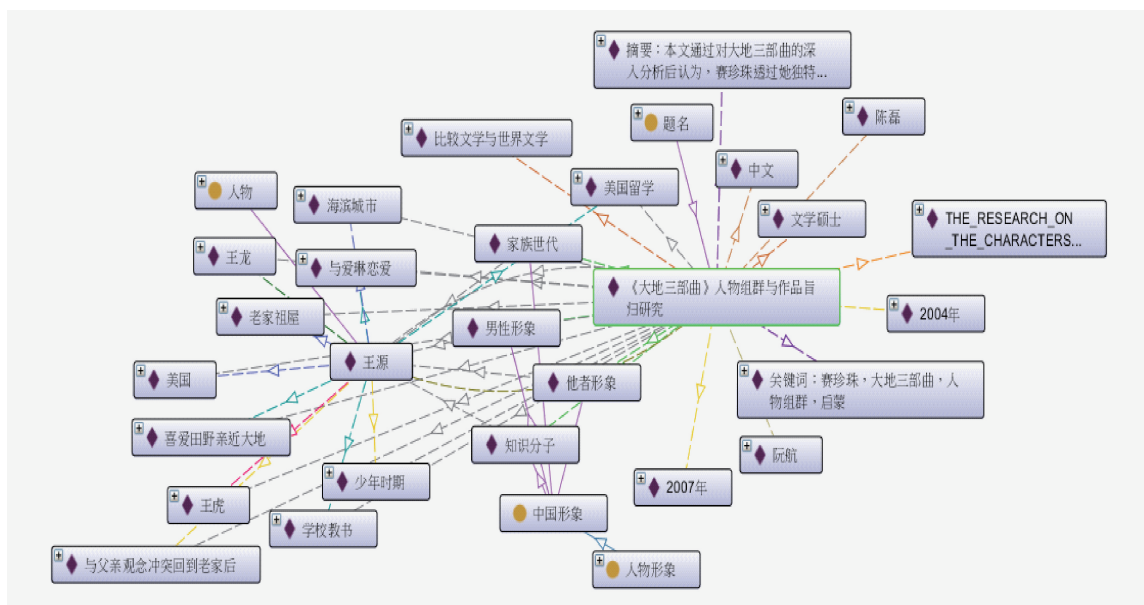


图5 《《大地三部曲》人物组群与作品旨归》检索可视化图谱

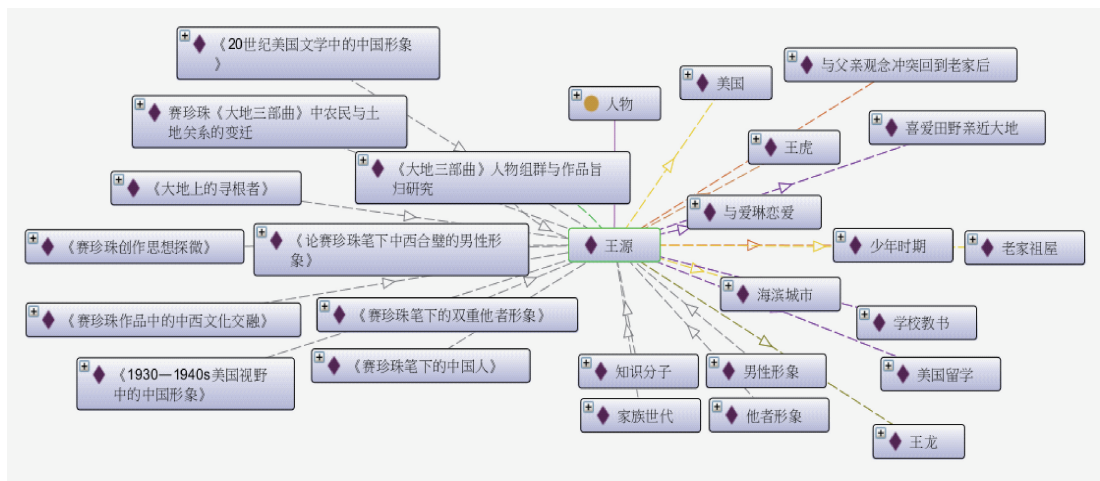


图6 研究对象人物“王源”检索可视化图谱

相关的研究学位论文文献有《论赛珍珠笔下中西合璧的男性形象》《赛珍珠笔下的中国人》等,涉及到的研究方向有知识分子、男性形象等。同时“王源”相关的事件有“美国留学”,相关的人物有“王虎”等。

(3)研究方向检索。检索某研究方向可以查询到该方向的研究文献及该研究方向涉及到的具体研究对象。

图7为研究方向“知识分子”实例检索OntoGraf可视化图谱。可以看到,研究方向“知识分子”的研究对象有人物“王源”,该方向相关的研究学位论文文献有《《大地三部曲》人物组群与作品旨归》《论赛珍珠笔下中西合璧的男性形象》等。

