文 献 与 数 据 学报
第 3 卷　第 4 期　　2021 年 12 月

智慧图书馆的文献资源描述再造框架 *

李玉海[1, 2]　田栩冉[1]　王常珏[1]
（1. 华中师范大学信息管理学院，武汉　430079；
2. 中国图书馆创新发展研究中心，武汉　430079）

摘　要：[**目的 / 意义**] 为提高文献资源的利用率，满足用户个性化的需求，针对智慧图书馆时代多种形式的文献资源，探讨基于本体构建的文献资源描述再造框架，为文献资源的多场景利用提供基础。[**方法 / 过程**] 梳理了智慧图书馆时代的文献资源类型及其智慧特性，借鉴素描绘画中“点－线－面－体”技法，结合现代信息处理技术的最新进展，提出“本体构建－资源描述－关联数据－知识图谱”的文献资源描述再造框架。[**结果 / 结论**] 基于本体构建的文献资源描述再造框架，具备多元知识互联结合的底层基础，可以支持智慧图书馆时代各类文献资源的跨场景运用。

关键词：智慧图书馆　文献资源描述　本体　知识互联
分类号：G253
DOI：10.31193/SSAP.J.ISSN.2096-6695.2021.04.06

0　引　言

随着以人工智能技术为代表的计算机技术与图书馆融合相交，图书馆迈入了 Lib4.0 时代——智慧图书馆（smart library）时代。传统的服务和技术已然跟不上用户快速变化的、场景导向的个性化需求。一方面，在这样一个充斥着各类信息的时代，用户常常会淹没在信息海洋中而找不到所需的真正有效信息；另一方面，传统的文献资源描述已无法适应多样化的资源形态。如何以更加智能的方式识别不同形态的资源内容、精准快速地满足用户的个性化需求，是智慧图书馆时代的重要命题[1]。

* 本文系国家社科基金重大课题“新时代我国文献信息资源保障体系重构研究”（项目编号：19ZDA345）研究成果之一。
[作者简介] 李玉海（ORCID：0000-0003-2256-0183），男，博士生导师，教授，研究方向为智慧图书馆，E-mail：yhli@mail.ccnu.edu.cn；田栩冉（ORCID：0000-0002-1414-7965），男，硕士研究生，研究方向为智慧图书馆，E-mail：tianxuran@mails.ccnu.edu.cn；王常珏（ORCID：0000-0001-9055-9796），女，博士研究生，研究方向为智慧图书馆，E-mail：49690205@qq.com。

1 智慧图书馆时代的文献资源及其描述

传统对文献资源的利用是建立在对相应文献资源的信息描述之上的，其核心原理是，将文献资源作为描述对象，基于一定的规则和标准，对其内容特征和形式特征进行描述，形成一条有关该资源的书目数据记录[2]。这就是传统的基于 MARC（Machine Readable Catalog，机器可读目录）格式的信息描述，是一种高度结构化、单维度的文献资源描述方法，使得对传统文献资源进行描述所需要的“一组特征”得以实现。

然而，文献资源经过多年的演化发展，已经拓展了更加广泛的内涵与意义。我国于 1983 年颁布的《中华人民共和国国家标准·文献著录总则》中指出，文献是记录有知识的一切载体。国际标准化组织于 1997 年 11 月给文献赋予了新的定义，认为文献是在文件的处理过程中，可作为一个单元处理的被记录的信息，而不论其现实的物理形式与特征。从这个意义上来说，文献资源应该包含一切以文字、图形、符号、声频、视频等方式记录在各种载体上的知识和信息资源。

传统图书馆收藏的文献资源，大致包括图书、会议记录、报纸期刊等连续出版物、学位论文、专利文件、政府出版物等，按照载体形式可以分为刻写型文献资源、印刷型文献资源、微缩型文献资源和视听型文献资源。

随着信息技术和图书馆服务理念的发展，在智慧图书馆时代，许多学者针对图书馆的文献信息资源提出了不同的思考。MARKUS AITTOLA 等人认为智慧图书馆是不受时空限制、能够感知的移动图书馆服务，可以通过连接互联网，帮助用户查找图书馆中的图书和其他类型资料[3]，此处的文献资源主要着眼于图书资料。夏立新等人认为，要想重构智慧图书馆的服务模式，需要以图书馆智慧服务策略为基础，将“人”“资源”与“空间”三者的内在逻辑进行深化与融合，进而发挥此三要素的协同促进作用[4]，此处将“资源”视为智慧图书馆的重要组成部分。王世伟认为智慧图书馆是一个书书相连的图书馆，其中的“书”是一个抽象概念，包括各类载体的多媒体文献[5]。

