

# 分类主题一体化词表的国内研究 进展述评

姜冠兰 张敏

(西南大学计算机与信息科学学院, 重庆 400715)

**摘要:** 分类主题一体化词表是分类主题一体化思想的具体表现形式, 可以更有效地组织信息资源。本文通过内容分析, 对我国分类主题一体化词表的构成模式、实现方法、应用方式等进行系统梳理, 旨在对当前理论研究及实践应用提供参考。研究结果显示: 分类主题一体化词表在网络环境中的应用包括标引与检索、词表互操作、本体构建、关联数据发布、知识发现等多个方面; 一体化词表的关联化、分面化以及集成化、可视化在未来研究中还需深入拓展。

**关键词:** 分类主题一体化; 一体化词表; 《中国分类主题词表》; 关联数据

**中图分类号:** G254    **DOI:** 10.3772/j.issn.1673-2286.2020.08.003

**引用格式:** 姜冠兰, 张敏. 分类主题一体化词表的国内研究进展述评 [J]. 数字图书馆论坛, 2020 (8): 15-21.

网络信息资源日益丰富, 单独使用分类法或主题法已经不能满足信息资源组织与利用需求, 理想的模式是使用分类主题一体化语言进行集成化组织与揭示<sup>[1]</sup>。分类主题一体化词表(以下简称“一体化词表”)是分类主题一体化的具体体现, 国外情报机构自20世纪60年代就开始了一体化词表的研究。1969年英国情报学家艾奇逊·琼编制了《分面叙词表》, 被认为是世界上第一部分类主题一体化词表, 1980年侯汉清<sup>[2]</sup>对其进行了介绍和论证。1983年《常规武器工业分面叙词表》的成功编制, 是我国一体化词表的最早尝试。1994年出版的《中国分类主题词表》(以下简称《中分表》), 是我国第一部大型综合性的分类主题一体化词表。2009年中国科学技术信息研究所牵头组织了《汉语主题词表》(以下简称《汉表》)的重新编制, 现已完成《汉语主题词表(工程技术卷)》和《汉语主题词表(自然科学卷)》的出版, 在词表规模、等同率、词间关系方面有了较大提升, 推动了一体化词表的创新发展。

分类主题一体化词表有机融合了分类表和主题词表, 可以同时满足分类标引与检索、主题标引与检索等需要, 能充分发挥原有的分类法和主题法的长处, 弥补

各自的缺点。网络环境下, 分类主题一体化语言依然是情报组织与检索中不可或缺的工具, 因其词汇控制的规范性、概念的丰富性、逻辑的严密性, 一体化词表仍然在不同领域得到广泛应用。本研究对我国一体化词表的构成模式、实现方法、应用方式等进行了系统梳理, 希望对当前理论研究及实践应用提供参考。

## 1 分类主题一体化词表的构成模式

按照对分类类目与主题词的映射、兼容、集成等控制手段的不同, 常见一体化词表的构成模式主要有如下3种(见表1)。

(1) 分面叙词表。通常包括分面分类表和字顺叙词表两部分, 每个叙词均同时出现在分类表和叙词表中, 用分类号将两部分连接, 使类目和叙词对应。通过对分类表和叙词表进行统一的词形、词义和词间关系控制, 来实现二者的兼容, 是最典型的一体化词表。典型代表如《农业科学叙词表》(以下简称《农表》)。

(2) 分类法-叙词表对照索引(或称分类主题映射词表)。《中分表》就是此类词表的代表。通常包括

分类号-主题词对应表、主题词-分类号对应表两部分内容。每个分类号下列出对应的主题词或主题词串,通过分类类目实现主题词的聚类、分类和浏览;每个主题词下列出对应的分类号,利用主题词对类目作进一步地注释与说明<sup>[3]</sup>。

(3) 集成词表。将若干叙词表与分类表融合汇编而成的一种词表,通常以某一部分类法或叙词表为主,

列出与分类号或叙词相对应的其他分类法或叙词表中的分类号或叙词,以实现分类语言与主题语言的兼容与互换。如重新编制的《汉表》就是一个包含分类、主题和概念等不同语义级别词汇和概念数据库的集成知识组织系统<sup>[4]</sup>,以《中国图书资料分类法》为基础组织叙词表的分类显示体系,将叙词表和分类表有机结合。