4.3 推理与知识发现

推理与知识发现是本体真正价值的体现。通过推理可以实现信息的聚合,完成深层次的知识发现和逻辑的验证。针对本体的推理,以OWL等标准的本体语言为基础,一些机构开发了RACER、HermiT、Pellet、FaCT等推理机。推理机是实现知识推理与发现的重要部件,是在计算机系统中完成推理与控制的重要组成部分。本文使用的Protégé5.0中提供了基于HermiT1.3.8.413推理机的推理功能。

根据定义的本体逻辑关系进行相关概念间的推理,结合知识发现需求,可以自主选择推理机的呈现结果,获取想要的知识内容,通过Direct Superclasses

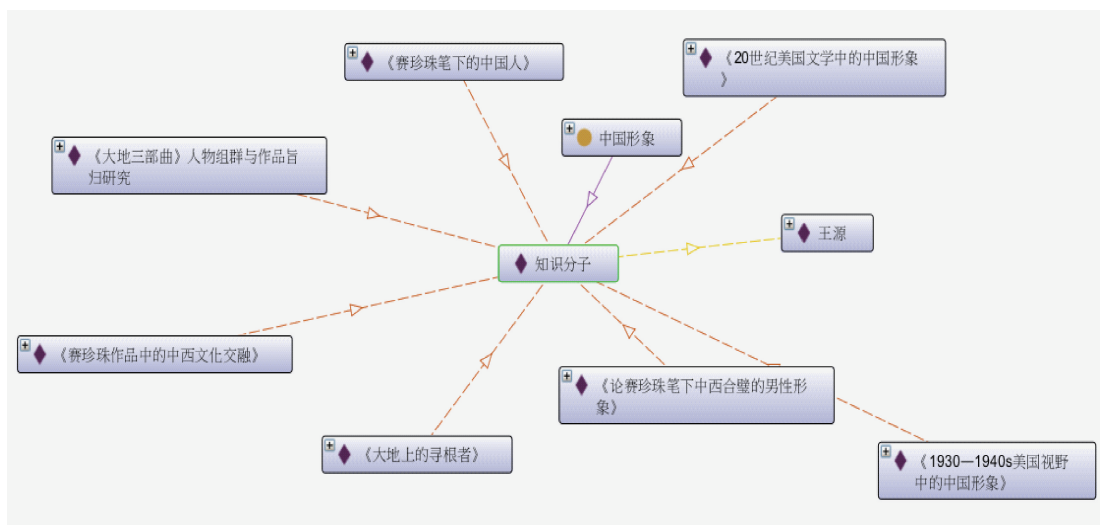


图7 研究方向“知识分子”检索可视化图谱

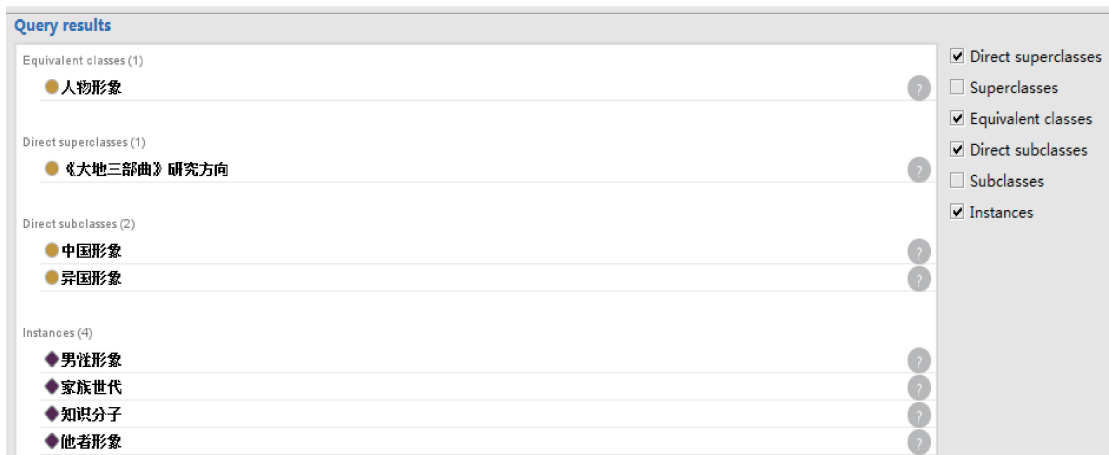


图8 研究方向“人物形象”推理界面

发现上级父类, Superclasses 发现全部父类, Equivalent Classes 发现相同类, Direct Subclasses 发现下级子类, Subclasses 发现全部子类, Instances 发现类的实例等, 从而进一步定制个性化的知识服务。本节仅以学位论文研究方向“人物形象”为例作相关说明。

图8为学位论文研究方向“人物形象”查询推理界面。在Query查询页面对研究方向“人物形象”进行查询, 使用Start Reasoner执行推理操作, 可获取推理结果。确定目标查询内容, 在完成对“人物形象”的知识推理后, 可以发现, “人物形象”是从属于《大地三部曲》研究方向的一个子类, 同时, 它包括下级子类中国形象和异国形象。还可以发现学位论文中针对该研究方向的具体研究内容包括实例男性形象、家族世代、知识分子等。