如此看来，相比传统图书馆，智慧图书馆时代的文献资源形式更加多样丰富，具体形式包括但不限于：

（1）口述资料。与文字资料和实物资料类似，口述资料也是人类知识的重要表现形式[6]。从便携式磁带录音机到现在的多功能录音设备，口述资料的记录与收集已经变得十分便捷。并且随着录像技术的发展，资料的采集并不仅仅停留在单一的口述形式，声像俱佳的视频记录方式逐渐增加。图书馆还可通过独立或与其他机构合作的方式参与口述历史和口头传统方面的数据挖掘，来“产生”和“创造”新知识[7]。

（2）开放获取资源。为了应对数据商“卡脖子”的问题，近年来开放获取运动兴起。2003 年，关于科学和人文科学知识开放获取的“柏林宣言”中提到，开放获取的内容应该是经过科学界专业人士认可并通过审查的人类知识性的综合性信息资源，具体包括原始实验数据、数字格式的图像和音视频、原创性科研成果、多媒体学术资源、元数据等[8]，而这些开放获取的资源往往较为分散，难以进行多学科的集中利用。

（3）UGC。UGC 即用户生成内容（user-generated content），泛指由网络用户自由创作的，以多种表达形式在开放网络平台上发表上传分享的语言文字、图片、音视频等开放性多媒体内容[9]。UGC

是当今 web2.0 时代最鲜明的时代特征，其中不乏具有科普性质和教育意义的内容。但由于用户在产生内容时的动机、时间因素、个人知识背景不同，从而导致内容的粒度、知识密度具有很大差异。目前 UGC 在不同网络社区相对较为集中并且呈现出各自社区的特征，ARMSTRONG 等人将网络社区分为关系导向型、兴趣导向型、娱乐导向型和交易导向型[10]，而跨社区的内容交互显然成为了难题。

（4）数字人文。近年来，运用计算机和多媒体等新兴技术开展人文领域研究的数字人文（digital humanities，DH）成果显著，包括但不限于古籍图书扫描版、虚拟古建筑模型、数字地图、计算机图像、在线网页、虚拟人物、各类在线专题音视频数据库等原生数字（born digital）资源[11]，涉及文学、语言学、音乐、艺术、建筑、历史等多个学科。

（5）专利文献资源

专利文献是技术信息最有效的载体，并且大部分的发明创造只通过专利文献这一种形式公开发表，故专利文献资源实效性、技术性更强，创新程度更大，通过对专利文献资源的挖掘可以快速了解相应的前沿技术信息。目前已有对专利文本进行计量分析，并对相关领域进行专利评价的研究，但还较少对专利信息进行关联共享。

智慧图书馆时代，需要将传统的文献资源与口述资料、开放获取资源、UGC、数字人文和专利文献资源等新时代文献资源结合在一起，形成包含各种资源类型与形式、横跨各个学科领域的立体化的文献资源体系。与此同时，智慧图书馆时代的用户服务，需要具有一定的智慧性，如：感知性、专业性、可变性、传递性、有用性、启发性。

（1）感知性。用户在进行资源检索时，并不一定充分了解自己的信息需求，智慧图书馆可以根据用户个人的知识背景、相关的历史浏览和点击记录，感知用户需求，推荐资源内容。

（2）专业性。指对检索结果的优化，在提高检全率的同时保障检准率。

（3）可变性。智慧图书馆最终呈现出来的知识成果并不是一成不变的，可以根据用户的需求和场景，利用相应关联的资源，自动组合转换成不同的媒介形式，如可以将简单的文字知识内容转变为图片式的知识内容。

