传统民俗文化知识的领域本体构建研究

李庆华
岭南师范学院　广东湛江　524048

【摘　要】目的：作为国家传统文化宝库重要组成部分的传统民族民俗文化，伴随着自然环境、社会环境以及人文环境的变迁，民俗传承和保存手段的局限，许多传统民俗技艺面临流变和消失，民俗传统文化的保护和传承面临日益严峻的挑战；办法：本文以本体构建理论为基础，以畲族传统民俗为研究对象，分析了畲族传统民俗文化知识的组织和管理，从民俗事件的发生时间、发生地点、表现形式、民俗类型、参与族群以及传承演变等六方面确定了畲族传统民俗知识的核心概念及层次关系，并在protégé平台设计并实现了畲族民俗本体及其实例；结果：基于畲族民俗本体及其属性关系知识库，并利用互联网的民俗信息资源作为语料库，采用基于本体覆盖度的方法对本文的本体概念进行了综合评价，评价结果表明该本体性能较好，达到了本体构建的要求；结论：民俗本体的构建可为传统民俗数字化提供新的方法和手段，提升民俗数字化的效率，有利于传统民族民俗文化的保护和传承。

【关键词】畲族；传统民俗；protégé；本体构建；本体评价

Research on the domain ontology construction of traditional folk culture knowledge

Li Qing-hua
Lingnan Normal University Zhanjiang，Guangdong　524048

【Abstract】Objective：As an important part of the national traditional culture treasure house，the traditional national folk culture，With the changes of the natural environment，social environment and human environment，The limitations of the means of folk-custom inheritance and preservation，Many traditional folk techniques face fluid change and disappearance，The protection and inheritance of folk traditional culture are facing increasingly severe challenges；Method：This paper is based on the ontology construction theory，Taking the She nationality traditional folk customs as the research object，Analyzed the organization and management of the knowledge of She traditional folk culture，The core concept and hierarchical relationship of she traditional folk custom knowledge are determined from six aspects：the occurrence time，occurrence place，form of expression，folk custom type，participating ethnic groups and inheritance and evolution of folk custom events，And designed and realized the she folk custom ontology and its examples on the protege platform；Results：Based on the knowledge base of she folk custom ontology and its attributes，And use the folk custom information resources of the Internet as a corpus，A comprehensive evaluation of the ontology concept in this paper using an ontology coverage-based approach，The evaluation results show that the ontology performs better，Achieve the requirements of ontology construction；Conclusion：The construction of folk custom ontology can provide new methods and means for the digitization of traditional folk custom，Improve the efficiency of folk custom digitization，It is conducive to the protection and inheritance of traditional ethnic folk culture.

【Key words】She nationality；traditional folk customs；protege；ontology construction；ontology evaluation

一、引言

随着世界经济的一体化和全球化，科学技术特别是传媒与网络技术的迅速发展，使人们对传统文化的发展投入了极大的关注，越来越多的国家和民族认识到传统文化对于国家软实力的重要作用。作为国家传统文化重要组成部分的少数民族传统文化，在和工业 4.0 和互联网+时代的邂逅过程中既有“爱”的火花，同时也出现不和谐的生态尴尬。一方面，随着全球化趋势的加剧，传统知识资源日益消减；另一方面，出于对文化多样性的价值认同，传统民俗文化又被作为一种遗产资源被公众消费，“越是民族的就越是世界的”，其独有的文化价值和不断显露的经济价值使其在国家社会经济发

展中的作用也愈发凸显[1]。然而，随着自然环境、社会环境以及人文环境的变迁，人类从农耕文化向工业文明再向 5G 社会的转型过程中，这些以部分族群民众口传身受的传统文化由于受到保存手段和传播方式的局限，使得许多传统民俗技艺面临流变和消失，民俗传统文化的保护和传承面临日益严峻的挑战[2]。

“传统”并不意味着守旧，传统民俗不排斥创新和发展，在浩瀚的历史长河中没有一成不变的传统民俗。当传统民俗邂逅现代信息技术，“爱”的火花一直在迸发，传统民俗数字化一直在进行。从早期的对民俗技艺的录音、录像到民俗数据的编码和数据库存储以及到近年来的民俗大数据系统的应用等，这些方法和应用都为传统民族民俗文化的保护和传承做出了一定的贡献。本体作为一种重要的形式化表示方法[3]，对其相关技术的研究是知识融合、知识推理、知识问答的基础，并已经在知识工程和自然语言研究领域得到广泛的应用。因此，本文采用本体及其相关技术对畲族民俗领域知识进行建模，并研究其属性及层次关系，并最终在 protégé 平台上实现了畲族民俗本体。民俗本体的成功构建，使得民俗研究者可通过文本挖掘分析、可视化等方式实现对民俗特征现象 “远读”[4]，有利于传统民俗文化的保护和传承。

