学科领域关联词汇集构建研究*

常　娥，孙文佳

摘　要　为了提供规范的资源属性、概念取值和关联类型描述，文章以间质性疾病领域为实验对象，构建了包含元数据元素集和取值词汇集在内的领域关联词汇集。首先，借鉴领域已有的词表、类表和规范文档等，创建了通用关联词汇集；其次，采用 N-gram 统计分词、命名实体识别、模式识别等多种技术方法，构建了领域核心关联词汇集，以更好地引出和关联该主题领域的各种资源与数据。

关键词　关联词汇集　元数据　知识本体　知识组织

引用本文格式　常娥，孙文佳. 学科领域关联词汇集构建研究[J]. 图书馆论坛，2016（8）：13-19.

Research on Construction of the Domain Linked Vocabulary

CHANG E，SUN Wen-jia

Abstract　In order to provide standard description of resource concepts，attributes and relationship，the domain linked vocabulary of interstitial disease is constructed in this article，which contains metadata element set and the vocabulary of domain concepts. Firstly，the domain general linked vocabulary is created using the existing thesaurus，classifications and normalized documents for reference. Secondly，the domain core lined vocabulary is built to link different resources and data in the field by using N–gram to count the word segmentations，to name the entity recognition and the pattern recognition.

Keywords　linked vocabulary；metadata；ontology；knowledge organization

0　引言

自英国学者 John Taylor 提出科学研究信息化(e- Science)概念后[1]，经过 10 多年的发展，世界各国已建成规模庞大的各种科研信息化平台和研究数据存储中心。近 10 年来，我国也投入大量资金，积极建设云计算、移动互联、物联网、数字图书馆，以及科学数据存储中心等项目，极大地推进了我国科学研究信息化工作。然而如今 e- Science 的发展已到了瓶颈期，它虽突破了传统科研封闭、重复劳动等局限性，却仍无法有效解决科研的可重复性以及科学数据共享问题[2]。为此，德国学者于 2011 年提出了关联科学(Linked Science)这一概念。关联科学指的是全面关联与共享包含出版物、研究方法和研究数据等在内的科学资源，以支持透明、可重复和跨学科的研究[3]，被认为是 e- Science 发展的新阶段。

关联科学是为组织科学资源而提出的一种新理论，与未来图书馆资源组织的目标一致。图书馆资源组织研究经历了分类法、主题法、元数据、知识本体、关联数据等多种知识组织方法

* 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

后，学者们认为，未来馆藏资源组织必将转向以关联数据为核心的细粒度、开放、关联与语义化的资源组织模式，并广泛关联图书馆外部网络资源。无论是关联科学，还是未来馆藏资源组织，两者都需深入研究关联数据、语义网、云计算等技术如何解决各种科学资源的组织、关联与发现。众所周知，关联数据仅是一套在网络上发布数据的原则，类似于行动指南，对于实际应用领域，需设计出具体数据关联模型才能有效组织资源。

综合借鉴各种知识组织理论与方法，本研究提出了3层架构的图书馆资源底层通用的整体数据关联模型(Bottom Common Organization Model of the Whole Library Knowledge Resource，BCOM)，可满足关联科学的所有愿景，无论是从资源出发，还是从知识单元出发，都能迅速关联并发现用户所需的知识或资源，因此增强了馆藏资源数据关联网络的整体连通性。无论是本体模型，还是元数据模型，最后都可形成一组组词汇集合，BCOM模型亦不例外。总体而言，目前BCOM模型仅是一个普适性的资源组织概念框架，针对不同应用领域，需建立更加具体的学科领域关联词汇集，以提供规范的资源属性、概念取值和关联类型描述，从而深刻揭示与表达馆藏资源、知识单元各自及其之间的复杂语义关系。

1 相关领域研究进展

1.1 图书情报领域

近年来图书馆界在开放数据网络中发布了大量的包含书目、分类词表、规范主题词、规范人名等资源在内的关联数据集，并积极寻求新一代资源组织框架，书目记录的功能需求FRBR是当前讨论最为热烈的馆藏资源组织模型。FRBR框架采用实体关系(E-R)模型，突破了传统MARC书目数据的线性资源模式。然而，由于FRBR模型的网状结构完全颠覆了编目员传统的思维模式，无法适应编目工作便利性的实践要求，并且它只是一个概念架构，新一代资源描述标准(Resource Description and Access，RDA)便应运而生，以支撑该框架的具体实施。RDA实际上是对FRBR模型中实体概念、属性，以及实体间关系的集中序化和表达，从而形成的一套附加了资源描述和检索的原则与说明的元数据词汇表。