（4）传递性。检索结果可以通过微信、微博、通讯录等实时传递给身边的人员，也可以传递到如打印机、音箱、大屏幕等智慧终端设备，以实现知识在人与物之间的充分流转。

（5）有用性。通过知识关联等技术操作，深入挖掘与关联文献资源的价值，以提高文献资源的利用率，从而使得智慧图书馆时代的文献资源具有高利用价值。

（6）启发性。通过关联推送最新的前沿知识，启发用户进行更为创新的探索。

纵看近几年的研究，如何智能地感知用户需求，从泛在的新型文献资源中提取有价值的信息，以便提供精准、满意的服务，是智慧图书馆建设的重要目标和挑战。为此，需要针对智慧图书馆时代的主要文献资源，以赋予其以上六大智慧特性为要义，构建文献资源描述再造的框架，以期实现文献资源的服务个性化与利用智能化。

2　文献资源描述再造框架

现如今，文献资源保存与流通的主要趋势是数字化，包括传统的纸质文献与其他新型网络资

源，智慧图书馆时代的文献资源基本上是以数字形式存在的。与传统图书馆相同，对文献资源加以利用，首先需要对其进行描述与组织。

数字图书馆时代主流的文献资源描述框架是元数据。目前各大主流的图书馆文献资源管理系统都采用都柏林核心元数据集（Dublin Core Element Set，简称 DC 元数据集），以实现对文献资源的数字化描述。但是 DC 元数据的构建主要是基于人工操作，从叙词表的构建到分类方法，基本都需要以人的理解认知为基础，描述的过程中还会存在诸如别名问题、歧义问题、关联推理问题和知识复用等局限[12]，文献之间也难以相互关联，无法满足诸如知识推送等智能服务需求。

为此，本文综合目前已有的一些计算机技术，模仿素描绘画技法中的“点 – 线 – 面 – 体”，提出“本体构建 – 资源描述 – 关联数据 – 知识图谱”这一文献资源描述再造框架（图 1），尝试构建具有智慧功能的文献资源基础架构。

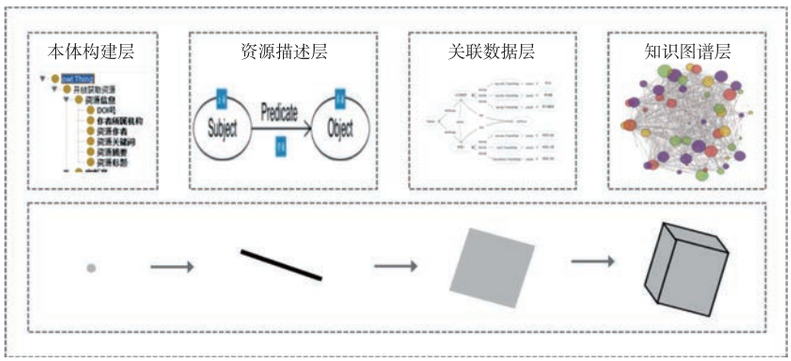

图 1　文献资源描述再造框架

2.1　本体构建——点

本体的概念起源于哲学领域，指“对世界上客观事物的系统描述，即存在论”，后被引入信息科技领域。GRUBER 认为，本体是使程序和人们共享知识信息的概念模型的规范说明[13]；后来，BORST 进一步指出，本体是一种形式化的，对于共享概念模型详细而又明确的说明[14]，其中，概念模型指对客观现实世界中部分现象的相关概念进行抽象而得到的模型，明确是指所抽象的概念以及使用这些概念都具有明确的约束和定义，形式化是指本体是计算机可处理的，共享是指本体中体现的是被一个团体共同认可的知识[15]。

简单来说，本体提供了一种结构化的可被机器理解的共享词表，一般可以用来针对该领域的属性进行推理，亦可用于定义该领域。本体设计的常用方法有 IDEF–5 法、TOVE 法、七步法等。其中，由 NOY N F 等人提出的七步法[16]在构建领域本体的方法中比较具有代表性，具体步骤包括：（1）明确本体的范围和适用领域；（2）考虑本体的重用；（3）列举领域中的重要术语；（4）定义类和类层次结构体系；（5）明确类的属性；（6）明确属性的分面；（7）实例的构建。

针对智慧图书馆背景下主要的五类文献资源，进行简要的领域区分和等级构建，形成不同的父类及其子类（图 2）。以专利文献资源为例，“专利详细信息”“专利摘要”“专利权人”“专利信息”为“专利”本体这一父类的子类，同时又包含若干下级子类。

