

# 面向语义出版的学术期刊数字资源聚合研究<sup>\*</sup>

■ 许鑫<sup>1</sup> 江燕青<sup>1</sup> 翟姗姗<sup>2</sup>

<sup>1</sup> 华东师范大学经济与管理学部信息管理系 上海 200241 <sup>2</sup> 华中师范大学信息管理学院 武汉 430079

**摘要:** [目的/意义]在语义出版发展的背景下,针对目前学术期刊数字资源聚合中存在的问题,提出实现面向语义出版的学术期刊数字资源聚合的方法,以期解决学术期刊数字资源提供与科研用户信息需求间的矛盾。[方法/过程]探讨面向语义出版的学术期刊数字资源聚合模型,构建以“冠心病”为具体应用背景的资源聚合实例,验证面向语义出版的数字资源聚合方法在学术期刊数字出版中的可行性。[结果/结论]以语义出版为导向进行学术期刊数字资源深度聚合,可以实现数字资源知识单元之间及与外部知识组织的语义关联,为科研用户提供更加高效的信息资源服务。

**关键词:** 语义出版 数字资源聚合 学术期刊

**分类号:** G250

**DOI:**10.13266/j.issn.0252-3116.2016.17.018

## 1 引言

随着信息化和数字化的不断发展,科研成果快速增长,科研效率逐步提高,数字出版物日益增加。一方面,数字化科研数据与学术文献的迅猛增长给科研用户的学术交流提供了诸多便利;另一方面,传统的 PDF 文档保存格式和知识表达方式已无法满足当下科研用户在新的科研环境下深度研读文献的需求<sup>[1]</sup>。为适应信息时代的要求,学术期刊加快了自身的数字化进程。然而在学术期刊数字化蓬勃发展的同时,其存在的问题也日益凸显。

由于不同来源的学术期刊所采用数据的规范和标准不同,使得其数字资源难以统一,给检索和利用带来困难。同时,不同出版商或服务提供商各自为政,学术期刊数字资源之间缺少必要的关联,传统的数字化方式难以提供高度集成的知识,科研人员在资源查找上需要花费大量时间。综上所述,当前学术期刊数字资源聚合程度较低,在内容揭示上也不甚理想,阻碍了学术期刊数字出版的纵深发展,传统的数字出版方式已无法适应用户对知识进行检索与利用的需求。

语义出版<sup>[2]</sup>是建立在资源深度揭示与高度集成基础上的信息组织方式,受到了越来越多的关注。作为数字出版的高级形态,语义出版丰富了学术期刊信息服务

的资源类型,为创新学术期刊组织与发布方式提供了新的思路;同时,通过对知识单元进行语义标识提高资源对象间的关联度,有助于实现资源内容的按需重组与发布。因此,本文以语义出版为导向,探讨面向语义出版的学术期刊资源聚合模型与应用,以期有效解决当前学术期刊在资源组织与服务中存在的集成维度关联问题,进而提高学术期刊在信息服务产业链中的地位,促进出版行业、图书馆以及新兴服务商等机构的协调发展。

## 2 相关研究

近年来,随着语义网技术的成熟与发展,基于本体和关联数据的聚合方式成为目前数字资源聚合方法的研究热点。数字资源在语义层面上的聚合重点关注的是其内容特征,并着力构建一个内容相互关联、多维度、多层次的资源体系。针对于学术期刊资源,目前主要有 4 种数字化语义聚合方法。

(1) 基于 XML 的数字资源聚合。基于 XML 的数字资源聚合方法充分利用了 XML 在数据处理方面的特点,将其作为数据交换与集成的中介,以提供多种数据之间的转换,实现数据集成与交互。王军<sup>[3]</sup>通过阐述 XML 在数字资信息整合方面的优势,提出了基于 XML 的数字图书馆 Web 信息资源整合系统,从而更好

