

# 基于 Drupal 的民国建筑知识库关联数据的组织与发布研究\*

张乐常娥

DOI:10.15941/j.cnki.issn1001-0424.2018.19.010

**【摘要】**文章对民国建筑知识库关联数据的组织聚合和关联化发布进行了实践研究,以期促进图书馆关联数据的开放与共享。在对图书馆关联数据的组织聚合进行研究的基础上,以民国建筑知识库为例,设计了关联数据本体,使用 Drupal 平台实现了关联数据的关联和发布。使用 Drupal 平台可以快捷实现关联数据的组织和发布,但还需进一步提升模块插件的适用性和大数据量转换效率。

**【关键词】**关联数据 图书馆 Drupal 民国建筑知识库

**Abstract:** This paper studies organizational and realization of the linked data in the architectural repository of the period of the Republic China, aiming at promoting the opening and sharing of the linked data in the library. On the basis of the study of the organization and aggregation of the linked data in the library, taking the architecture repository as an example, the paper designs the linked data ontology, and uses the Drupal platform to realize the linked data. Using Drupal platform can organize and publish linked data conveniently, but the applicability and conversion efficiency of the module plug-ins need to be improved.

**Key words:** linked data library Drupal architectural repository of the period of the Republic China

随着网络和数字技术的发展和应用,图书馆资源的类型增多、来源更加广泛。而这些数字资源分布在图书馆内部的数据库和网络空间中,以不同标准的结构化和非结构化数据形式存在着。这些数据含有很多值得揭示和复用的内容,但是它们都隐藏在实体内部,缺少结构化的独立标识和描述<sup>[1]</sup>,也未能揭示出实体之间的潜在关系。因此,如何将图书馆内部的数据组织聚合,实现多种类型资源的语义链接,开放共享资源,是开放型的现代数字图书馆亟待解决的问题。

为了解决现有万维网在信息表达、组织、检索中存在的问题,使互联网为大众提供更完善的服务,万维网联盟的蒂姆·伯纳斯-李在 1998 年提出了语义网的概念<sup>[2]</sup>。2006 年,伯纳斯-李在语义网的基础上提出了发展数据之网的构想,并指出核心和关键是关联数据。关联数据是语义网关键的一种实现方式,它侧重于数据相互的关联和人机之间理解的语义环境。随着关联数据应用的发展,越来越多的组织和个人认识到关联数据的重要性。2010 年,万维网协会(W3C)宣布成立的专门图书馆关联数据孵化小组(Library Linked Data Incubator Group)<sup>[3]</sup>,帮助图书馆使用关联数据,以便使图书馆将自有数据以关联数据的形式进行发布<sup>[4]</sup>。目前,很多具有影响力的国外图书馆已经将自己特有的数据转换为关联数据发布出来,如德国国家图书馆书目数据的关联数据发布等<sup>[5]</sup>。这些成功的案例对国内图书馆关联数据的组织聚合与发布实现提供了很有价值的参考。

Drupal 作为一种基于开源的 XAMPP 架构的内容管理系统,它特有的模块化结构特性可以方便地对功能进行扩展,通过一系列符合 RDF 标准的模块应用,完整地支持关联数据发布和消费,是实现关联数据应用较佳的选择。本文使用 Drupal 平台,以民国建筑知识库为例,从图书馆数字资源的组织聚合、关联数据本体的设计、元数据关联和发布几个方面进行了图书馆关联数据的组织和实现研究。

\* 本文系国家社会科学基金项目“图书馆资源组织中的数据关联机制研究”(项目编号:14CTQ005)研究成果之一。

## 1 关联数据与图书馆资源的组织聚合

伯纳斯-李认为语义网是一个使用现有技术根据语义来判断的智能网络。人们可以通过语义网中的计算机,在万维网上的海量资源中找到自己所需要的信息,从而将孤立的信息联系起来<sup>[6]</sup>。关联数据作为语义网一个重要应用,建立了一种集合多种元数据标准的框架,使机器能够理解被描述资源的元数据的信息<sup>[7]</sup>。