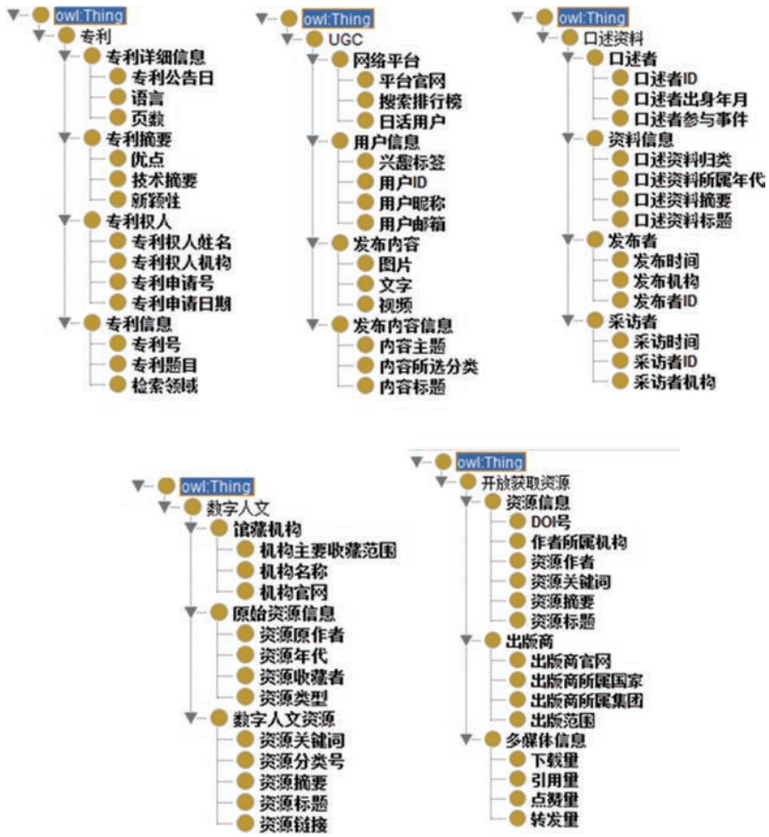

图 2　五类文献资源本体等级

通过这一本体构建，为后续智慧图书馆文献资源的综合利用提供基础：（1）在资源检索方面，可以从基于关键字的检索，上升到语义检索，用户的检索请求可以自动转换成语义检索的规定表达式，在本体数据集中匹配出符合条件的信息。（2）在资源集成方面，传统多媒体资源的描述方法不利于进行有效的集成和检索。借助本体构建，在进行语义对齐等操作后，可以在一定程度上解决语义异构的问题，将多种类型的文献资源信息关联起来，具有良好的扩展性和适应性。（3）在跨语言检索方面，在本体构建的过程中，会通过把各种语言中的词汇映射到同一个本体中，进而实现歧义的消解和检索结果页面的自动翻译。

2.2　资源描述——线

资源描述框架（Resource Description Framework，RDF），作为 XML（Extensible Markup Language）的一种衍生版本，是国际互联网协会（World Wide Web Consortium，W3C）推荐的用于描述和处理元数据的方案。RDF 定义了一个简单的模型，用于描述资源、属性和值之间的关系。其中，资源可以是能被统一资源标识符（Uniform Resource Identifier，URI）标识的所有事物；属性是资源的一个特定的特征；值可以是字符串，也可以是另一个资源。简单来说，一个 RDF 描述就是一个三元组：< 主语、谓词、宾语 >[17]。

传统图书馆利用 MARC、DC 等元数据技术对文献资源进行描述，只有一种格式规范，不同

的元数据之间无法兼容与关联，因而难以从语义层面加以利用。

而基于本体构建的资源描述方法，可以针对不同的文献资源本体，提取本体之间的关系，构建具有语义含义的三元组（图 3）。如：口述资料主要抽取“口述者”讲述（Tell）“口述资料”；开放出版资源主要抽取“出版商”分享（Share）“开放获取资源”；UGC 资源主要抽取“用户”发布（Release）“UGC”；数字人文资源主要抽取“馆藏机构”数字化（Digitalize）“数字人文资源”；专利主要抽取“专利权人”拥有（Has）“专利”。其中的几个谓语动词：讲述（Tell）、分享（Share）、发布（Release）、数字化（Digitalize）、拥有（Has），即为联结作为主语和宾语的相应本体之间的“关系”。