\* 本文系国家社会科学基金青年项目“面向语义出版的数字图书馆资源多维度聚合研究”(项目编号:15CTQ007)研究成果之一。

作者简介:许鑫(ORCID:0000-0001-7020-3135),教授,E-mail:xxu@infor.ecnu.edu.cn;江燕青(ORCID:0000-0002-8896-0094),硕士研究生;翟姗姗(ORCID:0000-0002-2787-0183),副教授。

收稿日期:2016-07-04 修回日期:2016-08-15 本文起止页码:122-129 本文责任编辑:王传清

地整合数字图书馆的信息资源。唐振宇<sup>[4]</sup>等人提出了基于XML的图书馆网络信息资源整合模型,并采用了JSP+XML的跨库检索系统,从而实现不同来源的信息资源的整合。

(2)基于元数据的数字资源聚合。基于元数据的数字资源聚合使用元数据技术来统一管理分散的数字资源,目前这一方法得到了普遍重视与应用。邱均平、方国平<sup>[5]</sup>提出基于元数据的馆藏资源深度聚合模式框架,在各层次之间实现数据系统的“自组织”。黄文碧<sup>[6]</sup>将元数据关联作为实现馆藏资源聚合的一种有效途径,从元数据仓库构建、元数据映射和元数据关联3个方面实现馆藏资源聚合。刘峰、张晓林<sup>[7]</sup>对当前科研领域中6种典型的元数据标准进行了深入分析,构建了科学数据元数据标准通用数据项设计模型。成全等<sup>[8]</sup>通过对《书目记录的功能需求》FRBR概念模型及《资源描述与检索》RDA编目规则核心思想及体系结构的深入分析,以FRBR、RDA与关联数据为基础,构建了馆藏资源元数据的语义描述与关联网络构建的框架模型,为馆藏资源元数据的语义描述及关联网络构建的实践研究提供了可能。

(3)基于本体的数字资源聚合。基于本体的数字资源聚合方法在国内的应用主要是构建各种类型的本体知识库和本体概念模型,在语义层面对资源内容进行描述和揭示,以此实现资源的深度聚合。何超等<sup>[9]</sup>建立基于本体的馆藏数字资源语义聚合与可视化模型,来解决数字资源孤岛和数字资源超载问题。肖希明等<sup>[10]</sup>研究基于本体的公共数字文化资源整合,构建公共数字文化资源顶层本体,将资源组织形式提高到了语义层面上,实现其语义互操作,从而为资源聚合提供便利。王文清等<sup>[11]</sup>基于OWL-DL标准的出版物数字内容资源本体模型PUBO,尝试设计与开发下一代数字资源内容管理系统。

(4)基于关联数据的数字资源聚合。关联数据在数字资源聚合方面具有独特优势,国内外学者展开了基于关联数据整合数字资源的一系列研究。如瑞典国家图书馆首先将书目(LIBRIS)开放为关联数据,美国国会图书馆将图书馆标题表(LCSH)发布为关联数据。我国对关联数据应用的研究目前尚处于起步阶段,田宁<sup>[12]</sup>详细阐述了基于关联数据的信息资源整合模型及如何实现,通过馆藏资源数字化、创建关联数据形式、发布关联数据等几个步骤,实现信息资源的整合;欧石燕、胡珊等<sup>[13]</sup>将本体与关联数据结合,提出了本体和关联数据驱动的图书馆信息资源语义整合框架,

以实现不同资源间语义关联以及关联数据的发布、检索与访问。赵夷平、毕强<sup>[14]</sup>利用关联数据的机器可读、语义表示、关联描述和网络资源属性的优势,初步实现具有相似文献查询功能的学术资源关联数据,便于用户快捷地获得与任何一篇文献高度相关的其他文献,从而实现学术资源深入聚合。夏翠娟等<sup>[15]</sup>以家谱数据作为起点,尝试利用关联开放数据技术重组图书馆传统资源,构建历史文献数据服务平台,该平台可支持面向万维网的书目控制,提供针对普通用户的寻根搜索服务和针对专业人士的数据挖掘服务,有效地实现了家谱信息在语义层面上的资源聚合。

