

民族服饰图案语义标签体系构建研究

银宇堃¹, 赵海英^{1,2}

(1. 北京邮电大学数字媒体与设计艺术学院, 北京 100876;
2. 北京邮电大学世纪学院移动媒体与文化计算北京市重点实验室, 北京 102101)

摘 要: 民族服饰图案种类繁多, 内涵丰富, 给出一套科学规范的语义标签体系极具挑战性。为此提出一种基于标签民俗语义模型的民族服饰图案语义标签体系, 首先这是一种具有层次的标签体系结构, 其次在标签民俗语义模型的基础上, 添加领域专家与大众协同标注, 一方面反映了标签系统协作化理念, 另一方面提升了标签专业性。最后运用构建的标签体系对壮族凤凰图案进行语义标注, 同时还分析并比较了标签体系的实用性、先进性。

关 键 词: 民族服饰图案; 语义标签; 核标签; 专业标签; 大众标签

中图分类号: TP 391

DOI: 10.11996/JGJ.2095-302X.2018050926

文献标识码: A

文章编号: 2095-302X(2018)05-0926-07

Study on Semantic-Tag System Constructing Based on National Costume Pattern

YIN Yukun¹, ZHAO Haiying^{1,2}

(1. School of Digital Media and Design Arts, Beijing University of Posts and Telecommunications, Beijing 100876, China; 2. Mobile Media and Cultural Calculation Key Laboratory of Beijing Century College, Beijing University of Posts and Telecommunications, Beijing 102101, China)

Abstract: Due to the great diversity and rich connotation of national costume patterns, it's super challenging to come up with a scientific and standard semantic-tag system. This paper proposes a semantic-tag system based on the folk classification model. Firstly, this is a hierarchical tagging system. Secondly, professional and public annotations are added into this system based on the folk classification model, which not only reflects the collaboration concept for tag system, but also enhances the professionalism of tags. At last, this paper annotates the Phoenix pattern from Zhuang by the newly-created tag system, and makes an analysis and comparison in terms of the practicability and progressiveness of tag systems.

Keywords: national costume pattern; semantic-tag; kernel-tag; professional-tag; public-tag

民族服饰图案是民族服饰中重要的装饰元素之一, 也是民族文化特征与审美情趣的核心表达载体。大部分博物馆、展览馆、学校等都只是将民族服饰进行实物性保护, 如建立服饰展馆, 收藏不同地区、不同民族的服饰, 对服饰及相关的实物资源进行保护, 然而对于民族服饰的数字化保护同样重

要^[1]。文化资源的信息数据本身是对中国传统民族文化的保护、继承和弘扬^[2], 民族服饰融入了多元文化的精粹, 风格独具、光彩夺目, 是最具代表性的非物质文化遗产。通过民族服饰图案的解读可以了解深层的文化内涵, 例如宗教、民族、习俗等。本文对服饰图案的语义标签进行梳理, 并基于标签

收稿日期: 2017-11-27; 定稿日期: 2018-03-02

基金项目: 北京市科委课题项目(D171100003717003); 甘肃省人才引进项目(2015-RC-47)

第一作者: 银宇堃(1994-), 男, 河南洛阳人, 硕士研究生。主要研究方向为文化计算与媒体数据挖掘。E-mail: yyk1015@sina.com

通信作者: 赵海英(1972-), 女, 山东烟台人, 副教授, 博士。主要研究方向为文化计算与媒体数据挖掘。E-mail: zhaohaiying@bupt.edu.cn

民俗语义模型构建了一套民族服饰图案语义标签体系,有利于更好地理解民族文化信息,同时也可提高对民族服饰图案语义的检索能力。

1 相关工作

社会化标注这种开放、灵活、有趣的信息分类组织方式反映了用户的主观思想,被普遍认为是解决大数据时代信息过载的有效手段^[3]。研究社会化标注系统中用户的标注行为,对于引导用户积极参与网络资源的标签标注、提高标注质量、促进社会化标注系统的发展具有重要意义^[4]。