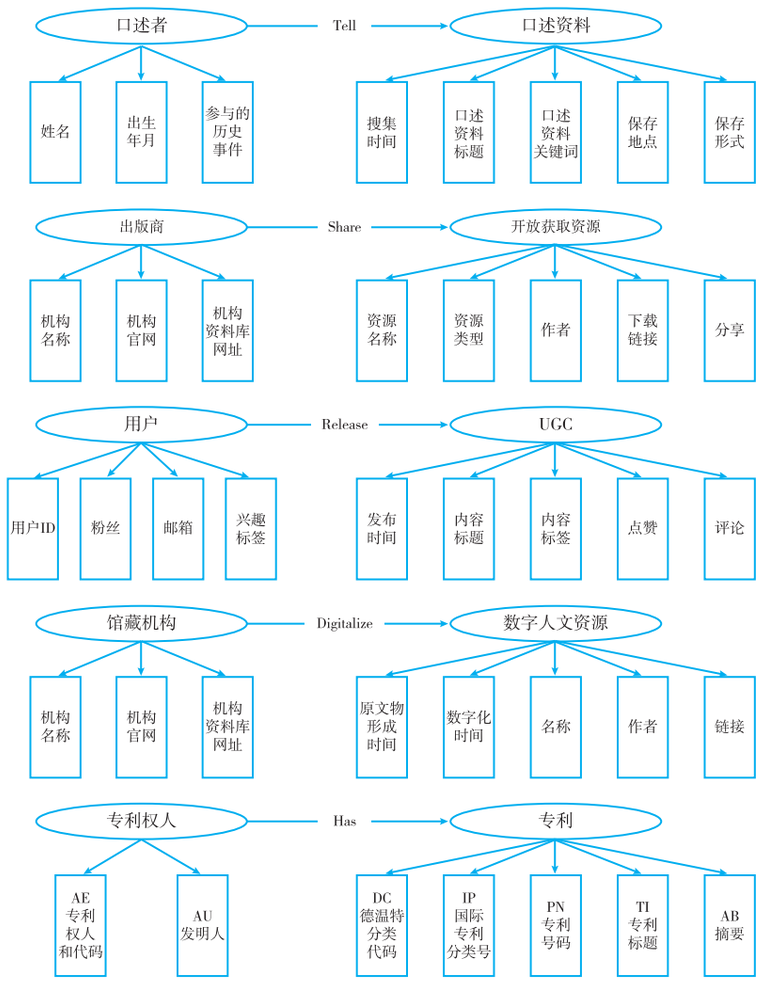

图 3　五类文献资源本体关系描述框架

之后，国际互联网协会推出了新的本体描述语言，RDF 模式语言和网络本体语言。RDF 模式语言（RDF schema，简称 RDFs），是在 RDF 的基础上引入模式层，定义类、属性、关系、属性的定义域与值域来描述与约束资源，构建最基本的类层次体系和属性体系，支持简单的上下位推理[18]。网络本体语言（Ontology Web Language，简称 OWL）进一步扩展 RDFs 的词汇，可声

明类间的互斥关系、属性的传递性等复杂语义，支持基于本体的自动推理，提供了一组适合 web 传播的描述逻辑的语法[19]。

2.3　关联数据——面

虽然本体的构建可以实现对文献资源的语义及其之间较为显性的关系（如 part-of 部分与整体的关系、kind-of 继承关系、instance-of 实例和本体间的关系、attribute-of 属性关系）的描述，但要将其他多媒体信息资源关联起来揭示一定的隐性关系还需要借助关联数据。关联数据（Linked Data），是指在语义网上发布、共享、连接各类数据、信息和知识的一种方式。关联数据包含三层要素：数据内容、描述数据内容的元数据格式、基于相应元数据格式转换为 RDF 格式发布的数据内容（或数据内容的描述信息集）[20]。

关联数据通过 URI 的方式表示并存取“资源”。URI 除了能够唯一标识资源对象之外，还能起到定位的作用，从而能够将互联网上异类（即具有不同内容类型和体裁，例如小说、戏剧、论文、专著、数据记录、微博消息、财务报表、博客文章等）、异构（即具有不同的数据格式及相应的语义规则）和分布的数据进行有效的关联[21]。如果这个资源是已经有 URI 标识的信息资源，则可以直接通过传统的 Web 方式获取；如果是还未有 URI 标识的信息资源，则链接到一个以 RDF/XML 编码的、用以指代该资源的数据文件[22]。