综上所述,针对数字资源在语义层面上的聚合方法研究,国内外已取得了丰硕的成果,但数字资源聚合仍缺乏明显的导向性,且未能深入到知识单元层面揭示资源间的关联关系,从而无法适应用户对知识发现和资源深度揭示的需求。作为数字出版的高级形态,语义出版的出现为创新学术期刊组织与发布方式提供了新的思路,同时,通过对知识单元进行语义标识提高资源对象间的关联度,将有助于实现资源内容的按需重组与发布。目前国外已逐步开展语义出版与数字资源聚合相结合的研究,但这一领域还处于不断发展与创新的阶段。因此,本文以语义出版为导向,通过探讨面向语义出版的学术期刊资源聚合的模型与应用,以期有效解决当前学术期刊在资源组织与服务中存在的集成关联问题。

### 3 语义出版对数字资源聚合的要求

王晓光、陈孝禹<sup>[16]</sup>提出语义出版的DBIU层次模型,将其架构层次分为数据层(data)、业务层(business)、交互层(interaction)和用户层(user),其中业务层是交互层和用户层功能实现的前提,依据业务层的要求来对数字资源进行聚合,才能更好地满足语义出版的要求,为交互层和用户层提供一个组织有序、整合良好的资源整体。因此业务层对数字资源聚合有着更基础和更迫切的要求,换句话说,语义出版层次模型对数字资源聚合的要求主要体现在业务层上,业务层需要借助语义的数字资源聚合方法来实现期刊内容的结构化描述和语义标注,从论文单元间的关联深入到论文所包含知识单元之间的关联,建立起语义元素之间的相互链接,形成具有语义关系的知识网络。

语义出版对数字资源聚合的要求可归纳为以下几方面。

(1)实现对期刊内容的结构化描述和对内部知

识单元的语义标注。不同来源的学术期刊往往采用不同的元数据标准,为资源聚合带来了语义异构问题,且难以实现资源间的语义互操作,信息孤岛问题严重。语义出版则要求通过语义技术为不同类型不同来源的资源提供统一的概念描述标准,从而实现资源的结构化描述,实现知识单元的语义标注,从而实现资源之间的语义互操作,解决期刊内容的语义异构问题。

(2)实现资源之间的相互关联,丰富出版物的内容。在传统互联网上,数字对象之间的关联主要靠超链接实现,这些超链接多是静态的,难于更新。语义出版可以通过借助语义技术来实现数字对象之间丰富的、多层次的、自动化的相互关联。这种关联不仅包括期刊与期刊之间的关联,还包括内部知识单元之间的关联、知识单元同外部解释性资源之间的关联等。通过创建链接,在不同来源、不同类型的数据之间建立语义关联,使其成为无缝链接、无限开放的整体,从而丰富出版物的知识内容。

(3)语义出版要求构建一个知识网络。在资源内容结构化描述和资源之间语义关联的基础之上,通过关联数据等技术构建一个庞大的、广泛的知识网络,使得出版物能够摆脱孤立的状态,实现资源之间以及同外部资源丰富的语义关联,促进信息和知识的自动发现和获取,为用户检索和利用资源提供便利。

## 4 学术期刊数字资源语义聚合模型的整体设计

基于语义出版的业务层对数字资源聚合提出的要求,本文构建了面向语义出版的学术期刊数字资源聚合模型,如图 1 所示:

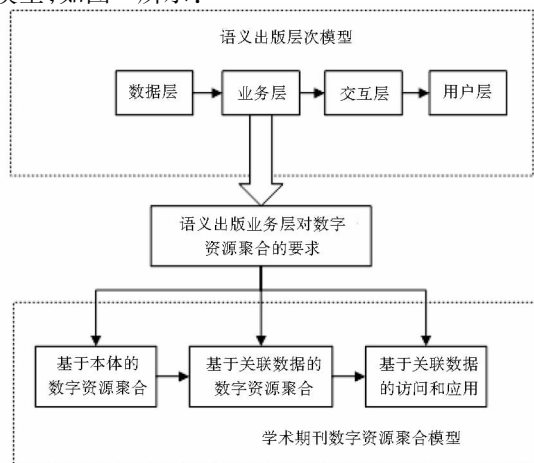


图 1 面向语义出版的学术期刊数字资源聚合模型

由图 1 可见,通过本体和关联数据等方法,能够实现学术期刊数字资源的结构化描述,并对其进行语义标注,揭示资源之间蕴含的内在联系,由此建立各个层面上的资源间丰富的、多层次的、自动化的语义关联。在此基础上,通过发布关联数据,构建一个无缝衔接的、广泛的知识网络,使得用户能够便捷地获取需要的知识,促进知识的自动发现,方便用户对信息知识的访问和检索利用。该模型包括 3 个层次架构:第一层,基于本体的数字资源聚合,结合学术期刊的著录特点,构建基于 DC 的核心元数据本体和基于 FOAF 的相关本体,并对主题词用 SKOS 的方法进行描述,以实现不同源数据的语义化描述和语义互操作;第二层,实现基于关联数据的资源聚合,在所构建的本体模型的基础上,将这些数据转换成统一格式,通过 RDF 连接各种相关资源,实现学术期刊数字资源中不同知识单元间的聚合以及同外部知识组织的链接,并采用静态文件的方式发布关联数据;第三层,基于关联数据的访问和应用,直接面向学术期刊信息用户,用于实现其对学术期刊数字资源的浏览以及查询,通过关联数据,用户可以得到简洁而清晰的查询结果,并在此基础上将相关的数字资源链接到一起,从而实现资源之间的无缝衔接。

### 4.1 基于本体的数字资源聚合

本体的数字资源聚合主要从以下 4 方面来阐述:明确资源的语义内容和含义、作为一种资源组织框架、作为查询模型以及实现语义逻辑推理。在资源组织中,通过对概念添加属性,在属性之间添加映射关系等使本体可以模拟人的思维结构和过程,对用户提交的查询进行语义推理,从中发现隐含其中的用户需要却未能表达出的需求。

不同来源、不同类型的学术期刊数据库资源通常会采用不同的元数据规范进行描述,这就导致了学术期刊数字资源的结构异构,给聚合和检索带来了困难。元数据虽然提供了数字资源的语义化基础,但是无法解决资源描述的异构性和语义性问题,而本体则可以实现不同格式、不同类型的元数据之间的互操作。采用本体来描述元数据,将其转换成统一 RDF 格式,可以实现不同源数据的语义化描述和语义互操作。同时,由于传统资源聚合中关键字匹配的方法无法解决一词多义、一义多词的问题,领域本体库的建立、同义词的描述以及语义元素间等级关系的明确,可以有效解决此类问题,为用户的信息获取带来便利。

### 4.2 基于关联数据的数字资源聚合

与传统网络相比较,以关联数据为核心的数据网



能够提供给用户更准确、更丰富的信息资源,更好地满足用户获取资源的需求。关联数据在学术期刊数字资源聚合中的作用主要体现为链接不同来源、不同格式的数据,实现资源聚合与关联;有效揭示资源之间的联系;在资源之间建立关联,丰富资源的知识内容。通过关联数据,为学术期刊连接更多的外部的开放资源提供技术框架,从而大大丰富期刊的知识内容。这样使得学术期刊出版不仅能够体现期刊本身的内容,还可以实现在语义层面上有关联的数据资源的联接。这也满足了语义出版中关联的需求。

依据关联数据的原则,在面向语义出版的学术期刊数字资源聚合模型的第二层中,采用 HTTP 协议中的 URI 地址命名所有资源,将数字资源在元数据本体化的模型基础上进行语义化描述。并采用 SKOS 语言进行语义化描述,将其他资源转化为统一的资源描述形式——RDF 格式,从而实现学术期刊数字资源语义聚合。

4.3 基于关联数据的访问和应用

经过数据资源聚合与关联数据发布之后,用户可以通过关联数据浏览器对数据集进行浏览,通过各数据之间的 RDF 链接,指引用户在不同的数据源之间进行浏览。用户还可以通过语义搜索引擎,在关联数据网络之中检索到需要的信息资源。