目前,关联数据在全球各个领域得到了广泛应用,图书馆界也加强了对它的关注和研究。W3C 成立了针对图书馆关联数据的孵化组织,逐步建立了诸如元数据模型、元数据架构等的标准和协议,鼓励将图书馆馆藏资源转换为关联数据发布出来,并链接语义网上其他领域的资源,以促进图书馆的资源共享。现代图书馆馆藏不再限于传统的纸质书本,更多的是电子资源,比如各种特色库、知识库、学者库等等。但这些资源都隐含在各个图书馆内部,这种孤立的分布式异构数据给用户带来了查找和使用方面的问题,需要对图书馆的资源重新组织聚合。

图书馆数字资源组织聚合有以下几种方式:基于数据层的资源组织聚合、基于应用层的资源组织聚合以及基于关联数据的资源组织聚合。基于数据层的资源组织聚合模式以 OPAC 集成为主要方式,主要针对图书馆内部的资源组织,通过 Z39.50 协议和添加 MARC 字段的方式来揭示和链接书目信息,以此来实现图书馆不同类型资源的内部组织。基于应用层的资源组织聚合模式主要通过跨库检索系统来实现,通过开发一个统一的检索工具,在后台链接异构的不同数据源来实现组织。清华同方的 CNKI 统一检索平台,以及目前比较流行的发现检索平台等都是较为典型的应用层的组织方式。但是这两种资源融合方式,都只能在局部范围内实现异构资源的集成,既不能解决多种资源语义异构的问题,也不能真正实现整体上的相互关联。随着以关联数据为核心的语义网的理念和技术的发展,为资源融合提供了更加合适的方法。基于关联数据的资源组织聚合模式弥补了异构网络信息的语义性缺失,通过定义、关联和发布结构化的数据,使得表面上分离的异构数据源实现了相互的语义关联,从而实现了真正的整体融合。

通过关联数据对图书馆资源元数据进行语义化重新组织聚合,建立与外部相关资源间的语义关联,以关联数据的方式发布到网上,可以大大加强图书馆内部数据外部的关联度和扩展度。

## 2 Drupal 与关联数据的发布实现

随着 W3C 逐步建立了一系列符合图书馆关联数据要求的标准和协议,图书馆将馆藏资源转换成关联数据发布出来变得日趋便捷。W3C 的 RDF 规定使用三元组 (subject, predicate, object) 三部分组成的资源描述通用模型,关联数据的三元组都应以 URI 表示。根据伯纳斯-李在《Linked Data》中定义的关联数据四项准则,关联数据的实现一般应包括以下几个步骤:

- (1) 选择合适本体词表,建立 RDF 数据本体模型描述数据资源。
- (2) 实例化数据模型,为资源数据生成 URI,建立相关联数据之间的 RDF 链接。
- (3) 将数据内容发布为关联数据,提供标准化接口对 RDF 数据库进行检索。

目前流行的关联数据实现工具有很多,比较常用的有 Drupal, D2R, Pubby 等。相比较而言,Drupal 作为一款建立在一个成熟的 Web 应用框架之上的 CMS (内容管理系统),它的内容结构定义很适用于对关联数据的支持。同时,它模块化的架构设计可以非常方便地对功能进行扩展,使用者可以通过模块的安装来控制功能的应用,安装新模块就可以增加新的功能。自 2008 年起,Drupal 逐步推出了支持关联数据的多个模块,例如针对 RDF 内容模型创建的 RDF 模块,作为 RDF 数据存储器的 ESMOL-ARC2 (Archive2) 等模块,以及针对本体定义和映射的 CCK 和 evoc 模块,分别实现 RDF 数据实体的内外部链接的 Reference 模块、Link 模块<sup>[8]</sup>,提供对关联数据检索支持的 SPARQL, RDF SPARQL Endpoint 模块等。相对于其他几种工具,Drupal 更符合本案例的小数据量关联数据实现的应用。