表1 常见分类主题一体化模式及代表性词表

| 词表类型        | 代表性词表   |
|-------------|---|
| 分面叙词表       | 《分面叙词表》《建筑工业叙词表》《基础叙词表》《教育主题词表》《社会科学检索词表》<br>《音像资料叙词表》《农业科学叙词表》等      |
| 分类法-叙词表对照索引 | 《中国分类主题词表》《中国图书馆分类法(R类)与医学主题词表(MESH)中医学主题词表对应表》<br>《装备科技信息分类主题一体化词表》等 |
| 集成词表        | 中医药一体化语言系统、《美国一体化医学语言系统(UMLS)》《汉语主题词表(工程技术卷)》<br>《汉语主题词表(自然科学卷)》等     |

## 2 分类主题一体化词表的实现方法

通过对分类表和叙词表各自的术语、参照、标识和索引实施统一的控制,可以实现二者的有机融合。在数字时代到来之前,主要是人工编制完成;随着计算机技术的发展,分类法与主题法的自动转换成为可能,词表的编制开始辅以计算机技术实现。针对分类语言和主题语言的兼容互换,目前主要实现方法有直接映射、自动映射和集成法<sup>[5]</sup>。

(1) 直接映射。需要人为主观判定类目、词汇之间的关系,建立不同词表中词汇间或词汇与分类号之间的对应联系,准确率较高,但是过多依赖于专家。《中分表》就是采用直接映射的方法,实现《中国图书馆分类法》类目与《汉表》主题词的对应,主要依靠手工对应表标引完成。

(2) 自动映射。这是利用计算机对相同的元数据或编目记录中,来源于不同词表的语词和分类号的共现频次进行统计,计算语词与分类号的相似度值,从而建立语词与分类号的映射。自动映射需要足够多的标引数据,并且同一文献需要用不同分类法和叙词表进行标注,对训练数据的要求较高。自动映射方法主要是为实现词表扩充与更新的自动化,使其跟上学科或资源的快速发展变化。

(3) 集成法。实质上是将某一特定主题领域的

多部叙词表或分类法进行融合,在各来源词表的基础上建立包含所有术语及相关参照的集成词表,通过识别等价词及准等价词建立词汇转换系统,实现分类表与叙词表的兼容转换。如中医药一体化语言系统(TCMLS)采用集成法,将各种主题词表、分类表、工具书中相关词汇集成,建立了与UMLS功能相似的中医药学及相关学科的一体化语言系统平台。

直接映射准确率高,但是过于依赖专家,耗费大量人力;相比直接映射,自动映射的效率更高,但存在准确率低等局限;集成法能较好地保留原有知识体系的逻辑内容,但是工作量较大,在具体的实践活动中,往往会采用多种方法相互配合。如《汉表》在重新编制过程中,就结合语义计算、共现聚类等技术辅助领域专家确立词间关系,最终构建了一个集成知识组织体系<sup>[6]</sup>。

## 3 分类主题一体化词表的应用研究

现有研究中,除了综合领域之外,分类主题一体化的研究成果主要集中在医学和农业等专业领域。在医学领域,利用TCMLS实现医学文本语义关系的发现、构建中药概念数据模型等;在农业领域,基于《农表》开展了大量研究,如构建农业领域本体、实现农业科技关联数据的构建和发布、构建农业知识服务平台等。一体化词表的应用促进了网络环境下信息资源组织及

服务的展开。除了医学和农业专业领域之外,其他领域学者也开始尝试引入分类主题一体化模式,如国防军事<sup>[3,7]</sup>、电子政务<sup>[8-9]</sup>等领域。总体而言,一体化词表主要应用于标引与检索、词表互操作、本体构建、关联数据发布、知识发现等多个方面。