具体来说,已发布的关联数据主要通过关联数据浏览器和关联数据搜索引擎等工具使得用户可以在数据网络中进行浏览和检索。通过关联数据浏览器,可以为用户提供资源之间的导航,借助 RDF 链接,用户可以在不同的数据源之间自由切换,实现对整个数据网络的浏览,提高浏览效率。用户对关联数据的查询,则可以借助关联数据搜索引擎来实现,在网络中抓取关联数据和相关信息。此外,在关联数据访问中,还可以将其浏览与检索技术结合在一起,为用户提供更加智能的数据服务。

5 面向语义出版的学术期刊数字资源聚合实现

语义出版的主要目标是通过数据抽取和信息聚合的方式促进期刊论文的知识发现,为科研用户提供一种更高效的信息利用方式。对于用户而言,信息检索与文献浏览是利用学术期刊资源进行科研活动最主要的方式。资源聚合的效果体现在信息检索上,并影响着最终的检索结果,而信息检索的最终结果直接关系到用户是否继续下一步的科研活动。从这个角度来

讲,实现基于语义的信息检索可以被看作是语义出版的实现目标之一。而目前数据库集成商提供的检索系统大多还是基于关键词进行信息查询,即将用户输入的检索词同数字资源的标题、全文、关键词等进行匹配查找。这样的机械匹配会由于忽略概念间的语义关系而漏检相关信息,无法精确满足用户查询需求。基于语义的资源聚合基础上的语义查询则可以避免这种问题的发生,提高用户获取信息的效率。因此,本文将从用户的体验角度着手,通过构造语义检索的实例来检测以语义出版为导向的数字资源聚合是否有效,如图 2 所示:



图2 学术期刊数字资源聚合方案

本文在构建的理论模型的基础上,将医学领域的常见心血管疾病——“冠心病”作为实例,以“冠心病”为主的系列主题词作为进一步查找期刊文献的检索词,将学术期刊数字资源与其他网络数据资源(作者、机构组织等相关信息)进行语义化描述转换成统一的数据,实现基于本体和关联数据的数字资源聚合,完成语义出版业务层对学术期刊数字资源聚合的要求。本文所设计的学术期刊数字资源聚合方案涉及4个步骤:①本体的构建。复用DC元数据构建核心元数据本体,描述涉及到的相关题录信息,并通过网络途径(学术机构库、专家信息库、机构主页、个人主页、维基百科等)查找相关作者和机构的信息,形成FOAF数据集;利用SKOS知识组织系统描述“冠心病”系列主题词表和分类号。②关联数据的构建。将已构建的本体数据转化为语义元数据,并用RDF格式来表示,用URI表示的资源替代RDF格式或者XML格式表示的语义元数据中数据类型属性的属性值,从而形成不同数据集数据之间的关联。③关联数据的发布与访问。采用静态文件形式来发布关联数据,以实现关联数据的发布和访问。④聚合结果评价与分析。将中国期刊全文数据库和万方数据库作为聚合结果评价的对比对象,

对“冠心病”实例进行聚合结果的分析与评价。

5.1 本体构建

本体是一种用于描述概念与概念之间关系的模型。随着语义网技术的发展,本体已具有很强的表达概念语义和获取知识的能力,通过明确表示概念的含义和相互的内在联系,从中推理出概念与概念之间隐含的关系,可以用来解决目前学术期刊数字资源出现的语义聚合问题。

(1)基于 DC 的核心元数据本体设计和描述。本文复用 DC 元数据标准的 15 个核心元素,构建核心元数据本体的属性,并在此基础上进一步添加了新的属性以及子属性(见表 1)。在表 1 中,对学术期刊的著录款项与 DC 核心元数据本体进行了相应的对照。本文构建的核心元数据本体的命名空间为 `http://example.ecnu.edu.cn/Ontology/metaonto.owl#`,缩写为“`metaonto:`”。