二、民俗资源知识化的理论基础

我们知道，相同专业或领域的人士在沟通时可能会使用他们自己的“行话”即专业术语进行交流，这样会导致专业背景不同的其他人员，因为其成长环境、专业背景以及现实需求的差异，出现对概念、结构和方法理解的差异，有时甚至会出现因为缺乏知识共享从而导致沟通不畅。另外，在当前各行各业数字化、信息化浪潮中，软件实现过程中不同的方法、范式、语言以及不同的软件实现工具因为缺乏共享的概念模型而导致软件互操作性差和相关数据信息难以重用和共享[5]。因此，为实现有效的信息沟通、高效的软件互操作能力和信息共享，作为实现领域知识共享和互联互通的知识框架术语–“本体”应运而生。根据 Studer 等人给出的定义，本体是“共享概念模型的明确的形式化规范说明”，其本质就是专业和领域知识的共享和复用，促进知识的传播和传承。

1、为什么要构建民俗领域本体

民俗，作为在特定时期，特定群体以及地域中形成、传播传承和演变的生活文化，它起源于人类社会群体生活的需要，为广大民众所创造、享用和传承。民俗是民间文化中带有集体性、传承性和模式性的现象，主要以口耳相传、行为示范和心理影响的方式来扩布和传承[6][7]。以民俗事项为载体的民俗文化，具备集体性、传承和扩布性、稳定和变异性以及规范和服务性等特征，表明民俗在形成、发展、传承和变异的过程中需要知识的共享和复用，这个共享和复用只是在特定地域和特定群体中通过特定的方式发生而已，不利于传统民俗文化的发展和繁荣，也不利于文化的传承和创新。而在互联网+时代，随着信息技术的发展，中国民俗文化已经被拉入到互联网+发展的语境之下，成为民族国家构建、经济转型升级、社会团结整合以及文化创新发展的资源[7]。但是当前大部分的民俗数字化的研究只是对民俗事项进行简单的录音、录像或者为某些文字信息建立数据库等低水平的数字化，难以适用当前对知识自动化、智能化的需求。而随着大数据、人工智能等信息技术的发展，当传统民俗文化邂逅人工智能，传统民俗知识如何组织和管理、如何实现形式化编码以及如何实现民俗文化资源的发掘保护，使得民俗资源在不同的知识空间环境下实现民俗知识的共享和复用，民俗本体的构建和实现将是不可或缺的过程。

2、民俗领域本体的知识组织过程

本体作为知识系统的核心概念，是不同系统之间通信、互操作的基础，本体构建的优劣直接影响知识系统的性能和使用。当前，针对各个专业领域的本体知识库很多，因为研究领域不一致，本体构建方法也不尽相同，目前还没有统一的本体构建标准方法。根据 Gruber 的研究论述，本文的本体构建遵循以下原则：1）本体概念清晰原则；2）概念定义一致原则；3）本体可扩展；4）最小编码原则；5）约定最小原则。基于以上 5 原则，结合本体构建过程常用的七步法来构建畲族传统民俗本体。同时，因为民俗本体属于领域本体的知识范畴，在民俗知识的组织和管理过程中需要民俗领域专家的参与，并在领域专家的指导下，确定民俗本体核心概念和关系属性，然后再由信息处理人员实现知识的存储和实现等过程。参考本体构建理论和领域本体的特点，本文的畲族民俗本体构建可以分为如下六个阶段：

（1）民俗本体需求分析阶段，在此阶段需要明确畲族民俗本体的领域范围，构建该本体可能应用的领域等；

（2）畲族民俗知识获取，在该阶段需要获取畲族传统民俗的领域范围、核心概念以及彼此之间的定性和定量关系等；

（3）确定核心民俗本体的分类层次关系，不同层次本体的属性关系等；

（4）创建本体对象实例，根据确定的民俗核心概念及其关系，创建不同层次的本体实例以及属性关系等；

（5）本体评价，利用现有的畲族民俗知识的文字描述材料，根据已创建的民俗本体对象，利用相关评价指标来判断本体创建的优劣；

（6）本体存储，基于某种数据格式和数据库平台来存储本体模型。

在完成上述六个阶段的分析处理后，民俗本体概念及属性的软件实现工具选择，我们可以从易用、易操作、跨平台和支持社区等方面考虑，选择在本体构建过程中使用较为广泛的开源软件 protégé，版本为较新的稳定版 5.5，本体构建语言选用 OWL-DL，本体数据存储平台使用 ACCESS。