在大众分类系统开放的环境下,标签是共享和可见的,这样基于“群体自然语言”机制,用户会采用相似或相同的标签来描述相关资源,并根据标签使用频率达成共识,从而形成在用户中共享相似兴趣的标签集^[5],提高标签协作性。传统标签模型^[6]结构单一,仅采用结构简单的 $\{U, T, R\}$ 结构,不适于做协作化处理,且大多数反映的是用户个人的观点,标签是用户主动添加的,仅表达了用户对资源的理解和意见^[7],既没有对本体进行客观全面的深层语义分析,也不做规范化处理。正是由于上述原因,导致传统的标签在大多数情况下过多的包含了用户的主观想法,不能充分反映本体的语义信息或标签存在大量冗余。

语义标签是基于文化属性空间考虑,表征文化的本质属性,图案标签解决的是图案描述问题。由于标签数量众多,且对于某个特定标签而言,与其真正存在关系的标签并不多,所以通常用体系结构的形式设计标签,便于解决数据之间的关联问题。传统对于图案信息的描述基本采用属性编目的方式,仅用图案的几个属性作为图案的检索数据,缺少图案的内容信息,所以泛化性较差。新华社数据库系统部门提出了富标签体系^[8],该体系从属性信息、内容语义、底层视觉特征3个层次进行全方位、多角度的标注,使得标注信息更加丰富、有层次。

构建标签体系通常需要多维反映本体的语义信息。文献[9]中提出的视频语义上下文树,采用一棵带有上下文标签的标签树作为镜头语义上下文层次结构的表征模型,以内节点的上下文标签表征镜头语义间的上下文关联,能为视频内容理解提供显著的信息增益。

构建语义标签体系是针对目标资源,将其概念类、属性、实例以及其他元数据标注出来的过程。

在海量数据情境中,大量资源需要多用户的协作标注。WAL^[10]在讨论 Flickr 和 Del.icio.us 所发展的信息架构时,首先将“folk”和“taxonomy”组合,创造出新的词汇“Folksonomy”,并称其为“由下而上的社会分类法”^[11],认为 Folksonomy 是“社会创造的典型平面化命名空间”。Folksonomy 分类法^[12]的普遍优点是具有灵活性、适应性、自由的协作能力及偶然性,人们可以在没有理解术语的情况下使用其他术语作为标签。Folksonomy 分类法通过人们感兴趣标签信息的聚合实现社交联系并改善社交偶然性^[13]。民族服饰图案具有丰富的语义信息,但辨识度较低,所以本文基于标签民俗语义模型进行改进,通过领域专家和大众协同标注,既增加了标签的协作能力,又使标签具有一定的专业性。

2 基于标签民俗语义模型的语义标签体系构建

2.1 传统三元组模型和标签民俗语义模型

目前标签标注(Tagging)多采用传统的三元组模型^[6],即

$$\text{Tagging: } \{U, T, R\} \quad (1)$$

其中, U 为标签标注的用户; T 为标签集; R 为被标注的资源集。

标签民俗语义模型^[12]是一种具有协作能力的标签模型。其除了包含与标签标注行为有关的对象外,还覆盖了协作层面的关系,其定义为 Folksonomy: {tag set 标签集, user group 用户组, source 来源, occurrence 频度, Tagging 标签标注}。

该模型把传统的标签标注作为协作标签行为过程的组成部分,使得模型具有了双层包含关系,标签集即为所使用的所有标签集合,用户组为参与标签标注的用户集合,来源为大众标签法的标签来源地址,频度用来表征标签普及程度,标签标注则为三元组模型。

2.2 一种融合作业性和专业性的服饰图案语义标签体系构建

本文考虑到民族服饰图案标注的专业性特点,服饰标签太过专业化,会影响标签的协同性及在运用中的遍历能力,如在标签检索中出现检索范围越来越窄的问题。为此在民族服饰标签体系中引入大众标注增加标注协作性,同时又增加了领域专家标注为标签提供专业性,使标签个体行为更加明确。在原有的双层模型上加以改进,抽取三元组模型中