### 3.1 标引与检索

分类主题一体化是提高检索效率的保证,用户可以通过分类或主题的方式进行浏览或检索,并随意切换检索方式,还可以通过分类与主题的相互限定改变检索范围。如在数字图书馆中,将《中分表》与OPAC链接,能够实现网络联机检索目录的功能,可以为用户提供学科分类导航和概念检索服务<sup>[10]</sup>。《汉语主题词表(工程技术卷)》和《汉语主题词表(自然科学卷)》则通过《汉表》服务系统提供相关服务<sup>[11]</sup>,实现不同颗粒度的智能查询和检索功能,既可以从分类层级类目入手批量获取文献信息,也可以通过主题概念进行缩检与扩检,还可以从主题和学科角度对文档进行聚类分析。胡昌平等<sup>[12]</sup>、陈果等<sup>[13]</sup>还通过对主题词表的分面化改造实现科技文献检索效果的提升,以及网络社区分面导航系统原型的构建。

利用分类主题一体化也可以同时完成文献信息的主题标引和分类标引,一次标引能同时获得主题和分类标引的结果,提升标引的规范性和准确性,也能提高标引工作效率。在图书馆中,当电子化和网络化的一体化词表与编目系统对接之后,信息查找更为便利,缩检、扩检的难度降低,能实现分类主题一体化标引与编目系统的无缝链接,提高编目效率。

在网络环境下,手工标引不能完全满足用户需求,人们开始探索利用计算技术进行自动标引,选择自动标引的知识库也要同时兼顾主题标引和分类标引的需要,因此一体化词表也为自动标引提供了有利条件。卜书庆<sup>[14]</sup>构建了基于《中分表》知识组织系统的自动标引服务系统,可以支持数字资源的自动标引与自动分类,以《中分表》为基础,不仅可以简化标引程序,还能进行分类标引与主题标引的相互对照与检验。类似的,在《汉表》服务系统中,输入需要标引的文献标题和摘要之后,可以输出代表性高的优选词作为标引词,并同时赋予文献合适的分类号。

针对一体化词表的自动构建和改造,王军<sup>[15]</sup>以《中分表》为基础,提出从已标引的结构化语料库中提取专

业领域词汇自动丰富现有词表的思路和方法,基于该方法,通过发现反映文献主题的关键词,并将其映射到规范的主题词上,能够实现自动标引和编目。何琳等<sup>[16]</sup>将标引经验和机器学习相结合,对《中分表》进行了基于自动标引的改造,测试结果表明,改造后的《中分表》在标引深度和专指度方面都有所提高。

### 3.2 词表互操作

信息资源在结构、领域、语言等方面的差异,导致了不同知识组织系统的异构性,为解决由异构系统带来的资源共享不便的问题,检索语言之间的兼容互换研究成为知识组织研究重点讨论的问题,通过对词表进行互操作研究则可以满足用户对信息资源进行跨库、跨领域、跨语言检索的需求。分类主题一体化词表本质上就是由分类语言和主题语言通过映射兼容形成的知识组织系统,基于一体化词表本身的集成特点,进一步对一体化词表开展互操作研究能使构成的知识组织系统功能更加全面。

针对不同检索语言之间的互操作,刘华梅等<sup>[17]</sup>以《中分表》为核心体系构建了教育集成词库,利用同现映射、相似度计算等方法,实现了多部分类法、叙词表和《中分表》的兼容。在不同语种词表间的互操作方面,部分学者研究了《农表》与AGROVOC叙词表的映射,并提出了跨语言搜索引擎的设计方案<sup>[18-19]</sup>;鲜国建<sup>[20]</sup>建立了《农表》与几大涉农知识组织系统映射互联的描述框架,包括AGROVOC、NALT、LCSH和EUROVOC叙词表,但还未进行互操作具体实践;常春等<sup>[21]</sup>等按照国际通用的标准规范,建立了《汉语主题词表(工程技术卷)》到英文超级科技词表间的映射关系,促进了中英文跨语言、跨库检索的研究和实施。在此基础上,邓盼盼等<sup>[22-23]</sup>等从不同角度探讨了中英文叙词表概念映射关系及方法。