## 3 民国建筑知识库的关联数据组织聚合研究

民国建筑知识库为图书馆自建的专题知识库,它系统地搜集和整理了南京现存的具有代表性的 200 余处民国建筑资料,包括民国建筑介绍及建筑相关的建筑师、文献、研究、历史事件等多种类型资源,例如相关的建筑文献包括研究性论文、出版书籍等资源类型。<sup>[9]</sup>因此,实现民国建筑知识库关联数据的组织聚合,需要先确定数据资源的实体和基本属性,设计和构建关联数据的元数据本体模型。

### 3.1 确定资源实体和属性

在设计元数据本体之前,需要详细整理民国建筑数据中作为实体的内容,以及实体的属性和实体之间的联系。参照《南京民国建筑》一书中对南京民国建筑的定义<sup>[10]</sup>,涉及到的资源实体包括建筑、建筑责任人(设计者)、

建筑描述、建筑年代、地理位置、建筑研究等。建筑实体的属性包括建筑类型、建筑年代、建筑描述等。建筑责任人包括建筑师的姓名、性别、生平、籍贯、年代、成就、思想、作品等，建筑描述包括建筑地点、流派、艺术风格、结构、材料等类型。地理位置包括了经度、纬度、高度等属性。相关研究包括馆藏书目、论文、图片、视频等类型资源。其中资源数据包括来源于图书馆内部的电子书目信息、馆藏书目、相关论文等资源实体，以及来源于图书馆外部的图片、视频、地理位置等资源实体。

### 3.2 元数据本体设计和构建

为了实现不同类型元数据的集成和融合，元数据的设计至关重要，目前主要有两种主流的构建方式：多本体型和混合型构建模式<sup>[11]</sup>。多本体型设计相对灵活，能够应对变化的元数据规范，而混合法设计方式比较容易实现应用层面元数据间的语义互操作。结合本案例的实际需求，选择多本体法作为元数据设计构建的基础，即采用不同元数据规范内的本体与属性，构建一个合成的元数据本体。考虑到元数据本体的易用和通用性要求，主要选择了现在广泛通用的元数据本体（DC，FOAF，Geonames，EVENT等），通过本体间相互映射，实现不同本体间的关联。本研究选择基于DC元数据（DC即Dublin Core，是全球应用最多的一种元数据标准）的12个DCTERMS的核心元素作为核心元数据本体，并复用Geonames本体作为建筑的地理位置属性，复用FOAF本体表示建筑建筑师的属性。精简的核心元数据本体基本可以描述民国建筑的共同属性。图1为一个简略的民国建筑本体模型。