目前较为常用的关联数据发布是借助 RDF 映射平台——D2RQ，借助核心映射机制，可以从关系数据库中，将概念或链接一致但 RDF 描述不同的数据进行映射关联[22]，并通过 D2R Server 对关联数据进行发布（如图 4）。

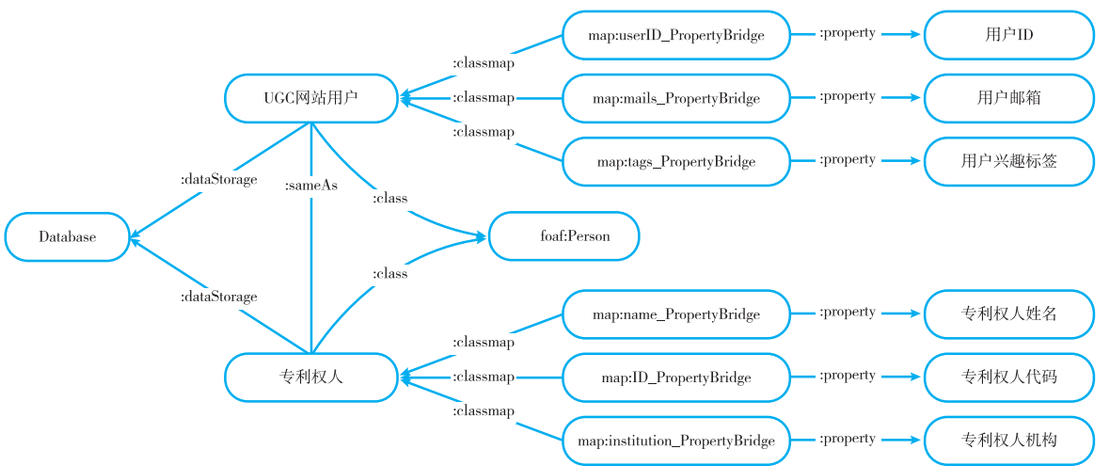

图 4　文献资源关联数据示意图

为了更好的利用与共享，还需要建立能够存储三元组关系的数据库，图数据库便应运而生。图数据库源起欧拉和图理论，是 NoSQL（Not Only SQL，不限于 SQL）中的一类数据库，是以“图形化的模型”这种数据结构来存储与查询，但并非存储图片的图像数据库。图数据库的数据模型主要是以节点和边（即节点之间的关系）来体现，具有丰富的关系表示和完整的事务支

持[23]，是快速处理复杂知识互联问题的底层基础。

2.4 知识图谱——体

通过以上描述组织步骤之后，最终主要是以知识图谱的形式呈现在用户面前。知识图谱（knowledge graph）为一种结构化的语义知识库，是对物理世界概念的抽象及其相互关系的描述，“实体 – 关系 – 实体”三元组是其基本组成单位，通过关系联结不同实体，构成知识网络，可以快速实现知识的响应和推理[24]。通过浏览知识图谱，能够以生动简明的图形方式向用户传递具有联接关系的知识，故而用户不必浏览大量网页，可以直接按照主题进行检索，从而实现语义检索并准确定位和获取深度知识。

3 结 语

在智慧图书馆时代文献资源多样化、用户需求个性化、服务模式场景化、国际挑战严峻化等一系列背景之下，借鉴素描绘画技法中的“点 – 线 – 面 – 体”，构建“本体构建 – 资源描述 – 关联数据 – 知识图谱”的文献资源描述再造框架，将各种不同类型的文献资源进行关联，为不同场景下的知识互通提供应用基础。

【参考文献】

［1］李玉海，金喆，李佳会，等 . 我国智慧图书馆建设面临的五大问题［J］. 中国图书馆学报，2020,46(2):17–26.

［2］钱鹏，郑建明 . 基于资源描述框架的图书馆科学数据组织初探［J］. 情报理论与实践，2012,35(3):100–102,108.

［3］AITTOLA M,RYHäNEN T,OJALA T. SmartLibrary–location–aware mobile library service［C］. International Conference on Mobile Human–Computer Interaction,2003:411–416.

［4］夏立新，白阳，张心怡 . 融合与重构：智慧图书馆发展新形态［J］. 中国图书馆学报，2018,44(1):35–49.