由于FRBR模型W(作品)、E(内容表达)、M(载体表现)、I(单件)分层的抽象与复杂性，让人难以理解，目前作为其简化模型的BIBFRAME框架备受关注，MARC、RDA、DC、VAR等业界标准均是其术语来源[4]。因此，FRBR模型本身并不完善，它将不断地被修补甚至完全改变，或许会成为一个过渡性工具，然而最基本的RDA元素集及其基本含义构成了图书馆资源组织的基础[5]。RDA词汇集详见国际图联(IFLA)网站[6]。传统知识组织领域已经积累的大量的诸如RDA词汇集、分类词表、主题词表等各种知识组织术语词汇表为本文领域关联词汇集的构建奠定了基础。

1.2 关联科学领域

关联科学的概念被提出后，研究者们设计了关联科学核心词汇规范(Linked Science Core Vocabulary Specification，LSCVS)，以提供概念将科学研究中涉及的研究人员、数据、方法、假设、结果、出版物等互联起来，并将这些事物同时间、空间以及主题联系起来[7]。LSCVS的概念类包括研究者、研究、数据、假设、预测、出版物、结论、地区和时间，属性关系包括“参与”“产生”“利用”“测验”“证实”“证伪”等17种。LSCVS旨在为科学研究定义基本的概念或类，是一个轻量级、简单的词汇集，需要在此基础上进行扩展，以满足不同学科领域研究需要的更为具体的概念、类别和属性关系。总体而言，LSCVS属于通用词汇集范畴，侧重于从内容层面揭示科学研究出版物中的知识单元及其语义关系。

1.3 数据管理领域

目前元数据是科学数据管理的主要方式，在科学数据的保存、检索与利用中发挥了重要的作用[8]。学术界普遍认为，描述科学数据的元数据项越丰富，越有利于用户共享、发现与再利用科

学数据[9]。各学科领域构建了复杂的科学数据元数据标准，如 FGDC(地理空间)、DDI(社会科学)、EML(生态学)、CF(气象学)。在实践应用领域，科学数据元数据的学科通用性与专指度始终无法得到统一，目前解决方案主要有两种：(1)忽略科学数据的学科差异性，构建通用元数据标准。国际上的典型代表是 DataCite[10]，其核心必备元素包含：识别符、创作者、题名、出版者和出版年。此外，学者们还总结出描述科学数据生命周期中所有特征的元数据维度。例如，Ball 等人认为应包含 6 类元数据信息：标识、责任、数据存档、主题覆盖和派生、时间与空间覆盖、来源[11]；张华等人则总结出科学数据元数据的 7 个重要属性维度：主题内容、存储、责任、质量、使用与评价、来源、关联[12]；(2)构建元数据之元数据，以实现不同学科领域科学数据的集成与共享问题。元数据编码和传输标准 METS 是这一解决方案的典型代表，它可将数字资源库中各种形式的元数据进行打包，目前已成为诸如 Hathitrust 等大型数字图书馆项目资源组织的基础[13]。

除元数据外，生物医学领域还构建了知识本体来组织与关联科学数据，如 Tan 等人构建了生物芯片本体[14]，徐坤等人构建了中医药数据本体[15]。通过本体构建，将科学数据存在于一个知识环境中进行完整描述。

2 学科领域关联词汇集的构建

BCOM 模型包含 3 层结构：资源层、中间层和知识层。资源层显示了粗粒度资源与资源之间的关系，主要表现为引用和同现关系，学科领域差异性较弱；中间层描述了资源与知识单元之间的关系，由于学术论文通常具有固定的内容框架结构，因此学科差异性亦较弱；知识层则揭示了细粒度知识单元之间的关系，由于不同学科领域的知识概念及其语义关系千差万别，因此学科领域差异性显著。

有鉴于此，从词汇通用性来看，学科领域关联词汇集包含领域通用关联词汇集和领域核心关联词汇集两种类型。其中，领域通用关联词汇集主要建立在 BCOM 模型的资源层和中间层，与具体应用领域无关，提供通用的资源属性、概念取值和关联类型描述，而领域核心关联词汇集则建立在知识层，与具体应用领域相关，提供该领域特有的资源属性、概念取值和关联类型描述，以更好地揭示与关联该主题领域的各种资源与数据。