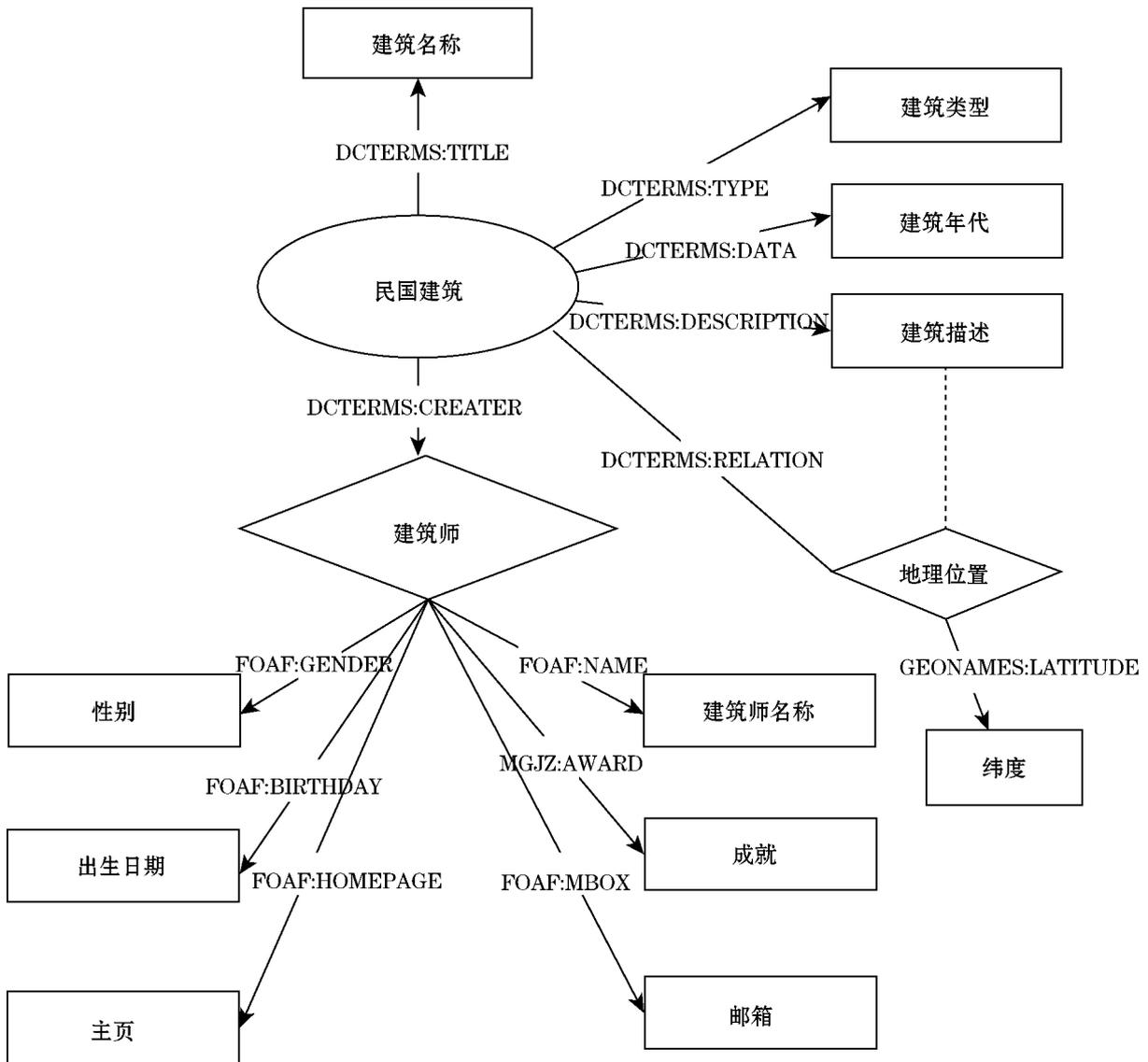


图1 民国建筑本体模型

如图 1 所示, 民国建筑本体除了包含了 DCTERMS: TITLE, DCTERMS: TYPE, DCTERMS: DATA 等属性, 还关联了建筑师本体, 由于其中涉及到人物、位置、事件等多种资源类型, 因此复用了 FOAF、GeoNames 等本体词汇表对这些属性进行了扩展。FOAF 本体即 Friend-of-a-Friend<sup>[12]</sup>, 是主要用来描述人及其基本特性, 如姓名、电子邮件地址的本体词汇表。GeoNames 本体集成了各种地理信息数据, 比如经纬度、异种语言的地名、海拔, 以及其他一些信息, 能够对民国建筑的地理位置做出完善的表达。在此例中使用 LATTUDEDE 属性表示民国建筑实体的地理位置坐标属性。

如图 2 所示, 本案例建筑师实体复用了 FOAF 本体, 如: FOAF: NAME, FOAF: GENDER, FOAF: MBOX, FOAF: BIRTHDAY, FOAF: HOMPAGE 等表示建筑师实体的姓名、性别、生日、邮箱、主页等主要属性, 并根据实体属性需求扩展了网络资源链接的属性 MGJZ: CONNECT 和代表作品属性 MGJZ: AWARDS。

FOAF复用和扩展属性

FOAF:AGENT

- ◆ FOAF:NAME
- ◆ FOAF:GENDER
- ◆ FOAF:MBOX
- ◆ FOAF:BIRTHDAY
- ◆ FOAF:PHONE
- ◆ FOAF:HOMEPAGE
- ◆ DC :TITLE
- ◆ MGJZ:CONNECT
- ◆ MGJZ:AWARDS