三、畲族传统民俗知识资源的本体构建

畲族传统民俗文化随着自然条件、社会生活环境以及人文生态的变化，由于受到保存手段和传播方式的局限，使得许多在浩瀚历史长河流传了许多年的传统生活习俗和民间技艺面临流变和消失，民俗的保护和传承迫在眉睫。在现代信息技术语境下，虽给面临挑战的传统民俗文化带来了新的机遇，但是也会出现掌握民俗知识内容的人不懂技术，懂技术的没有内容资源的窘境[8]，民俗知识资源的本体构建就是为了跨越民俗资源数字化的鸿沟，实现民俗知识与技术的有机结合。

1 畲族传统民俗知识空间体系

畲民在长期的农业生产和生活实践中，创造了大量的物质和精神财富，形成了极具民族特色的传统文化和生产生活习俗，如“三月三”畲歌会、畲族歌舞、畲族传统婚礼以及畲族传统服饰等，其中的畲族三月三和畲族传统服饰在 2008 年 6 月入选第二批国家级非物质文化遗产目录。不同地域、不同时期的畲族民俗既有个性也有共性，要列出畲族传统民俗的所有知识领域有一定困难。本体，作为一种通用的知识形式化表示方法，在自然语言理解、语义形式化、知识工程以及几乎所有专业领域包括社会科学领域的知识形式化都得到了较广泛的应用。在民俗数字化已成为趋势的今天，构建畲族民俗文化本体，对于保护和传承畲族传统文化，促进传统民俗的共享与复用都将有非常重要的意义。根据本体知识空间构建的要求，确定领域知识范畴、提取核心概念并按其性质和特征进行分类然后按照本体层次关系，抽取概念和属性。我们根据畲族民俗的特征和社会功能等将畲族传统民俗知识体系分为发生时间、发生地点、表现形式、民俗类型、参与族群以及传承演变等，具体如下图 1 所示：

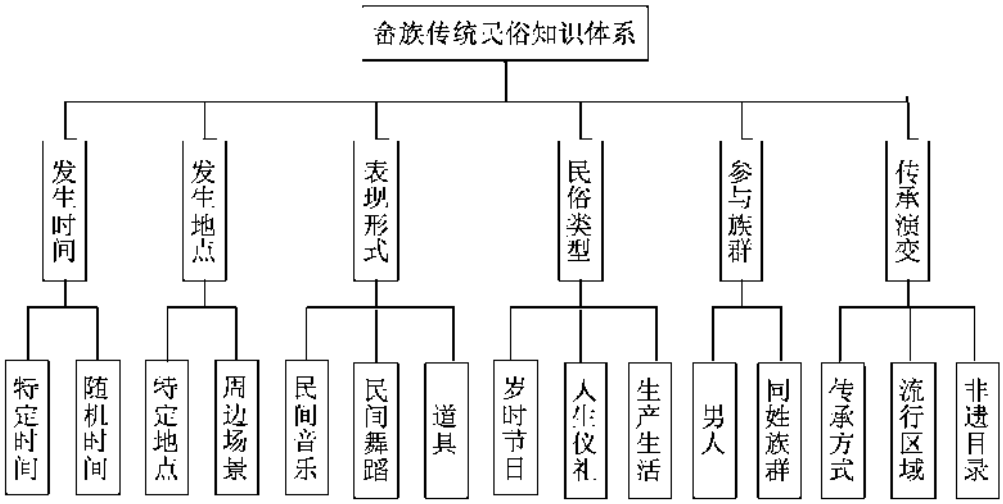

图 1　畲族传统民俗知识体系

以第二批国家级非物质文化遗产“畲族三月三”为例，在每年的农历三月初三这天，畲民都会穿上节日盛装，举行各种祭祀活动，参加民俗表演、山歌对唱等各种文娱活动，而且在三月三这天，畲民都会吃乌米饭等等这些风俗可以看出，畲族三月三牵涉到民俗文化体系中的发生时间、发生地点、表现形式以及民俗类型等方面，即民俗实例是传统民俗概念的具体体现。