一体化词表能为构建本体等知识组织系统提供语义来源等术语服务<sup>[17]</sup>。随着语义网的发展,OWL、SKOS、Web Service等技术被引入术语服务研究,曾新红等<sup>[24]</sup>以《中分表》《社会科学检索词表》等为例,构建了中文叙词表本体共建共享系统(OTCSS),提供面向应用程序的Web Service术语服务,实现了术语的浏览、标引与检索。范炜等<sup>[25]</sup>以经过SKOS表征的《中分表》主题词表数据为基础,构建了面向用户和机器的术语服务原型系统,可支持关键词检索和可视化服务。此



外,《汉表》服务系统以网页服务和接口调用的方式,提供术语检索和浏览。

从现有研究来看,将一体化词表与其他词表进行互操作的项目并不多,且多是相同领域词表的互操作,对于跨领域知识融合等问题的研究还较为缺乏。针对一体化词表提供的术语服务,主要是提供术语的检索与浏览功能,信息检索之外的更多应用还需扩展。

### 3.3 本体构建

本体规范度高、语义丰富,可以提高异构系统之间的互操作性,能够较好地满足新网络环境下信息组织需求,有助于知识共享。本体与分类表和主题词表有相一致的特性,构建本体的方式之一就是传统的分类法、主题词表等知识组织体系中的相关概念改造成本体。一体化词表同时提供了知识的等级体系和主题概念的语义关联,为本体的构建提供了便利条件。利用一体化词表构建本体时,本体概念可以直接复用词表已有的选词,并且可以根据词表中的各类关系获得概念关系,直接参照词表中的参照关系、分类号以及限义词以确立概念属性<sup>[26]</sup>。

在相关研究中,由于《中分表》能在各学科领域中广泛使用,且能够统一描述各学科领域内的知识,因此较多研究选择《中分表》作为知识源。目前一体化词表用于本体构建大致包括完全复用型和语义辅助型两种方式<sup>[26]</sup>。完全复用型是指在构建本体时完全以词表中的主题词、分类类目词作为术语集,通过对一体化词表的完全复用,学者们构建了不同的领域本体,如民乐本体、图书情报本体、旅游本体、教育领域本体等。语义辅助型是指构建本体时词表不再用作术语集的全部来源,而是开始逐渐加入主题词以外的自然语言,如标签。张云中等<sup>[26]</sup>将《中分表》和大众分类法进行融合,复用《中分表》的语义关系、分类主题词作为构建本体的语义关系和术语集,再结合标签语义关系和高频标签辅助构建了一个散文领域本体,可以对网络资源进行深入的揭示,也可以确保本体构建所用术语集的新颖性。

### 3.4 关联数据发布

一体化词表具有丰富的概念及语义关系,能够有效地对信息资源进行组织和利用,提高信息获取效率。但是作为传统知识组织系统,一体化词表是相对封闭

和孤立的系统,目前还没有提供便捷获取和利用的途径,缺乏与外部网络应用进行开放链接的能力,妨碍了其在网络资源索引和检索方面的应用能力<sup>[25]</sup>。关联数据的出现为其指出了一个新的发展方向。关联数据主张在不同数据间建立联系,将孤立的资源关联起来,因此,可以将关联数据的理念和技术方法与一体化词表结合。

在关联化的一体化词表相关研究中,将一体化词表直接应用于信息资源关联数据化的现有研究还较为少见,在这类研究中,一体化词表主要作用是构建关联数据的基本语义关系模型,为其他关联数据提供语义结构和关系描述框架。如鲜国建<sup>[20]</sup>基于《农表》等知识组织体系,建立适用于描述多类型信息资源的多维语义关联框架模型,为多维农业科技语义关联数据的构建和发布提供基础。任瑞娟等<sup>[27]</sup>依托《中分表》叙词及词间关系建立关系型本体库,实现了学位论文、书目信息等多类型学术资源的语义化组织与关联化发布。