的标签和资源作为核标签模型；增加领域专家标签标注，获得专业标签模型；在原有用户组、标签集的基础上结合效用度概念，建立大众标签模型，最终构建“核标签模型、专业标签模型、大众标签模型”语义标签体系(图 1)。

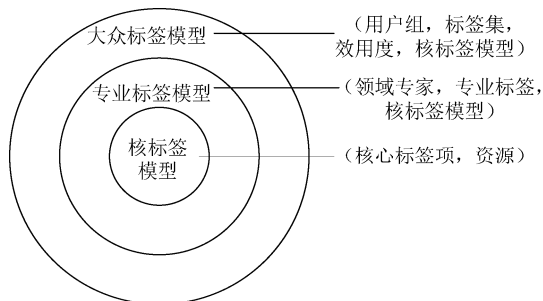


图 1 语义标签体系示意图

2.2.1 核标签模型

语义体系中最关键的一层是核标签模型，其是服饰图案语义标签的核心描述，即

$$K(T, R) \quad (2)$$

其中， T 为标签项； R 为待标注服饰图案资源。

2.2.2 专业标签模型

三层结构的中间一层为专业标签模型，在传统“用户、标签、资源”的三元组模型的基础上引入领域专家作为标引者，即

$$P(U, T_1, K(T, R)) \quad (3)$$

其中， U 为用户，这里是进行标注的领域专家； T_1 是由领域专家按核标签项讨论后给出的一组专业标签，第 3 个参数与式(2)中参数有同样的含义。

2.2.3 大众标签模型

三层结构的最外一层是大众标签模型层，与专业模型不同，领域专家较少而大众用户数量庞大，所以将标签民俗语义模型基础上将频度改进为效用度，即

$$F(G, T_2, E, K(T, R)) \quad (4)$$

其中， G 为用户组； T_2 为用户组按核标签项给出的标签集； E 为标签效用度，效用度结合了多个评价指标集，比频度更能反映用户对服饰图案标签的依赖程度；最后一个参数为式(2)中的核标签项和待标注服饰图案资源。

大众标签集的最终确定结合了标签效用度的计算方法^[14]。该方法通过标签评判指标集，确定指标集下单个指标的评判矩阵。然后确定权重集，并对每个指标赋予权重，权重计算公式^[14]为

$$a_i = \sigma_i / \sum_{i=1}^l \sigma_i \quad (5)$$

其中， σ_i 为第 i 项评判指标的标准差， σ_i 越大，说明第 i 项越能代表用户对不同标签效用的差异，这项的权重就越大。之后指定权重分配，并根据隶属函数^[15]计算模糊综合评价。最后考虑综合评价的结果，进行去模糊化，以将输出变为效用度的值，即第 k 个标签的效用度，通过上述方法并综合标注频度从而确定最终大众标签 T_2 。

3 核标签结构

首先收集民族服饰图案相关资料文章，然后使用基于语义的关键词提取(semantic-based keyword extraction, SKE)算法^[16]提取关键词，该算法首先对文本进行预处理，然后进行词语语义贡献值计算和词语统计特征值计算，最后通过上述计算结果加权得到词语关键度，从而得到关键词。之后经过专家筛选并结合实际情况确定作为标签字段的关键词：名称(Title)、民族(Nationality)、宗教(Religion)、地域(Region)、工艺(Crafts)、颜色特征(Color)、类别(Category)、构成(Structure)、寓意(Means)、适用人群(Users)、时间(Time)、语言(Language)。