表 1 期刊著录款项与 DC 核心元数据元素对照表

著录内容	核心元素		
	元素名称	修饰词	复用 DC 元素
文章题名	题名	—	Title
文章责任者	主要责任者	机构	Creator
		职称	
		学历	
文章关键词	主题	主题词表	Subject
文章分类号		分类号	
摘要	描述	摘要	Description
基金项		资助	
其他责任者	其他责任者	责任方式	Contributor
		机构	
收稿日期	日期	接受日期	Date
修改日期		修改日期	
出版日期		发布日期	
期刊属性	资源类型	—	Type
纸张大小	格式	—	Format
唯一标识符	标识符	—	Identifier
论文语种	语种	—	Language
参考文献	相关资源	—	Relations
文章的版权信息	权限	—	Right

(2)基于 FOAF 的相关本体设计和描述。对于人物和组织机构,主要基于 FOAF 本体<sup>[17]</sup>进行描述,并对该本体进行必要的扩展(扩展部分为 `foafx`),如表 2 所示。本文通过万方数据库提供的专家信息库,来查找本文样本文献中提及的人物和组织机构;如万方数据库中未收录,则从相关的个人主页和相关机构中查找,并描述其语义化信息。

表 2 FOAF 本体及相关扩展属性

姓名	foaf:name
性别	foaf:gender
生日	foaf:birthday
邮箱	foaf:mbox
电话	foaf:phone
职称	foaf:title
工作主页	foaf:workplaceHomepage
个人主页	foaf:homepage
兴趣爱好	foaf:topic_interest
当前研究项目	foaf:currentProject
出生地	metaonto:birthPlace
荣誉	metaonto:awards
出版物	metaonto:publications
研究方向	metaonto:researchDirection
社交账号	metaonto:Chat_ID

5.2 基于 SKOS 的知识组织资源描述

对于本文中涉及到的相关知识组织资源(主题词、分类号),将采用 SKOS 语言进行描述,以明确描述概念的含义及其相关关系。SKOS 实质上是一套预定义的词汇集,用于描述各类结构化受控词表的结构和概念,例如叙词表、分类法、术语表等,以机器可理解的方式来表达其结构、概念以及语义关系,从而提供交换和重用<sup>[18]</sup>,当用户输入检索词后,系统将在检索词的同一等级中进行自动搜索,若用户认为信息量过少时,则系统可以根据用户的反馈信息提供该词的上位类相关文章,使用户可以在该词的上位类词中继续搜索,扩大检索范围,提高信息查全率;若用户认为信息量过多时,系统则可以提供该词的下位类,以提高查准率<sup>[19]</sup>。SKOS 知识组织资源描述能够帮助用户在查询时实现语义上的蕴涵扩展(如查找“心血管疾病”时,也能查询到“心脏病”“冠心病”的信息)、语义同义扩展(如查询“冠心病”时,也能查到“冠状动脉痉挛”“coronary heart disease”等同义信息)、语义相关扩展(如查询“心绞痛”时,也能查询“冠状动脉狭窄”“冠状动脉闭塞”等),满足用户在信息查全和查准上的不同需求。

用 SKOS 描述题名为《冠状动脉粥样硬化性心脏病》的文献样例见图 3。

5.3 关联数据的构建

关联数据的构建是依据本体将数据转化为 RDF 格式的语义元数据,并在数据集间形成关联。在上述构建本体数据的基础上,利用 Altova XML Spy 软件将其全部转换成 XML/RDF 格式,采用数据类型属性,即属性值均为文字值,并基于 FOAF 本体描述个





表 3 CNKI 中各检索表达式与检索结果的对应关系

检索表达式(A)	检索结果(A)(篇)	检索表达式(B)	检索结果(B)(篇)
TI = ‘冠心病’	62 089	TI = ‘冠心病’ NOT TI = ‘冠状动脉粥样硬化性心脏病’	62 089
TI = ‘冠状动脉粥样硬化性心脏病’	1 178	TI = ‘冠心病’ NOT TI = ‘冠状动脉心脏病’	62 089
TI = ‘冠状动脉心脏病’	22	TI = ‘冠心病’ NOT TI = ‘冠状动脉疾病’	62 088
TI = ‘冠状动脉疾病’	662	TI = ‘冠心病’ NOT TI = ‘冠状动脉痉挛’	62 086
TI = ‘冠状动脉痉挛’	417	TI = ‘冠心病’ NOT TI = ‘缺血性心脏病’	62 086
TI = ‘缺血性心脏病’	1 437	TI = ‘冠心病’ NOT TI = ‘冠状动脉供血不全’	62 089
TI = ‘冠状动脉供血不全’	4	TI = ‘冠状动脉疾病’ NOT TI = ‘冠心病’	661