［5］王世伟 . 未来图书馆的新模式——智慧图书馆［J］. 图书馆建设，2011,(12):1–5.

［6］王子舟，尹培丽 . 口述资料采集与收藏的先行者——美国班克罗夫特图书馆［J］. 中国图书馆学报，2013,39(1):13–21.

［7］尹培丽 . 口述资料收藏——图书馆的新领地［J］. 大学图书馆学报，2013,31(4):14–18.

［8］魏蕊，初景利 . 学术图书开放获取与美国大学图书馆出版服务［J］. 大学图书馆学报，2014,32(3):17–22.

［9］赵宇翔，范哲，朱庆华 . 用户生成内容（UGC）概念解析及研究进展［J］. 中国图书馆学报，2012,38(5): 68–81.

［10］ARMSTRONG A,HAGEL J. The real value of online community［J］. Harvard Business Review,1996,74(5):134–141.

［11］刘炜，叶鹰 . 数字人文的技术体系与理论结构探讨［J］. 中国图书馆学报，2017,43(5):32–41.

［12］夏天，钱毅 . 面向知识服务的档案数据语义化重组［J］. 档案学研究，2021,(2):36–44.

［13］GRUBER T R. Toward principles for the design of ontologies used for knowledge sharing?［J］. International Journal of Human–Computer Studies,1995,43(5):907–928.

［14］BORST W N,BORST W N. Construction of Engineering Ontologies for Knowledge Sharing and Reuse［D］. University of Twente:Centre for Telematics and Information Technology (CTIT),1997.

［15］STUDER R,BENJAMINS V R,FENSEL D. Knowledge engineering:Principles and methods［J］. Data & Knowledge Engineering,1998,25(1):161–197.

［16］NOY N F,MCGUINNESS D L. Ontology development 101:A guide to creating your first ontology［J］. Knowledge Systems Laboratory,2001,32(1):1–25.

［17］邹磊 , 彭鹏 . 分布式 RDF 数据管理综述［J］. 计算机研究与发展 ,2017,54(6):1213–1224.

［18］MCBRIDE B. The resource description framework (RDF) and its vocabulary description language RDFS［J］, Handbook on Ontologies,2004:51–65.

［19］MCGUINNESS D L,VAN HARMELEN F. OWL web ontology language overview［J］. W3C Recommendation, 2004,10(10):1–12.

［20］沈志宏 , 张晓林 . 关联数据及其应用现状综述［J］. 现代图书情报技术 ,2010(11):1–9.

［21］EISENBERG V,KANZA Y. D2RQ/update:updating relational data via virtual RDF［C］. Proceedings of the 21st International Conference on World Wide Web,2012:497–498.

［22］刘炜 . 关联数据：概念、技术及应用展望［J］. 大学图书馆学报 ,2011,29(2):5–12.

［23］欧石燕 . 面向关联数据的语义数字图书馆资源描述与组织框架设计与实现［J］. 中国图书馆学报 ,2012,38(6):58–71.

［24］刘峤 , 李杨 , 段宏 , 等 . 知识图谱构建技术综述［J］. 计算机研究与发展 ,2016,53(3):582–600.

Reconstruction Framework of Literature Resources Description in Smart Libraries

LI Yuhai[1,2]　TIAN Xuran[1]　WANG Changjue[1]

(1. School of Information Management,Central China Normal University, Wuhan 430079, China;
2. China Library Innovation and Development Research Center, Wuhan 430079, China)

Abstract:［**Purpose/significance**］In order to improve the utilization rate of literature resources and meet the personalized needs of users,this paper discusses the reconstruction framework of literature resources description based on ontology,aiming at various forms of literature resources in the era of smart library,and provides a foundation for multi-scene utilization of literature resources.［**Method/process**］The literature resource types and characteristics in the era of smart library are summarized. Referring to the technique of "point-line-face-body" in sketch painting and combining with the latest progress of modern information processing technology,the literature resource description reconstruction framework of "ontology construction - resource description - linked data - knowledge graph" is proposed.［**Result/conclusion**］The ontology-based literature resource description reconstruction framework,with the underlying foundation of multi-knowledge interconnection,can support the cross-scene application of various literature resources in the era of smart library.
Keywords: Smart library; Literature resources description; Ontology; Knowledge interconnection

（本文责编：王秀玲）