同时，根据标签字段的所属性和专业性，又将核标签模型分为基础标签、高级标签、扩展标签，并在下文中给出核标签的内容描述(图 2)。

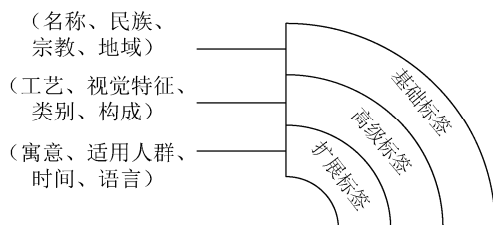


图 2 核标签结构示意图

3.1 基础标签

基础标签描述了民族服饰图案的根本属性，是民族性的基础表达，具有较高的专业性，其包含名称、民族、宗教、地域 4 种标签(表 1)。

3.2 高级标签

高级标签是指在基础标签之上，说明服饰图案细致特征的具象标签，具体包含工艺、颜色特征、类别、构成(表 2)。

3.3 扩展标签

扩展标签是图案语义的扩充及抽象表达，具体包含寓意、适用人群、时间、语言(表 3)。

表 1 基础标签

标识	名称	民族	宗教	地域
定义	该服饰图案的标准名称	该服饰图案所属的民族	该服饰图案象征的宗教	该服饰图案常用的地区
示例	龙纹、八角纹等	壮族、苗族等	佛教、基督教等	云南、贵州等

表 2 高级标签

标识	工艺	颜色特征	类别	构成
定义	该服饰图案的制作手法	人眼可观察到的图案颜色	图案的纹样种类	该服饰图案的构成方式
示例	刺绣、挑绣等	红、黑等	植物纹、几何纹等	对称、四方连续等

表 3 扩展标签

标识	寓意	适用人群	时间	语言
定义	服饰图案蕴含的深层涵义	该服饰图案的使用人群	该服饰图案起源或盛行的时期	该服饰图案包含的民族语言
示例	吉祥、幸运等	儿童、妇女、通用等	清朝、沿用至今等	藏语、蒙语等

4 民族服饰图案语义标签构建案例分析

4.1 图案语义标签标注流程

基于语义标签体系的民族服饰图案语义标签标注由领域专家和大众两部分组成，首先是领域专家对核标签项的标注，其次是大众对核标签项的标注，最终结合专业标签和大众标签得到标签集，标注流程如图 3 所示。

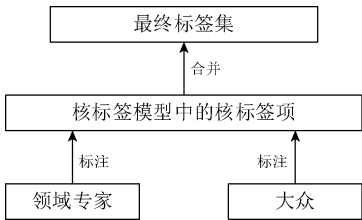


图 3 服饰图案语义标签标注流程图

4.2 标注案例

本文以壮族织锦中的凤凰图案(图 4)为例进行标注。



图 4 壮族织锦中的凤凰图案

具体的标注示例如图 5 所示(由于用户太多没有全部列出)。

由于大众标签的选择具有随意性，本文选取了 30 位大众用户对凤凰图案进行标注，并展示其对工艺字段的标注情况，如图 6 所示。

由图 6 可知，“织物”标签词标注次数最多，“编织”次之，“线织”、“布艺”等的标注次数也相对较高，结合标签标注次数并利用效用度计算方法^[14]得出这些标签的效用度，从而选出相应标签作为该字段的大众标签。最后通过该方法得出大众标签集，并结合专业标签构成完整标签集。

4.3 分析与比较

4.3.1 标签字段覆盖度分析

通过邀请民族服饰研究的领域专家，对 5 幅民族服饰图案进行标注，标注结果如图 7 所示。

标注完成后分析标签内容与本文体系标签字段的吻合度，见表 4。

由表 4 可知，名称、民族、工艺、类别、构成字段覆盖度较高；地域、颜色特征、适用人群、时间、语言字段的在标注过程中覆盖度较低，说明在未知标签字段标注过程中标签信息较少，且没有反映出图案的关键性信息、有大量冗余。其次，一方面民族服饰图案中不乏出现少数民族的语言，如蒙语、藏语等，民族服饰图案的使用人群及用色也比较考究，同时服饰图案往往可以反映少数民族的宗教信仰，并可追溯图案起源时间；另一方面由于标签字段较多，在一幅图案中可能无法全部反映，综上所述本文体系中设置的标签字段具有合理性。