2 畲族民俗本体核心概念定义及确定

在确定了畲族传统民俗的知识体系和范围后，需要对民俗知识进行分类抽象，抽取出构成畲族民俗本体的基本概念和层次关系。然后按照经济合作与发展组织（OECD-1996）的解释，根据“知识是什么-WHAT，知识为什么-WHY，知识怎么做-HOW，谁的知识-WHO”的方法来进行分类[6]183，结合定义的本体概念必须覆盖整个畲族民俗的知识体系且本体概念不能有二义性和交集等原则，将畲族民俗知识空间定义为生产生活类、岁时节日类、人生仪礼类和民间艺术类等四个一级分类，具体如下表 1 所示：

表 1　畲族民俗知识核心概念分类

一级分类	二级分类	民俗实例
生产生活类	农业狩猎类	猎神节、闹春舞、载竹舞等
	饮食服饰类	吃乌米饭、畲族服饰等
	居住建筑类	茅寮等
岁时节日类	宗教节日类	招兵节、封龙节、迎祖节等
	生产节日类	牛歇节、尝新节等
	年节类	畲族春节、畲族三月三等
人生仪礼类	诞生仪礼类	供“奶婆娘”、报喜、满月酒等
	婚姻仪礼类	畲族婚俗
	丧葬仪礼类	畲族葬丧仪式
	宗教仪礼类	功德舞、安祖舞等
民间艺术类	音乐类	杂歌、历史歌、解放歌、哭嫁歌等
	舞蹈类	猎捕舞、载竹舞、龙头舞等
	传统工艺类	畲族刺绣、畲族编织、畲族银器等

确定了畲族民俗本体四个一级分类后，另外还有宗教节日类等 13 个二级分类，在二级分类后就对应具体的民俗事件，如“畲族三月三”等 30 个具体民俗事件实例。通过以上概念和实例的确定，畲族民俗本体在 Protégé 中的实现如下图 2 所示：

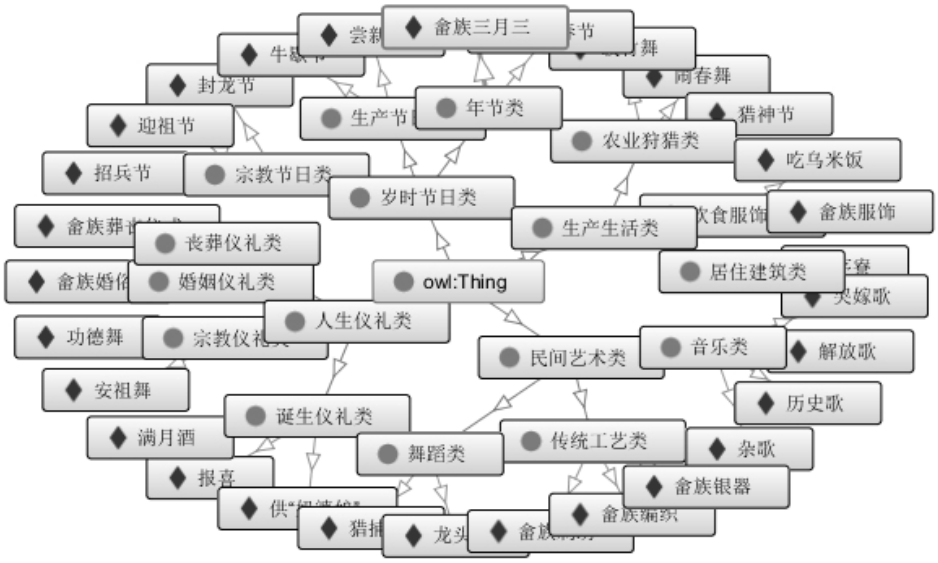

图 2　畲族传统民俗核心本体及实例的 protégé 实现

3 民俗本体实例属性关系构建

根据本体建模 protégé 的建模规则，以及对畲族民俗本体对象的抽象分析，可以将民俗本体的属性分为对象属性和数据属性。其中本体的对象属性是用来描述民俗本体以本体之间的关系，可以分为时间、地域、社会功能、表现形式以及传承演变等五个方面；数据属性则是描述本体的内在属性，表示本体实例的数量或其他量化的性质等，据此可以将数据属性分为数据类型和数据值两个方面。