表 4 万方数据库中各检索表达式与检索结果的对应关系

检索表达式(C)	检索结果(C)(篇)	检索表达式(D)	检索结果(D)(篇)
题名 = ‘冠心病’	53 812	题名:(“冠心病”) ^ 题名:(“冠状动脉粥样硬化性心脏病”)	53 811
题名 = ‘冠状动脉粥样硬化性心脏病’	1 089	题名:(“冠心病”) ^ 题名:(“冠状动脉心脏病”)	53 812
题名 = ‘冠状动脉心脏病’	17	题名:(“冠心病”) ^ 题名:(“冠状动脉疾病”)	53 811
题名 = ‘冠状动脉疾病’	371	题名:(“冠心病”) ^ 题名:(“冠状动脉痉挛”)	53 812
题名 = ‘冠状动脉痉挛’	206	题名:(“冠心病”) ^ 题名:(“缺血性心脏病”)	53 812
题名 = ‘缺血性心脏病’	878	题名:(“冠心病”) ^ 题名:(“冠状动脉供血不全”)	53 812
题名 = ‘冠状动脉供血不全’	0	题名:(“冠状动脉疾病”) ^ 题名:(“冠心病”)	370

通过以上对比可以发现,现有的资源聚合方式多是针对文献资源或是文献资源中的几种进行整合,这样的资源聚合方式没能使用户链接和发现更多的外部相关信息,也没能揭示出资源间的隐含关系,资源聚合的广度和深度不够。

面向语义出版的学术期刊数字资源聚合有效实现了期刊内容语义上的融合,因此可以为用户提供基于语义的检索查询。本体提供了对概念和概念之间关系的明确定义,因此可以揭示出资源间隐含的内在关系(同一、整体与部分等),从而可以在检索过程中发挥作用,提高检索效率。如当用户输入检索词“冠心病”之后,检索系统可以依据本体对概念的描述来查找与这个词具有相同含义但表述不同的其他词汇(冠状动脉粥样硬化性心脏病、冠状动脉痉挛等),并将其作为检索词进行进一步的查询,从而可以大大降低以往单一依靠关键词检索的漏检率,提高文献检索中查全率和查准率。

6 结语

本文为解决学术期刊数字资源提供与科研用户信息需求的矛盾,提出了一种面向语义出版的学术期刊数字资源聚合模型,该模型通过本体和关联数据等方法,能够实现对学术期刊数字资源的结构化描述,并对其进行语义标注,揭示资源之间蕴含的内在联系,由此建立各个层面上的资源间丰富的、多层次的、自动化的

语义关联。在此基础上,通过发布关联数据,构建无缝衔接的知识网络,使得用户能够便捷地获取需要的知识,促进知识的自动发现,方便了用户对信息知识的访问和检索利用。此外,本文构建了以“冠心病”为例的资源聚合实例,实例展示了基于本体、SKOS 和关联数据等方法实现学术期刊数字资源聚合的优势,探讨了在当前语义出版环境下利用语义网等相关技术进行学术期刊数字资源聚合的有效性和可行性。在此基础上,真正实现了学术期刊出版领域中,不同类型资源之间的深度聚合,以及发现并无缝链接外部广阔的相关资源,可以为用户提供更加智能的服务。

笔者认为,面向语义出版的学术期刊数字资源聚合的相关实践还处于发展阶段。实现资源的深度聚合是学术期刊语义出版发展的必然趋势,在未来随着本体、关联数据等相关语义技术的不断成熟和发展,以及科研用户对语义出版强烈的需求推动,更多隐性的知识特征将被挖掘和发现,语义层面上也将集成更丰富的信息源,用户获取信息也会获得更大的便利。

参考文献:

[ 1 ] 王晓光,陈孝禹. 语义出版的概念与形式[J]. 出版发行研究, 2011(11):54-58.