图 2 FOAF 复用和扩展属性

#### 4 民国建筑知识库关联数据的发布实现

本案例使用 Drupal7 为关联数据实现工具, 民国建筑知识库关联数据实现包括以下几个步骤: 构建内容节点类型和属性、节点与本体库的关联映射, 以及 RDF 数据的生成、发布与查询。

##### 4.1 构建内容节点类型和属性

Node (节点) 是 drupal 的基本存储单元, 在 Drupal 中几乎所有的内容都被存储为一个节点, 在内容节点之下包含各种字段。通过使用 Drupal7 中的 CCK (Content Construction Kit) 模块可以自由创建各种节点的类型和字段。根据上节元数据本体的设计, 共建立 7 个内容类型, 如: 建筑、建筑师、地理位置、历史人物等。以建筑类型为例, 根据元数据中的建筑名称、建筑结构、建筑单位、建筑类型等属性添加了 Arch\_Name, field\_arch\_structure\_type, field\_arch\_structure\_type, field\_arch\_type 等字段。其中 field type 代表字段属性类型是 (文本型、字符型、节点引用等类型)。通过使用 FIELDS 模块下的 Node Reference 功能, 设置建筑师、地理位置、历史人物等复用本体类型的字段属性为 (Node reference) 节点引用。内容节点建立完成后, 系统会自动给每个节点分派一个 URI, 地址表示为: <http://localhost/drupal7/node/序号数字.rdf>。

建筑内容节点创建如图 3 所示。

Architecture						
		EDIT	MANAGE FIELDS	MANAGE DISPLAY	RDF MAPPINGS	COMMENT FIELDS
Home » Administration » Structure » Content types » Architecture						
LABEL	MACHINE NAME	FIELD TYPE	WIDGET			
+ Title	title	Node module element				
+ 建筑名称	field_arch_name	Text	Text field			
+ 建筑师*	field_architect_name	Node reference	Check boxes/radio buttons			
+ 地理位置*	field_address	Node reference	Check boxes/radio buttons			
+ 建筑类型	field_arch_type	Text	Text field			
+ 建筑结构	field_arch_structure_type	Text	Text field			
+ 历史人物*	field_historical_figures	Node reference	Check boxes/radio buttons			
+ 使用单位	field__i	Text	Text field			
+ 历史事件*	field_history_id	Node reference	Check boxes/radio buttons			
+ 建筑描述	field_common	Text	Text field			
+ 建设年代	field_arch_found_year	Text	Text field			
+ 相关论文*	field_book	Node reference	Check boxes/radio buttons			
+ 馆藏目录*	field_architecture_book	Node reference	Check boxes/radio buttons			
+ 使用单位	field_use_unit	Text	Text field			
+ 建筑材料	field_arch_material	Text	Text field			
+ 关联网络资源	field_architecture_resouces	Text	Text field			
+ 建筑单位	field_construction_unit	Text	Text field			

图3 建筑内容节点创建

#### 4.2 关系 RDF 数据的关联资源内容类型与本体库之间的映射

内容节点创建完成以后,需要对复用本体的内容类型与本体之间进行映射关联。以建筑师节点为例,映射如图4所示。

建筑师		EDIT
Home » Administration » Structure » Content types » 建筑师		
Manage the way this bundle and its fields are represented		
Type	sioc:Item, foaf:Document	RDF Type
Title	dc:title property	sioc:Item
姓名	foaf:name property	Enter a co
个人简介	property	
个人图片	property	
性别	foaf:gender property	
主页	foaf:hhomepage property	
出生日期	foaf:birthday property	
邮箱	foaf:mbox property	
成就	MGJZ:award property	
联系方式	foaf:phone property	