在确定属性类型后，需要对属性做进一步的规范和限制，可以通过属性定义、确定属性取值范围、确定属性值域等属性分面来实现。上述本体对象属性和数据属性的定义如下：

时间：是指民俗事件发生的年月日或者某个时间段；

地域：是指民俗事件发生的行政区域或者某个特定的场所；

社会功能：是指民俗事件产生的起因，缘由以及对畲民的影响和功用等；

表现形式：是指民俗事件的活动形式，音乐、舞蹈或者山歌以及祭祀等形式；

传承保护：是指民俗事件在持续传播演变过程中，传承的方式和手段，是否已列入非物质文化遗产保护目录等；

数据类型：是指民俗本体属性中以数值形式表示的属性的类型，如数值数据类型，字符串数据类型等；

数据值：就是某个具体的数值数据或者字符串等。实例属性的详细信息如下表 2 所示：

表 2　本体实例属性分面

本体对象	属性名	属性类别	属性值
民俗本体实例（如畲族三月三）	时间	对象属性	可以是年、月、日或者某个时间点
		数据属性	对应年月日的数值
	地域	对象属性	行政区划省、市、县及乡镇等
		数据属性	对应行政区划级别的地点数量
	社会功能	对象属性	对应的具体民俗功能描述
		数据属性	涉及到的功能数量
	表现形式	对象属性	民俗事件的表达方式
		数据属性	各种具体表达方式的数量等
	传承保护	对象属性	传承保护的方式，是否进入非遗目录等
		数据属性	某种方式的数量，是、否等表现形式

根据图 1 的本体定义及表 3 的属性分面，在 protege5.5 中实现了畲族民俗本体各个实例的属性构建，选择较为典型的“畲族三月三”民俗，其属性关系如下图 3 所示：

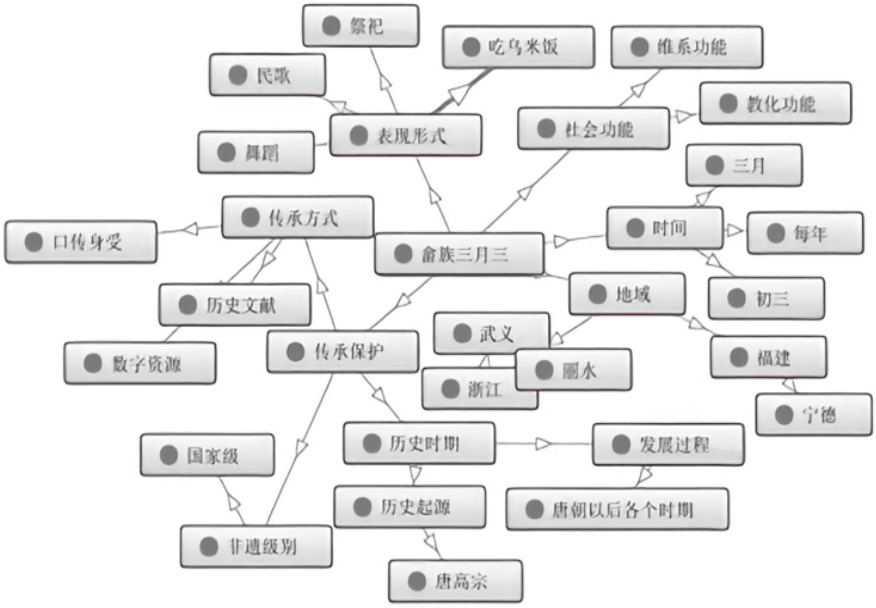

图 3　民俗本体实例“畲族三月三”属性架构

四、民俗本体评价

本体评价是本体构建过程中的一个比较重要的环节，其目的就是用来评价本体的构建过程是否合理、方法是否先进、构建的本体概念是否符合知识的实际使用需求以及本体的可扩展性和可重用性等方面。目前，本体评价的方法较多，如有主观意愿因素占比较多的基于用户评价的方法、基于构建原则评价的方法[9]等；也有使用定量评价的方法，如基于应用场景的评价方法和基于语料库的评价方法[10]等。

本文的评价对象即为创建的畲族民俗本体，语料库选用互联网中常用的一些文字介绍网站，评价的语料库 1 为百度百科中的“畲族三月三”民俗知识介绍页面，具体网址为“https：//baike.baidu.com/item/%E7%95%B2%E6%97%8F%E4%B8%89%E6%9C%88%E4%B8%89/16954292?fr=aladdin”，另外的一个语料库 2 选用东南网的畲族服饰知识介绍专栏网