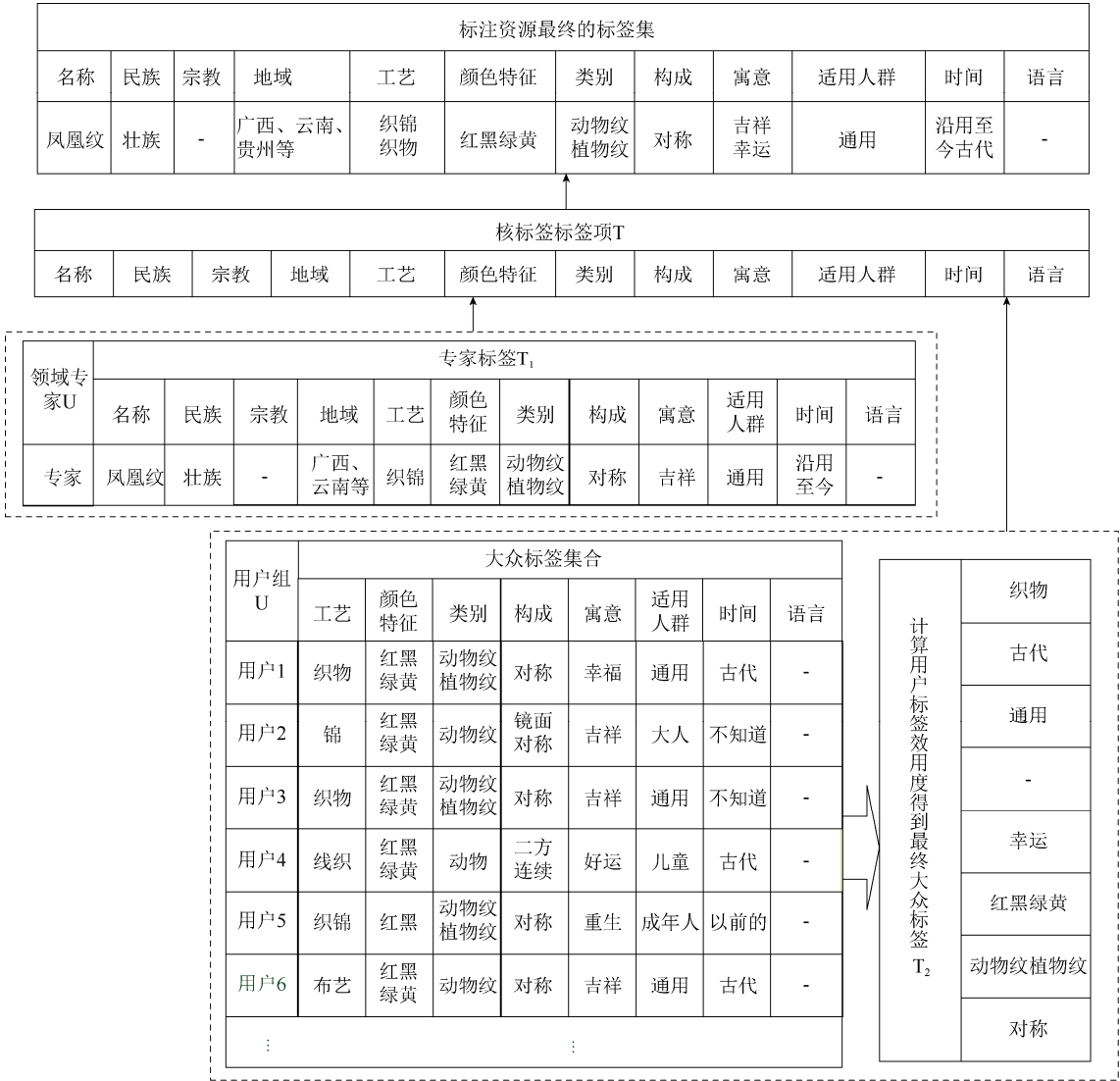


图 5 利用服饰图案语义标签体系标注壮族服饰图案示例

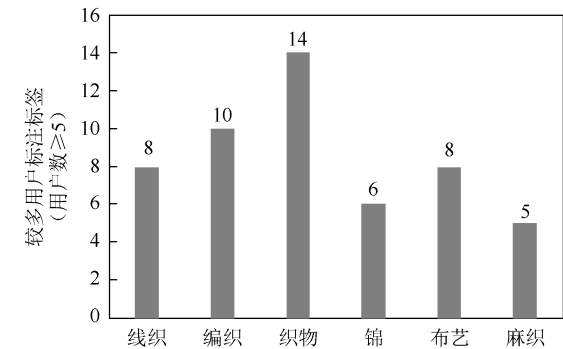


图 6 大众标注服饰图案标签情况

同时本文标签体系采取大众和专家共同标注，既具有协作标注能力，也排除了一些标签属性上，专家标签过于专业而标签不易理解的现象，另一方面大众标签丰富了标签集，使标签范围更广。