图4 内容类型建筑师 (Architect) 部分映射图

如图 4 所示,其中用到 DC, FOAF 等本体库。对于复用 FOAF 本体库的 name, gender, homepage, phone 等属性使用 RDF 模块中的 MAPPINGS 功能与相对应的 FOAF <http://xmlns.com/foaf/0.1/> 中的属性建立关联。将属性 title 与 DC 本体库中的 title 属性建立关联映射。其他内容类型如地理位置 (Address) 中的复用属性与“GeoNames”本体库相应的属性建立关联映射,历史事件 (field\_history\_id) 中的复用属性与“EVENT”本体库相应的属性建立关联映射等。

#### 4.3 RDF 数据的发布与查询

所有内容类型的节点创建和关联完成以后, DRUPAL 系统会自动生成相对应 RDF 数据,通过对所有内容模型的实例化,完成民国建筑关联数据的最终实现。

图 5 所示为实例化的“历史人物”内容模型 RDF 数据。

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/ele,emts/1.1/"
  xmlns:foaf= http://xmlns.com/foaf/0.1/>
  <rdf:Description rdf:about="http://localhost/drupal7/node/53">
    <foaf:name>郭秉文</foaf:name>
    <foaf:gender>男</foaf:gender >
    < foaf:Mbox>non </foaf:Mbox >
    < foaf:birthday>1880</foaf:birthday >
    <dc:description>郭秉文 ( 1880—1969 ) ,男,字鸿声,南京浦口人,生于上海青浦,早年毕业于上海清心书院,1908 年赴美留学,1914 年获哥伦比亚大学教育学博士学位;回国即参与“南高”创办,先后任南京高等师范学校教务主任、校长和国立东南大学校长。
    </ dc:description >
    <foaf:homepage>https://www.baidu.com/baidu?tn=monline_3_dg&ie=utf-8&wd=%E9%83
    %AD%E7%A7%89%E6%96%87</foaf:homepage>
  </rdf:RDF>
```

图 5 “历史人物”内容模型的 RDF 数据

RDF 数据发布以后,通过打开 Drupal 的模块“SPARQL Endpoint”,可以提供 SPARQL 访问接口,使用标准的查询语句,获得相关的 RDF 数据集。

#### 5 结语

通过对民国建筑知识库关联数据的关联化组织和发布实现的研究发现,使用 Drupal 平台实现关联数据的组织和发布,具有操作简单、扩展容易的特点,但也存在着一些不足:

(1) Drupal 中的第三方模块更新与维护得不到及时保障,兼容性需要使用者自行测试。

(2) 在通过本体和内容模型之间的映射关系实现语义关联过程中,如果映射关系定义有疏漏就会产生语义的缺失。

(3) Drupal 的关联数据存储模块 ARC2 更加适合于小型数据库关联数据的实现。而在图书馆的一些大型复杂的数据库,如学者库、论文库的关联数据转换过程中, D2R 这类转换工具的转换效率更加适合。随着关联数据技术的不断应用和发展,灵活使用多种类型关联数据转换工具如: Drupal, D2R, Pubby 等,将会更加高效地应用和共享图书馆的关联数据。

#### 注释

[1]刘炜. 关联数据: 概念、技术及应用展望[J]. 大学图书馆学报, 2011 ( 2): 5-12.

- [2]语义网 [EB/OL]. [2018-05-20]. [https://baike.baidu.com/item/%E8%AF%AD%E4%B9%89%E7%BD%91/118508?fr=aladdin#reference- \[1\]-1967-wrap](https://baike.baidu.com/item/%E8%AF%AD%E4%B9%89%E7%BD%91/118508?fr=aladdin#reference- [1]-1967-wrap).
- [3] Knight, F. Tim. Library Linked Data Incubator Group: Review of the Final Report [EB/OL]. [2018-05-20]. <https://yorkspace.library.yorku.ca/xmlui/handle/10315/15524>.
- [4]曹玉平, 龚主杰, 陈德容, 康阳光. 关联数据技术及其研究现状[J]. 图书馆理论与实践, 2014 (11): 42-45.
- [5]Linked Data Service of the German National Library [EB/OL]. [2018-05-20]. [www.dnb.de/EN/Service/DigitaleDienste/LinkedData/linkedata\\_node.html](http://www.dnb.de/EN/Service/DigitaleDienste/LinkedData/linkedata_node.html).