此外，从词汇功能上看，学科领域关联词汇集包含属性元素集和取值词汇集两种类型。其中，属性元素集的功能在于建立资源、概念各自及其之间的语义关系，取值词汇集的功能在于提供资源概念、实例的各种表达形式。

2.1 资料来源

本文将以间质性疾病主题领域为例，尝试构建 BCOM 模型框架下的学科领域关联词汇集。间质性疾病涉及呼吸病学、风湿病学、心肺血管学，以及病理学、影像学等多个学科领域，形成了以疾病为中心的多学科联合研究模式，内容丰富而新颖，是医学领域的前沿课题之一。我国对于间质性疾病的研究，除了借鉴西医技术外，还较多地融入了中医理论与方法，从而形成具有中医特色的间质性疾病研究成果。因此，本研究将重点收集国内间质性疾病领域的研究资料，尝试构建其中文关联词汇集，以探索 BCOM 模型的具体应用方法与实施路径。学科领域关联词汇集的构建语料主要来源于以下 3 个方面：

(1)国内知识组织领域已有的各种词表、分类表、元数据标准等，以及国内生物医学领域特有的叙词表、分类表、术语表等词汇表中与间质性疾病有关的概念、词汇、术语等，如《中文医学主题词表》(CMeSH)、《中国图书馆分类法》(医学专业分类表)。

(2)国内间质性疾病相关研究论文。首先分别从中国知网(CNKI)、重庆维普(Cqvip)以及中国生物医学文献数据库(SinoMed)，下载、合并及去重后，获得间质性疾病研究论文题录 11245 篇，其中核心期刊 4948 篇，数据采集日期截止到 2015 年 9 月底。考虑到研究语料应尽可能权威和完整，本文选择了 4948 篇核心期刊论文题录，以及 676 篇期刊论文全文作为研究语料，被选论

文全文要求至少同时出现在3大数据库中的某两个数据库中。

⑶美国国立医学图书馆的统一医学语言系统(Unified Medical Language System，UMLS)。UMLS提供了生物医学领域最为丰富的知识源，包括3个核心部分[16]：①超级叙词表，词汇来源于生物医学领域上百种叙词表、分类法、代码集等，具有空前的广泛性、异构性和多语言性；②语义网络，为超级叙词表中的概念、术语、词汇等提供统一分类体系，并定义它们之间的语义关系；③专家词典，主要用于自然语言处理，以生成、规范生物医学领域词汇。

由于UMLS并不是高度概念化的生物医学领域本体，因而无法保证它在某个领域中的应用是最优的，需要根据特定任务进行个性化定制[17]。国内以UMLS为基础构建了中文一体化医学语言系统(Chinese Unified Medical Language System，CUMLS)，然而CUMLS还未完全实现统一结构体下的词表整合机制，其开放性和成熟度远不如UMLS[18]。因此，本研究仍以UMLS为基础，下载了超级叙词表和语义网络作为本文的另一项重要研究语料。

2.2 领域通用关联词汇集构建

在BCOM模型的资源层和中间层，领域通用关联词汇集只需提供资源属性与关系类型描述，无需提供知识概念及其不同取值，因此其实质是构建由属性关系构成的元数据元素集。资源层主要以同现、引文为基础建立文献之间的各种关联；中间层主要利用学术元数据和书目元数据建立文献与知识单元之间的关联。

同现关联实现路径主要包含作者同现、机构同现、期刊同现和关键词同现这4种方式。引文是学术文献特有的结构，然而为进一步揭示引文之间的语义关联，需深入研究引文动机，并对其进行分类。目前，一般从引用功能和观点倾向两个角度出发确定引文的类别，但引文类型定义仍未形成统一标准[19]。本研究根据学术论文写作方式的规律性，结合学术文章的内容结构，从引文内容、引文位置等方面对引文功能进行分析，将其分成背景引用、数据引用、方法引用、理论引用、观点引用、结果引用这6种类型。值得注意的是，引文关联是人为主观选择而建立资源之间的主题关联，而关键词关联则是资源之间客观存在的主题关联。换言之，资源之间若存在引文关联，则一定存在关键词同现关联，而存在关键词同现的资源，则不一定存在引文关系。