4.3.2 标签本体中类的比较

类是对现实世界中个体的抽象，类的完整性体

现了本体的完备性^[17]。本文的语义标签体系在 7 个基本类^[17]的基础上增加了专业标签和领域专家类，并与 Tag Ontology、SCOT、MOAT 等 3 种标签模型中的类进行了比较(表 5)。

从表 5 可看出，前 3 种标签本体均有资源、标签、标引者类，满足前文提到的三元组模型，SCOT 在 Tag Ontology 的基础上增加了用户组、标签集及来源类，具有协作标注能力；MOAT 虽然增加了含义类，但不具备协作标注的能力。本文提出的民族服饰语义标签体系，包含资源和标签类的同时将标引者分为领域专家和大众用户组。既包含用户组、标签集，具备协作标注能力；又引入领域专家、专业标签类，使得标签更具专业性、可信性；虽然在体系中没有含义类，但在核标签模型中包含对服饰图案寓意的解读，能够反映服饰图案语义信息，具有较好的表述能力。

标注图案	标签内容
	白族、传统图案、民族图案、白族元素、民族元素、中国传统文化、民族的、白族刺绣、白族服饰图案、白族刺绣、十字绣、挑绣、文字、植物、动物、符号、喜、福、花、四季常青、狗牙边、锁齿边、八角花纹样、装饰成双成对、祈祷、幸福、吉祥、写意、言情、爱情、感情、审美、和谐、向往、寄托、赠物、人与生活、四季常青、永不衰落、永恒、大理之花
	白族、传统图案、民族图案、白族元素、民族元素、中国传统文化、民族的、白族刺绣、白族服饰图案、白族刺绣、十字绣、挑绣、白族服饰边饰、植物、符号、二方连续、花、米字花、方形、十字花、对称、散点、波纹、自然、装饰、美好、吉祥
	白族、传统图案、民族图案、白族元素、民族元素、中国传统文化、民族的、白族刺绣、白族服饰图案、白族刺绣、十字绣、挑绣、植物、动物、符号、二方连续、纹样、八角花纹、十字花、雪花、叶子、蝴蝶、对称、菱形、自然、抽象、美好追求、吉祥如意
	白族、传统图案、民族图案、白族元素、民族元素、中国传统文化、白族刺绣、白族服饰图案、白族刺绣、十字绣、挑绣、二方连续、几何纹样、底纹、十字绣编织图案、十字、方形、对称、装饰、服饰边饰、室内装饰、包装、经典、实用、生活
	白族、传统图案、民族图案、白族元素、民族元素、中国传统文化、民族的、白族刺绣、白族服饰图案、白族刺绣、十字绣、挑绣、二方连续、几何纹样、底纹、十字绣编织图案、方形、对称、图形、装饰、服饰边饰、包装

图 7 5 幅民族服饰图案及其标签内容

表 4 标签字段覆盖度分析

组别	名称	民族	宗教	地域	工艺	颜色特征	类别	构成	寓意	适用人群	时间	语言
A	√	√	-	-	√	-	√	√	√	-	-	-
B	√	√	-	-	√	-	√	√	√	-	-	-
C	√	√	-	-	√	-	√	√	√	-	-	-
D	√	√	-	-	√	-	√	√	-	-	-	-
E	-	√	-	-	√	-	√	√	-	-	-	-

表 5 Tag Ontology、SCOT、MOAT 与民族服饰语义标签本体中类的比较

类	Tag Ontology	SCOT	MOAT	民族服饰语义标签本体
资源	√	√	√	√
标签	√	√	√	√
标引者	√	√	√	-
用户组	-	√	-	√
标签集	-	√	-	√
来源	-	√	-	-
含义	-	-	√	-
领域专家	-	-	-	√
专业标签	-	-	-	√