较多研究探讨了一体化词表本身的关联数据化实现。要实现一体化词表的关联化,首先要实现词表的规范化描述,SKOS是形式化描述的首选方法。SKOS简洁、通用、易扩展,提供的语义关系比RDF更加精确,不像OWL那样苛求较复杂的逻辑。《中分表》主题词表部分、《农表》均已进行了基于SKOS的语义转换实践。针对《中分表》一体化结构的描述也有学者提出了解决方案,如曾新红<sup>[28]</sup>提出将《中分表》当成两个独立的概念体系进行描述,然后进行映射集成。

在发布方式上,目前已经有许多工具支持关联数据的转换发布,如D2R Server、Virtuoso universal server、Triplify等,鲜国建等<sup>[29]</sup>和蔡颖<sup>[30]</sup>基于Virtuoso,分别实现了《农表》以及《中分表》关联数据发布系统的构建,Virtuoso可以将RDF储存到关系型数据库中,还支持SPARQL语法查询。关联数据集着眼于机器处理的便利,用户理解上还有一定难度,因此可视化十分必要。Relfinder、Graphviz、RDFGravity、Gruff等都是常见的关联数据可视化工具。如范炜等<sup>[25]</sup>结合Graphviz和Protovis类库实现了《中分表》主题词部分的关联数据可视化,能更直观地展示相关数据,增强用户对数据的理解。

基于SKOS的关联数据发布能够较好地表达概念之间的关系,但是对复杂等级体系的概念关系表达不够充分。曾新红<sup>[28]</sup>构建了中文知识组织系统形式化语义描述标准体系,包括通用CNKOS语义描述规范、高

受控词表的OntoThesaurus描述规范以及领域本体的OWL描述规范三部分。其中通用CNKOS语义描述规范对现有SKOS词汇无法细致描述的语义元素进行了扩展,如受控词表中组配概念及特种概念,分类法中交替类目、类目注释的具体类型及其隐含语义、类号范围等问题的处理;在此基础上发布了《中分表》的关联数据服务<sup>[31]</sup>,但目前该服务还未提供到其他词表的关联,可视化等功能也还在建设中。

### 3.5 知识发现

随着信息资源的爆炸式增长,用户的需求也从信息检索转向知识的有效获取。知识发现是从以各种形式表示的信息中,发现知识之间内在的联系,为用户提供更好的服务。从现有的概念关联体系中,直接获取概念术语的半监督知识发现技术具有良好的效果<sup>[32]</sup>。一体化词表将分类法中的层级关系与主题词表中的概念关系有机结合,能够为知识发现提供以概念为中心的同义词汇聚、概念及实体识别、基于范畴类目及等级体系的概念分类组织等信息,有助于优化知识发现服务<sup>[33]</sup>。

刘爱琴等<sup>[34]</sup>以《中分表》为受控词表,设计了面向非相关文献的知识关联发现系统,依据主题词的等级结构以及语义关联或相似程度,构造文献的隶属和相关结构,最后计算文献之间知识关联程度。赵瑞雪等<sup>[35]</sup>综合运用《农表》及其他词表建立索引,构建了基于元数据搜索的统一发现服务,能够实现一站式检索、多维分面、学科导航及语义拓展功能。

随着网络的发展,大量的网络社区也成为用户获取知识和解决问题的重要途径,用户在网络社区的交流内容通过一定的挖掘和组织后,能实现更深层次的知识服务。陈果<sup>[32]</sup>以心血管领域为对象进行实验,构建了基于《中文医学主题词表》和基于百科的结构化概念关联体系,在丁香园心血管论坛中引入领域概念关联体系,从概念间的细粒度关联角度对用户发帖内容建立相关关系,以实现知识关联发现。

## 4 研究结论及展望

随着一体化词表的应用不断拓展和延伸,从传统环境下最基本的标引与检索、词表的互操作研究,发展到语义网环境下的本体构建和关联数据发布,并为知识关联提供概念支撑;从文献信息组织扩展到网络信息

组织,并在不同的领域得到应用。其中,《中分表》《农表》的数字化建设领先于其他词表,为一体化词表的多方面应用奠定了较好基础。未来还可以从以下方面实现理论与应用扩展。