中间层学术元数据实质上是学术文章内容的另一种结构化表达，同时反映的亦是资源与知识单元的一种映射关系。本研究已撰文详细论述了资源与数据的关联特征，此处不再赘述。本文通过广泛调研学术文章的篇章结构，抽取、统计、分析并总结出一般性表达结构，形成了学术元数据框架中4种基本属性关系：研究问题、研究方法、研究数据和研究结果。其中，研究问题包含研究背景、研究目的、研究意义等元素项；研究方法包含研究方法、研究材料、研究过程等元素项；研究数据包含原始数据、校准数据、验证数据等元素项；研究结果包含实验结果、研究结论、未来工作等元素项。

中间层书目元数据旨在描述文献信息资源的一般外部特征。迄今为止，MARC是描述文献信息资源特征最为详尽的通用元数据标准，因此本文主要参考MARC元数据标准，同时综合考虑BCOM模型中各种关联属性的统一性，为中间层书目元数据定义了8种核心元数据项，分别为：题名、作者、机构、语种、国别、日期、类型和统一标识符。资源层和中间层中领域通用的元数据元素集详见表1。

2.3 领域核心关联词汇集构建

表1 领域通用关联词汇集

	属性元素
资源层关联属性	引用(背景引用、数据引用、方法引用、理论引用、观点引用、结果引用)
	同现(同作者、同机构、同期刊、同关键词)
中间层学术元数据	研究问题(研究背景、研究目的、研究意义等)
	研究方法(研究方法、研究材料、研究过程等)
	研究数据(原始数据、校准数据、验证数据等)
	研究结果(实验结果、研究结论、未来工作等)
中间层书目元数据	题名、作者、机构、语种、国别、日期、类型、统一标识符(ISBN、ISSN、DOI等)

知识层完全由细粒度的知识单元构成，其语义关系最为丰富和复杂，且具有较强领域特点。诸如通用知识本体、学科分类法等通用知识组织模型，由于其知识概念和语义关系抽象层次高，无法满足 BCOM 模型知识层相关概念及其关系细致、具体而深入的表达需求，进而影响 BCOM 模型的整体功能。因此，本研究在知识层构建了领域核心关联词汇集，丰富了领域知识概念及其关系的表达。该词汇集主要包含属性元素集和取值词汇集，其实质是领域本体的另一种表达形式。囿于研究时间和精力，本文将以间质性疾病为例，构建中文领域核心关联词汇集，探索 BCOM 模型知识层概念联通的实现路径。具体构建方法如下：

第一步，获得本领域知识概念的基础取值词汇。首先，在《中文医学主题词表》和全国科学技术名词术语数据库(http：//www.cnctst.cn)中查找筛选出与间质性疾病直接相关的知识概念共 39 个；其次，采用 N- gram 统计分词、命名实体识别、模式识别等多种技术方法，从 4948 篇核心期刊论文题录中，挖掘出更加具体的与间质性疾病相关的其他知识概念；最后，由领域专家对所有相关概念词汇进行筛选并给出 UMLS 语义类型，构建核心取值词汇集。为了包含中医治疗间质性疾病有关知识概念，本研究在 UMLS 材料(substance)语义类型中增加了“中草药”语义类型，在完整的解剖结构(fully formed anatomical structure)中增加了“经络”和“腧穴”两个语义类型。领域核心取值词汇集共包含 4292 个词语，表 2 列举了该词汇集前 10 项词语示例。

第二步，抽取本领域知识概念之间的各种关联关系，构建属性元素集。知识概念间的关系主要包含同义关系、等级关系和相关关系这 3 种类型。比较而言，同义和等级关系较易识别，而相关关系则较难识别。相关关系提供了不同层面知识概念间细致而深入的语义关联。例如，“治疗”关系建立了药物和疾病这两组不同知识概念间的联结；若没有“治疗”关系，药物和疾病仅可能具备各自知识树上的等级关系而已，而无法实现概念间的连通。因此，相关关系抽取是本研究的重点。文献全文是语义关系最丰富的来源，本文以 676 篇间质性疾病研究论文为基础，同时结合 UMLS 中语义关系，抽取间质性疾病领域的语义关系，构建领域属性元素集。具体方法如下：