站“http：//fuding.fjsen.com/2016-05/23/content_17852788.htm”。本体的具体评价指标为：概念覆盖度、属性覆盖度和实例关系覆盖度，对于每个评价指标的权重分配，我们采用三种不同的方式来分配。方案一：每个指标的权重都一样，即为 0.333；方案二：考虑各个不同概念的重要程度，各指标的权重设置为 0.4，0.4，0.2；方案三：各个指标的权重设置为 0.5，0.3，0.2。具体如下表所示：

表 3　各指标权重 w 分配

方案名	概念覆盖度	属性覆盖度	关系覆盖度
方案一	0.333	0.333	0.333
方案二	0.4	0.4	0.2
方案三	0.5	0.3	0.2

根据文献[11]的语料库评价模型，本体综合评价可表示为：

$$\text{Evaluation}(O) = \sum_{i=1}^{3} w_i \times Coverage[i]$$

其中 $Coverage[i]$ 为本体在语料库中的覆盖度，如为 $Coverage[1]$ 则表示本体概念在语料库中的覆盖度，w_i 表示对应方案的权重，如为 w_2 则表示方案 2，对应属性覆盖度的权重为 0.4，据此计算得出畲族民俗本体在语料库 1 和语料库 2 中的综合评价分如下表 4 所示：

表 4　畲族民俗本体覆盖度综合评价分

语料库名	方案一	方案二	方案三
语料库 1	0.733	0.729	0.743
语料库 2	0.704	0.701	0.698

最终本文构建的民俗本体在上述两个语料库中的综合覆盖度为 0.717，如果本体的覆盖度达到 1，则表示设计的本体可以完全覆盖语料库的内容，性能优秀；如果覆盖度为 0，则表示本体完成达不到设计的要求，性能很差。本文构建的畲族本体覆盖度评分为 0.717，根据五级制评分规则，性能介于一般与良好之间，还有较大的改进空间。

五、结语

本文基于本体构建的理论和方法对畲族民俗传统文化资源进行知识组织和整合，吸收民俗领域专家的意见构建了畲族民俗本体及其属性关系知识库，并利用互联网的民俗信息资源作为语料库，采用基于本体覆盖度的方法对本文的本体概念进行了综合评价，评价结果介于一般与良好之间。下一步我们将利用本文创建的畲族民俗本体来创建畲族民俗知识图谱，通过知识图谱和大数据的方式来发掘民俗传统文化传播和传承的规律，更好的促进民俗文化的保护和传承。

参考文献

[1]潘鲁生.民间手工艺的知识产权保护与文化传承问题[J].民间文化论坛，2012（03）：58-66.

[2]康保成.中国非物质文化遗产保护发展报告（2012）[M].北京：社会科学文献出版社，2012，87.

[3]Neches，R.，Fikes，R.E.，Finin，T.，Gruber，T.，Patil，R.，Senator，T.，& Swartout，W.R.Enabling Technology for Knowledge Sharing[J].AI Magazine，1991，12（3），36

[4]郭翠潇.计算民俗学[J].民间文化论坛，2017（6）：124-128

[5]Uschold M，Gruninger M.Ontologies：Principles，methods and applications[J].The Knowledge Engineering Review，1996，11（2）：93-136.

[6]A，Rudi Studer，V.R.B.B.C，and D.F.A.Knowledge engineering：Principles and methods[J].Data & Knowledge Engineering，2008，25（1-2）：161-197.

[7]钟敬文.民俗学概论[M]上海：上海文艺出版社，1998，313.

[8]李松，王学文.跨越数字鸿沟—信息化时代中国民俗文化数字化的现状、问题与对策[J].西南民族大学学报（人文社科版），2014（6）：155-160

[9]疏礼兵.组织知识、知识分类和知识特性[J].情报杂志，2008，27（1）：76-79

[10]马文峰，杜小勇，et al.领域本体评价研究[J].图书情报工作，2006，50（10）：68-71

[11]李江华，时鹏，郑剑.一种基于语料的本体评价方法[J].情报杂志，2012，31（7）：187-190

作者简介：李庆华（1976.10-），男，湖南湘乡人，博士，副教授，研究方向：数字化文化遗产。

基金项目：本论文为 2021 年度广东省教育厅科研项目“民俗大数据助力粤西乡村旅游研究” 的研究成果，项目编号：2021ZDZX4055。