(1) 推动主题词表的分面化改造。国内对于分类主题一体化的研究主要集中在医学、农业领域,其他专业领域较少涉及。分面叙词表是兼容性最好的分类主题一体化模式,将分面分类体系引入不同领域的主题词表,可以弥补主题词表缺乏内在逻辑性和分类体系的缺陷。另外,概念属性的分面分析是本体的重要特征,因此在本体构建中分面叙词表也能得到良好的应用。有研究调查了我国目前已编制出版的140多部主题词表,这些主题词表涉及多个学科领域,但现有词表对新的信息和技术环境的不适应,使得其中70%都已处于休眠状态,没有得到利用与维护<sup>[36]</sup>。因此,加强主题词表的分面化改造,是实现各类信息资源分类主题一体化的切实可行方案,可以进一步扩展分类主题一体化的应用领域。

(2) 深化一体化词表的关联化研究。语义网络环境中,一体化词表要发挥语义组织与检索、知识发现等功能,实现词表的机器可理解、可操作、可执行是基本的前提和保障<sup>[37]</sup>。因此,如何将一体化词表加工成形式化的、四星甚至五星的开放关联数据是必须解决的关键问题。国内学者积极探索一体化词表的关联数据发布格式,如SKOS模型化。但许多研究聚焦于一体化词表中主题词表部分,对分类法部分重视不足,这也一定程度上导致了一体化词表集成特性发挥的作用有限。部分学者曾对此提出进一步的转换解决方案,如Zeng等<sup>[38]</sup>曾提出将OWL和SKOS结合,借助OWL的形式化逻辑解决类号组配与复分推理等问题,但是这些方案还未在实践中得到验证。此外,还需要对一体化词表与其他词表之间的关联匹配进行深入探索,关联数据五星标准要求就是不同RDF数据集之间构建关联,因此还需探索不同词表之间的语义映射方法,如引入带有机器学习的新AI或许可以提高不同词表间的语义互操作性<sup>[39]</sup>。

(3) 加强一体化词表的集成化、可视化研究,提升信息服务质量。在词表的集成研究中,由于不同词表的语义、层次结构,甚至是语言、文化的差异较大,导致词表的集成和映射任务较为困难。邓仲华等<sup>[40]</sup>曾提出多语种词表的映射和扩展方案,可以为一体化词表解决跨语言信息检索的问题提供参考。还有学者指出利用本体和语义框架开发的叙词表集成工具,可以解决叙

词表的层次结构模糊的问题,如VISTA工具旨在通过可视化词汇层次结构来帮助用户对两个术语之间的赋值进行智能处理<sup>[41]</sup>。

用户需求的满足和使用的便利是词表研究的最终目的,知识可视化和图谱表示越来越受到欢迎,如何将这些源于其他实践领域,但能够提高用户体验的知识表示形式加以整合,也是一体化词表需要考虑的。引入人工智能、机器学习、自然语言处理等领域的技术,探索构建知识图谱的方法,通过一体化词表本身的层级和概念关系,可以满足部分推理需求,并且可以基于知识图谱构建自动问答系统,为用户查找问题并提高检索质量提供帮助<sup>[42]</sup>。

探索一体化词表实现方法与技术看新,从概念形成、概念类目映射到规范描述,均需图书情报机构与专业机构、知识组织体系建设机构与应用机构等相关群体密切协作,提高信息资源共享程度,从而实现协同攻关。中国科学技术信息研究所提出的国家叙词库系统发展战略<sup>[43]</sup>,倡导加强全国性科研协作机制,对推动我国分类主题一体化词表的发展具有重要意义。