表 2　间质性疾病领域核心取值词汇集样例

取值词汇	语义类型
阿米替林	临床药物(clinical drug)
阿莫西林	临床药物(clinical drug)
阿莫西林钠	临床药物(clinical drug)
阿奇霉素	临床药物(clinical drug)
阿糖胞苷	临床药物(clinical drug)
阿糖腺苷	临床药物(clinical drug)
非小细胞肺癌	疾病或综合症(disease or syndrome)
鳞状细胞癌	疾病或综合症(disease or syndrome)
支气管原癌	疾病或综合症(disease or syndrome)
癌基因	基因或基因组(gene or genome)

首先，以领域核心取值词汇集为基础，对 676 篇间质性疾病研究论文进行全文扫描，提取至少包含两个取值词汇的句子；其次，以 UMLS 中的 54 种语义关系为识别模式，自动过滤、删除候选句，即若 UMLS 中的语义关系出现在候选句中，则将该候选句删除，并记录语义关系；再次，由人工判别剩余候选句中的语义关系，并进行记录。最后，将所有语义关系合并、去重后，添加进属性元素集中。经研究发现，间质性疾病领域不仅包含了 UMLS 的 54 种基本语义关系，还有所拓展，本文共筛选出 73 种属性元素。为了完整定义这些元素，在属性元素表中给出了各元素的使用范围和取值范围说明，样例详见表 3。

3　结果讨论

在知识组织领域，一直无法有效解决知识组织系统的学科通用性和专指度的问题。例如元数据的学科专指度越高，其通用性就越差，可跨学科使用的可能性越低[20]。反之，元数据的通用性越强，其学科专指度就越差，关联与发现特定领域资源与数据的能力越弱。本文以间质性疾病为主题领域，构建了领域通用关联词汇集和领域核

表 3　间质性疾病领域核心属性元素集样例

属性元素	使用范围	取值范围
…是一个… (is a)	指一个实体或事件属于另一个实体或事件的一种；用于连接具有属分关系的两种知识概念	取值词汇包括生物、解剖结构、解剖异常、制造对象、生物活性物质等类型的概念
是…一部分 (part- of)	指由一个或多个物理单位可组成一些更大的整体；该属性可以表达整体的某个部分、片段或者分层单元；用于连接可以进行分解的物理实体或概念实体	取值词汇包括身体器官、细胞、组织、胚胎结构以及核苷酸序列等类型的概念
由…组成 (consists- of)	指在整体物质中的各组成物质；用于连接具有组成、生成或形成关系的两种概念	取值词汇包括身体器官、有机化学药品、基因或基因序列、氨基酸或蛋白质等类型的概念
包含 (contains)	指持有或包含液体或其他物质的容器；用于连接容器类概念与被包含物概念	取值词汇包括身体空间、身体物质、给药装置、临床药物等类型的概念
连接 (connected- to)	指直接连接到身体物理单元的肌腱或肌肉；该属性可以表达具有连接关系的物理实体类概念	取值词汇包括身体器官、身体空间、身体物质、组织等类型的概念
治疗 (treats)	指引用药物或医疗器械治疗疾病	取值词汇包括医疗器械、药效物质、损伤或中毒、体征或症状、疾病或综合症等类型的概念
破坏 (disrupts)	指对已存在的状态或状况产生负面影响，从而使其发生改变	取值词汇包括生物活性物质、有害或毒性物质、胚胎组织、生理功能等类型的概念
并发 (co- occurs with)	指症状或疾病同时发生或同时存在	取值词汇包括解剖异常获得性异常、生理功能、病理功能、损伤和中毒等类型的概念
预防 (prevents)	指停止、阻碍或消除一个动作或状态	取值词汇包括临床药物、中草药、疾病或综合症等类型的概念
诊断 (diagnoses)	指区别或标识疾病的性质或特征	取值词汇包括医疗器械、疾病或综合症、生理功能等类型的概念

心关联词汇集，尽可能兼顾了知识组织系统的学科通用性和专指度，增强了 BCOM 模型的整体功能。主要体现在以下 3 个方面：

⑴领域通用关联词汇集提供了规范的资源属性，建立了资源与资源、资源与知识概念间的关联。通用词汇集与具体应用领域无关，增强了 BCOM 模型的学科通用性。

⑵领域核心关联词汇集则提供了丰富的领域概念取值和概念关系描述，可以更好地引出和关联来自某主题领域的各种资源与数据。核心词汇集与具体应用领域相关，使 BCOM 模型可满足学科特定知识的组织需求，具有专指性。