5 结 束 语

本文提出了基于标签民俗语义模型的三层语义标签体系来准确系统地获取民族服饰图案的文化内涵。

民族服饰的纹样作为中国传统文化的载体,有着悠久的历史,语义标签作为文化资源的一种显性表示,在文化的传承中起到很重要的作用。对民族服饰图案的标注过程,既是对服饰图案的解析,又是对民族服饰图案从古至今演变过程的整理与具象化,也是向大众推广民族服饰纹样的过程,所以民族服饰语义标签体系要求更加精准和丰富。不同于以往的标签体系,本文的语义标签体系引入了大众和领域专家协作标注的概念,在确定民族服饰图案核心标签模型之后分别引入了专业标注和协作化标注方法,一方面使得获得的标注标签更加专业,另一方面也允许大众用户参与标注工作,体现出语义标注的协同性,使得语义标签准确的同时更易于理解,表现出更丰富的语义信息。

(感谢中央民族大学的领域专家提供图 7 的数据支持!)

参 考 文 献

- [1] 王真. 少数民族服饰教育资源库构建研究[D]. 昆明: 云南师范大学, 2016.
- [2] 陈寅飞. 基于语义的少数民族民间文化资源库研究[D]. 昆明: 云南大学, 2014.
- [3] 刘礼锋, 潘旭伟. 社会化标注中用户标签的主题鲜明性研究[J]. 情报理论与实践, 2017, 40(8): 135-138.
- [4] 邵杨芳, 陈新国. 社会化标注系统中用户的标注行为及差异分析[J]. 图书馆, 2017(10): 42-49, 61.
- [5] 熊回香, 王学东. 大众分类体系中标签与本体的映射研究[J]. 情报科学, 2014(3): 121-126.
- [6] KIM H L, SCERRI S, BRESLIN J G, et al. 最新标签本体: 标签和民俗分类法语义模型[J]. 数据分析与知识发现, 2009, 3(3): 30-37.
- [7] 吴晓芳. 社会标注中标签语义分析研究[D]. 大连: 大连理工大学, 2011.
- [8] 王民. 富标签体系在中国全球图片总汇中的研究与实现[C]//中国新闻技术工作者联合会 2012 年学术年会、五届四次理事会暨第六届“王选新闻科学技术奖”的“人才奖”和“优秀论文奖”颁奖大会论文集. 河南: 中国新闻技术工作者联合会, 2012: 8.
- [9] 余春艳, 苏晨涵. 视频语义上下文标签树及其结构化分析[J]. 图学学报, 2015, 36(5): 747-755.
- [10] WAL T V. Folksonomy [EB/OL]. [2015-12-10]. <http://vanderwal.net/folksonomy.html>.
- [11] 张锐. 面向期刊论文的多粒度语义标注方法研究[D]. 武汉: 华中师范大学, 2015.
- [12] KIM H L, SCERRI S, BRESLIN J G, et al. The state of the art in tag ontologies: a semantic model for tagging and folksonomies [C]//Proceedings of the 2008 International Conference on Dublin Core and Metadata Applications. New York: ACM Press, 2008: 128-137.
- [13] QUINTARELLI E. Folksonomies: power to the people [J]. ISKO Italy-UniMIB Meeting, 2005, 38(6): 289-299.
- [14] 王珉, 王永滨. 网络音乐标签效用度和稳定性研究[J]. 计算机科学与探索, 2015, 9(5): 535-545.
- [15] 黄卫华, 方康玲, 章政. 典型模糊控制器的隶属函数设计及分析[J]. 模糊系统与数学, 2010, 24(5): 83-90.
- [16] 王立霞, 淮晓永. 基于语义的中文文本关键词提取算法[J]. 计算机工程, 2012, 38(1): 1-4.
- [17] 熊回香, 廖作芳, 蔡青. 典型标签本体模型的比较分析研究[J]. 情报学报, 2011, 30(5): 479-486.