在完成推理后, 通过OWL Viz本体结构图可以

明确地显示概念之间的关联, 使用推理机制实现对本体逻辑性的验证。对比自定义与推理后的层次结构模型, 可以检验本体设计是否逻辑无误, 从而达到对本体的修正和完善。图9为推理机制下的学位论文研究方向“人物形象”与中国形象、外来思想等在《大地三部曲》研究方向本体中的关系结构。由图可见, “人物形象”与《大地三部曲》研究方向, 中国形象、异国形象是逻辑上的从属关系, 与外来思想、艺术风格、地理特征、传统文化等是逻辑上的并列关系, 与自定义的层次结构一致, 因此可以判断该本体的设计是逻辑合理的。

5 结语

以赛珍珠《大地三部曲》专题研究学位论文为例, 探索深入内容层面的学位论文本体库构建, 设计

解决了以往高校图书馆一般按照时间、作者等属性提供知识服务所导致的信息资源之间属性关联与语义关联缺失,内容揭示不够深入的问题,利用本体技术建立知识导航图谱,图谱中的节点、连线等都具有一定的属性含义,通过知识节点间的关联方便检索,用户在使用时可以根据自己的需求进行知识的获取与选择,利用推理机制进行知识发现,结合自身特点定制个性化的知识图谱,便捷地获取个性化的知识服务,最终实现对馆藏学位论文资源的全面挖掘与利用。

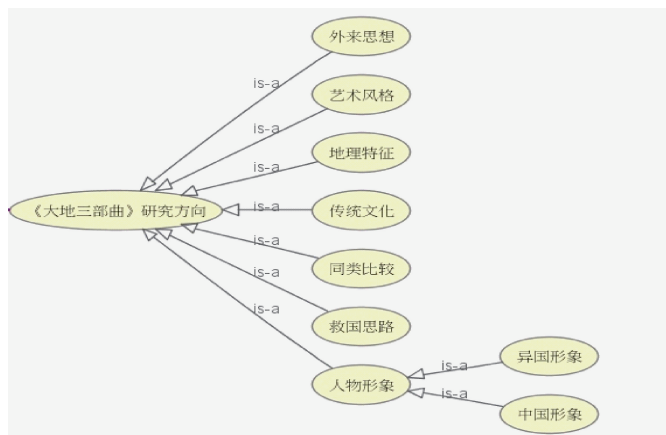


图9 研究方向“人物形象”相关概念之间的逻辑关系

- 王莉. 基于 Protégé 的本体建模方法研究——以学位论文服务系统为例[J]. 现代图书情报技术, 2006(10):55-59.
- 夏翠娟,刘炜,张磊,朱雯晶. 基于书目框架(BIBFRAME)的家谱本体设计[J]. 图书馆论坛, 2014(11):5-19.
- Uschold M, King M. Towards a methodology for building ontologies[C]. Workshop on Basic Ontological Issues in Knowledge Sharing, held in conjunction with IJCAI-95, Montreal, Canada, 1995.
- Natalya F. Noy, Deborah L. McGuinness. Development 101: a guide to creating your first ontology[EB/OL]. [2016-07-11]. <http://www.cin.ufpe.br/~in1116/aulas/ontology101.pdf>.
- Ian Davis, Eric Vitiello Jr. Relationship: a vocabulary for describing relationships between people[EB/OL]. [2016-07-20]. <http://vocab.org/relationship/>.html.
- Yves Raimond, Samer Abdallah. The event ontology [EB/OL]. 2007-10-25. [2016-07-20]. <http://mootools.sourceforge.net/event/event.html>.
- Jerry R. Hobbs, Feng Pan. Time ontology in OWL [EB/OL]. 2006-09-27. [2016-07-09]. <http://www.w3.org/TR/owl-time/>.
- GeoNamesTeam. GeoNamesOntology [EB/OL]. [2016-07-25]. <http://www.geonames.org/ontology/documentation.html>.
- Dan Brickley, Libby Miller. FOAF vocabulary specification 0.99[EB/OL]. [2016-07-25]. <http://xmlns.com/foaf/spec/>.
- 耿萍萍. 基于本体的图书情报学知识地图的构建研究[硕士学位论文][D]. 河北:河北大学, 2015.

作者单位:江苏大学科技信息研究所,镇江,212013

收稿日期:2016年8月24日

参考文献

- 陆敏,杨发毅. 基于本体的医学学位论文检索系统的开发与应用[J]. 中华医学图书情报杂志, 2009(2):64-66.

Research on Construction of Dissertation Ontology Library Based on the Literature Content

—Taking the Dissertation about *The Good Earth* as Example

Lu Zhangping Shao Shuyun

Abstract: Most of the existing dissertation databases in University libraries only provide features search and can not reveal the contents of the literature, this paper studies on how to resolve these problems by constructing a dissertation ontology library based on the literature content. Using the dissertation about *The Good Earth* as an example, It mainly discusses the design of the dissertation ontology library of *The Good Earth*, which contains the basic informations, the research objects and directions of dissertation and try to provide the comprehensive knowledge connection and knowledge discovery.

Keywords: Literature Retrieval; Dissertation; Ontology; Knowledge Discovery