- [6]Heath. Linked Data: Evolving the Web into a Global Data Space [EB/OL]. [2018-05-20]. <http://linkeddatatoolkit.com/editions/1.0/>.
- [7]冯玲. 一种基于语义的 P2P 网络资源检索机制[D]. 武汉: 华中科技大学, 2007.
- [8]Entity reference [EB/OL]. [2018-05-24]. <https://www.drupal.org/project/entityreference>.
- [9]顾建新, 袁曦临, 常娥. 南京民国建筑与文化数字博物馆建设研究[J]. 新世纪图书馆, 2011 (12): 37-39.
- [10]卢海鸣, 杨新华. 南京民国建筑[M]. 南京: 南京大学出版社, 2001.
- [11]王薇. 基于关联数据的图书馆数字资源语义融合研究[D]. 南京: 南京大学, 2013.
- [12]Foaf Vocabulary Specification 0.99 [EB/OL]. [2018-05-20]. [http://xmlns.com/foaf/spec/#term\\_Agent](http://xmlns.com/foaf/spec/#term_Agent).

张 乐 东南大学图书馆馆员, 硕士研究生。通讯作者。

常 娥 东南大学图书馆副研究馆员, 博士。

---

(上接第 61 页)

- [8]蒲筱哥, 乔亚铭, 胡亚敏. 基于网络分析法的高校图书馆电子资源服务绩效评价模型及实证研究[J]. 大学图书馆学报, 2014 (4): 41-49.
- [9]Li Yingjie. Research on Application of BP Neural Network in Library Electronic Resource Quality Evaluation [J]. Revista de la Facultad de Ingenieria, 2017 (5): 605-613.
- [10]Gkoumas Georgios, Lazarinis Fotis. Evaluation and Usage Scenarios of Open Source Digital Library and Collection Management Tools [J]. Program -electronic Library and Information Systems, 2015 (3): 226-241.
- [11]刘佳音. 国内高校数字资源利用效能综合评价研究[D]. 武汉: 武汉大学, 2013.
- [12]Younghee Noh. Evaluation of the Resource Utilization Efficiency of University Libraries Using DEA Techniques and a Proposal of Alternative Evaluation Variables [J]. Library Hi Tech, 2011 (4): 697-724.
- [13]左利云, 曹志波, 董守斌. 云计算虚拟资源的熵优化和动态加权评估模型[J]. 软件学报, 2013 (8): 1937-1946.
- [14]Tsuen Ho Hsu, Li Chu Hung, Jia Wei Tang. A hybrid ANP Evaluation Model for Electronic Service Quality [J]. Applied Soft Computing Journal, 2011 (1): 72-81.
- [15]肖珑, 章琳. 引进数据库的发展趋势与价格成本策略[J]. 大学图书馆学报, 2015 (1): 5-13, 20.
- [16]向林芳. 高校图书馆电子资源投入产出绩效评价体系构建[J]. 图书馆建设, 2010 (10): 94-97.
- [17]贺秀英, 王晓文, 呼翠侠. 基于模糊语义法的高校图书馆电子资源绩效评价研究[J]. 情报理论与实践, 2016 (2): 113-115.
- [18]蒲筱哥, 胡亚敏. 图书馆电子资源绩效多指标综合评价方法评析[J]. 图书情报工作, 2013 (7): 14, 139-143.
- [19]严安, 严亚兰. 高校图书馆电子资源持续使用意愿影响因素实证研究[J]. 图书馆论坛, 2013 (3): 30, 43-50.
- [20]Kassim Ahmad Rafee Che, Kochtanek Thomas R. Designing, Implementing, and Evaluating an Educational Digital Library Resource [J]. Online Information Review, 2003 (3): 160-168.

王 力 西安邮电大学图书馆, 工程师, 硕士。研究方向: 数字图书馆及计算机应用。