⑶学术元数据框架和书目元数据框架关联的知识概念只是知识层概念的一小部分，大量的领域知识概念将由领域核心取值词汇集提供，同时借助领域核心属性元素集在知识层可构建更广泛的知识关联，增强了 BCOM 模型的整体功能。

换言之，知识层的概念节点要明显多于资源可关联到的概念节点，即存在着大量未与资源建立关联的概念节点。但由于概念节点间总是存在着某种关联，通过知识推理，这些概念节点可与资源实现间接连通，其功能类似于主题词的入口词。

4　结语

为了提供学科领域内规范的资源属性、概念取值和关联类型描述，本文以间质性疾病领域为实验对象，构建了包含属性元素集和取值词汇集在内的关联词汇集。该词汇集是构建知识本体的核心要素，其中取值词汇集对应着知识本体的概念类，属性词汇集对应着知识本体的概念间语义关系，利用关联词汇集可以搭建领域知识网络，即知识本体的基本框架。由于关联词汇集暂未涉及领域关系的函数和公理约束条件，而且没有进

行形式化，因此它又不完全等同于知识本体。接下来，本研究将在 BCOM 模型框架下，以关联词汇集为基础，继续深入研究包括 URIs 命名处理、RDF 创建与复用等在内的馆藏资源关联数据集的发布问题。

参考文献

[1] 曾伟忠. 科学研究的信息化：e-Science 的产生和发展[J]. 现代情报，2006（2）：6-8.

[2] 唐义，肖希明. 关联科学：一种全新的科研支撑方式[J]. 图书馆杂志，2013（8）：4-11.

[3] 关联科学 [EB/OL]. [2016-03-25]. http：//Linked-Science.org.

[4] 刘炜，夏翠娟. 书目数据新格式 BIBFRAME 及其应用[J]. 大学图书馆学报，2014（1）：5-13.

[5] 编目精灵. “重温永恒的价值”以及关于 RDA 的观点对立[EB/OL]. [2016-03-25]. http：//catwizard.net/posts/20151031172826.html.

[6] ISBD Profile in RDA[EB/OL]. [2016-03-25]. http：//www.ifla.org/publications/international-standard-bibliographic-description.

[7] 唐义. 关联科学核心词汇规范：提出、优化及展望[J]. 图书馆杂志，2013（3）：55-60.

[8] 黄如花，邱春艳. 图书馆参与科学数据管理中的元数据应用实践研究[J]. 图书与情报，2014（6）：65-69.

[9] 赵华，王健. 科学数据元数据功能与内容分析[J]. 科技管理研究，2015（3）：232-235.

[10] STARR J，GASTL A. IsCitedBy：A metadata schema for DataCite [J]. California Digital Library，2011，17（1）：1-6.

[11] Ball Alexander. Overview of scientific metadata for data publishing “citation” and curation[C]. Eleventh International Conference on Dublin Core and Metadata Application (DC-2011). Bath：University of Bath，2011.

[12] 赵华，周国民，王健. 基于元数据的科学数据属性特征分析[J]. 情报杂志，2015（7）：173-178.

[13] 李蓓. 数字化图书馆资源仓库的基础[J]. 情报科学，2004（11）：1375-1379.

[14] Tan C S，Ting W S，Mohamad M S，et al. A review of feature extraction software for microarray gene expression data [J]. BioMed Research International，2014，8：1-15.

[15] 徐坤，蔚晓慧，毕强. 基于数据本体的科学数据语义化组织研究[J]. 图书情报工作，2015（9）：120-126.

[16] UMLS [EB/OL]. [2016-03-25]. https：//www.nlm.nih.gov/research/umls/.

[17] Metathesaurus[EB/OL]. [2016-03-25]. http：//www.ncbi.nlm.nih.gov/books/NBK9684/.

[18] 李丹亚，胡铁军，李军莲，等. 中文一体化医学语言系统的构建与应用[J]. 情报杂志，2011（2）：147-151.

[19] 祝清松，冷伏海. 引文类型识别研究进展[J]. 图书情报知识，2013（6）：70-76.

[20] 常颖聪，何琳. 科学实验数据元数据模型构建研究——以植物学基因表达实验为例[J]. 图书情报工作，2015（13）：117-125.

作者简介 常娥，东南大学图书馆副研究馆员；孙文佳，东南大学 2015 级图书情报专业硕士研究生。

收稿日期 2016-04-06

责任编辑：付伟棠