## 参考文献

- [1] 梅伯平. 网络信息组织的分类主题一体化研究[J]. 情报科学, 2003, 21(4): 385-387.
- [2] 侯汉清. 分类法与主题法结合的成功尝试——分面叙词表[J]. 图书情报工作, 1980, 24(6): 16-21.
- [3] 赵建国, 韩丽影. 分类表与主题表一体化集成研究——《军用分类主题映射表》集成构建[J]. 情报理论与实践, 2016, 39(5): 73, 86-89.
- [4] 曾建勋, 常春. 网络环境下新型《汉语主题词表》的功能定位与发展[J]. 情报学报, 2010, 29(6): 973-977.
- [5] ZENG M L, CHAN L M. Trends and issues in establishing interoperability among knowledge organization systems[J]. Journal of the American Society for Information Science and Technology, 2004, 55(5): 377-395.
- [6] 曾建勋, 常春, 吴雯娜, 等. 网络环境下新型《汉语主题词表》的构建[J]. 中国图书馆学报, 2011, 37(4): 43-49.
- [7] 韩丽影, 真漆. 推进装备科技信息资源组织创新发展——装备科技信息分类主题一体化词表编制初探[C]//第二十届全国计算机信息管理学术研讨会论文集, 第二十届全国计算机信息管理学术研讨会. 南昌: 中国科学技术情报学会, 2010: 55-60.
- [8] 高文飞, 赵新力. 我国政务信息资源主题分类研究[J]. 情报科学, 2008, 26(9): 1340-1343.
- [9] 贾君枝, 武晓宇. 基于FAST的综合电子政务主题词表分面式改造[J]. 图书情报工作, 2014, 58(8): 105-110.
- [10] 卜书庆. 基于《中分表》的国家数字图书馆知识组织思考[J]. 图书馆论坛, 2009, 29(6): 194-198.
- [11] 中国科学技术信息研究所. 《汉语主题词表》服务系统[EB/OL]. [2020-06-20]. <https://ct.istic.ac.cn/site/organize/index>.
- [12] 胡昌平, 林鑫. 科技文献检索中基于主题词表分面化改造的分面构建[J]. 情报学报, 2015, 34(8): 875-884.
- [13] 陈果, 肖璐, 孙建军. 面向网络社区的分面式导航体系构建——以丁香园心血管论坛为例[J]. 情报理论与实践, 2017, 40(10): 112-116.
- [14] 卜书庆. 中国分类主题词表网络化发展与关联数据建设研究[EB/OL]. [2020-04-23]. <http://www.chinalibs.net/ArticleInfo.aspx?id=406653>.
- [15] 王军. 词表的自动丰富——从元数据中提取关键词及其定位[J]. 中文信息学报, 2005(6): 38-45.
- [16] 何琳, 何娟, 阎素兰. 基于自动标引的《中分表》改造及测评研究[J]. 中国图书馆学报, 2012(6): 81-88.
- [17] 刘华梅, 侯汉清. 基于受控词表互操作的集成词库构建研究[J]. 中国图书馆学报, 2010, 36(3): 67-72.
- [18] LIANG A C, SINI M. Mapping AGROVOC and the Chinese agricultural thesaurus: definitions, tools, procedures[J]. New Review of Hypermedia and Multimedia, 2006, 12(1): 51-62.
- [19] 常春. 基于叙词表映射的农业跨语言检索系统设计[J]. 情报学报(增刊), 2008: 294-296.
- [20] 鲜国建. 农业科技多维语义关联数据构建研究[D]. 北京: 中国农业科学院, 2013.
- [21] 常春, 曾建勋, 吴雯娜, 等. 《汉语主题词表》与英文超级科技词表概念映射构架设计[J]. 数字图书馆论坛, 2012(12): 28-32.
- [22] 邓盼盼, 常春. 基于聚合关系的中英文词表概念映射方法及实证[J]. 数字图书馆论坛, 2014(10): 32-39.
- [23] 邓盼盼, 常春. 基于精确匹配的概念映射关系规则研究[J]. 图书情报工作, 2013, 57(16): 25-29.
- [24] 曾新红, 林伟明, 明仲. 中文叙词表本体的检索实现及其术语学服务研究[J]. 现代图书情报技术, 2008(2): 8-13.
- [25] 范炜, 邹庆. 《中国分类主题词表》的术语网络服务探索——以主题词规范数据为例[J]. 图书情报工作, 2012, 56(14): 40-46.
- [26] 张云中, 张从昱. 利用《中分表》与大众分类法构建本体: 基于融合的视角[J]. 情报科学, 2016, 34(10): 49-55.
- [27] 任瑞娟, 濮德敏, 王剑宏, 等. 基于Drupal实现多类型学术资源的语义化组织与关联化聚合[J]. 情报科学, 2015, 33(5): 63-67.



- [28] 曾新红. 中文知识组织系统形式化语义描述标准体系研究(一)——扩展SKOS实现传统受控词表全描述[J]. 中国图书馆学报, 2012, 38(3): 57-68.
- [29] 鲜国建, 赵瑞雪, 寇远涛, 等. 农业科学叙词表关联数据构建研究与实践[J]. 现代图书情报技术, 2013(11): 8-14.
- [30] 蔡颖. 国家图书馆关联数据注册与服务系统设计及实现[J]. 数字图书馆论坛, 2019(11): 36-43.
- [31] 黄华军, 曾新红, 林伟明. OTCSS关联数据服务的研究与实践[J]. 现代图书情报技术, 2012(Z1): 40-47.
- [32] 陈果. 基于领域概念关联的网络社区知识聚合研究[D]. 武汉: 武汉大学, 2015.
- [33] 李晓瑛, 李军莲, 李丹亚. 一体化医学语言系统及其在知识发现中的应用研究[J]. 数字图书馆论坛, 2019(9): 24-29.
- [34] 刘爱琴, 安婷. 面向非相关文献的知识关联发现系统设计与实现[J]. 国家图书馆学刊, 2019, 28(3): 75-83.
- [35] 赵瑞雪, 鲜国建, 寇远涛, 等. 大数据环境下的农业知识发现服务探索[J]. 数字图书馆论坛, 2016(9): 28-33.
- [36] 吴雯娜. 我国叙词表的编制历史与发展模式[J]. 情报理论与实践, 2018, 41(6): 39.
- [37] ZENG M L, MAYR P. Knowledge Organization Systems (KOS) in the Semantic Web: a multi-dimensional review [J]. International Journal on Digital Libraries, 2019, 20(3): 209-230.
- [38] ZENG M L, PANZER M, SALABA A. Expressing classification schemes with OWL 2 Web ontology language [EB/OL]. [2020-04-24]. <http://192.102.239.53/lis/stdnrds/grps/acoc/documents/EPC132-38.3.doc>.
- [39] ZENG M L. Interoperability [J]. KO Knowledge Organization, 2019, 46(2): 122-146.
- [40] 邓仲华, 罗玲, 王琴, 等. 多语种叙词本体研究[J]. 情报学报, 2009, 28(5): 664-671.
- [41] CHIO Y. Vocabulary Integration Reexamined [J]. International Journal of Librarianship, 2018, 3(2): 96-102.
- [42] QIAO B, FANG K, CHEN Y, et al. Building thesaurus-based knowledge graph based on schema layer [J]. Cluster Computing, 2017, 20(1): 81-91.
- [43] 周杰, 丁道劲, 吴雯娜, 等. 网络环境下国家叙词库的构建研究[J]. 图书情报工作, 2013, 57(16): 6-10.

## 作者简介

姜冠兰, 女, 1995年生, 硕士研究生, 研究方向: 信息组织服务与用户行为。

张敏, 女, 1974年生, 博士, 教授, 通信作者, 研究方向: 信息组织服务与用户行为, E-mail: zhangwu@swu.edu.cn。

Review on Research of Classification-Subject Integration Thesauri in China

JIANG GuanLan ZHANG Min

(School of Computer and Information Science, Southwest University, Chongqing 400715, China)

Abstract: The Classification-Subject Integration Thesauri is a form of classification-subject integration, which can organize information resources effectively. By using a content analysis method, this paper makes a systematic and comprehensive study of the construction mode, implementation method and application mode of the Classification-Subject Integration Thesauri in China, aims to provide reference for current theoretical research and practical application. The related research results show that the integration thesaurus is applied in the network environment in various ways, such as indexing and retrieval, interoperation, ontology construction, linked data released, knowledge discovery, etc. It is concluded that the future research can be carried out on the construction of the linked data, facet, integration and visualization of the integration thesaurus.

Keywords: Classification-Subject Integration; Integration Thesaurus; *Chinese Classified Thesaurus*; Linked Data

(收稿日期: 2020-06-05)