[ 2 ] SHOTTON D, PORTWIN K, KLYNE G, et al. Adventures in semantic publishing: exemplar semantic enhancements of a research article[J]. PLOS computational biology, 2009, 5(4):e1000361.

[ 3 ] 王军. 基于 XML 本体描述语言的数字图书馆 Web 信息资源整合[J]. 现代情报,2008,27(11):84-86.

[ 4 ] 唐振宇,陈凤岩,冯玉强. 基于 XML 的图书馆网络信息资源整合研究[J]. 哈尔滨工业大学学报,2007,39(7):1135-1137.

[ 5 ] 邱均平,万国平. 高校图书馆语义化馆藏资源深度聚合模式及其应用研究[J]. 图书馆学研究,2014(21):64-71.

[ 6 ] 黄文碧. 基于元数据关联的馆藏资源聚合研究[J]. 情报理论与实践,2015,38(4):74-79.

[ 7 ] 刘峰,张晓林. 科学数据元数据标准述评及其通用化设计研究[J]. 现代图书情报技术,2015(12):3-12.

[ 8 ] 成全,许爽,钟晶晶. 馆藏资源元数据语义描述及关联网络构建模型研究[J]. 情报理论与实践,2015,38(4):124-129.

[ 9 ] 何超,张玉峰. 基于本体的馆藏数字资源语义聚合与可视化研究[J]. 情报理论与实践,2013,36(10):73-76.

[ 10 ] 肖希明,完颜邓邓. 基于本体的公共数字文化资源整合语义互操作研究[J]. 国家图书馆学刊,2015(3):43-49.

[ 11 ] 王文清,刘春彤,张月祥,等. PUBO:面向出版的数字资源本体建模[J]. 大学图书馆学报,2015,33(3):88-95.

[ 12 ] 田宁. 基于关联数据的信息资源整合[J]. 图书馆学刊,2014,36(1):37-39.

[ 13 ] 欧石燕,胡珊,张帅. 本体与关联数据驱动的图书馆信息资源语义整合方法及其测评[J]. 图书情报工作,2014,58(2):5-13.

[ 14 ] 赵夷平,毕强. 关联数据在学术资源网相似文献发现中的应用研究[J]. 现代图书情报技术,2016(3):41-49.

[ 15 ] 夏翠娟,刘炜,陈涛,等. 家谱关联数据服务平台的开发实践[J]. 中国图书馆学报,2016(3):27-38.

[ 16 ] 王晓光,陈孝禹. 语义出版:数字时代科学交流系统新模型[J]. 出版科学,2012(4):81-86.

[ 17 ] FOAF vocabulary specification [EB/OL]. [2016-03-22]. <http://xmlns.com/foaf/spec/>.

[ 18 ] W3C. SKOS Simple Knowledge Organization System [EB/OL]. [2016-03-22]. <https://www.w3.org/2004/02/skos/>.

[ 19 ] 李欣. 基于概念检索的智能信息检索技术研究[D]. 武汉:华中师范大学,2004.

作者贡献说明:

许鑫:提出研究思路与方法,设计研究方案,审核修订论文;  
江燕青:采集与处理数据,撰写论文初稿;  
翟姗姗:整理资料,撰写与修订论文。

Research on Digital Resources Aggregation in Academic Journals Oriented to Semantic Publishing

Xu Xin<sup>1</sup> Jiang Yanqing<sup>1</sup> Zhai Shanshan<sup>2</sup>

<sup>1</sup> Department of Information Management, Faculty of Economics and Management,  
East China Normal University, Shanghai 200241

<sup>2</sup> School of Information Management, Central China Normal University, Wuhan 430079

**Abstract:** [Purpose/significance] With the development of the semantic publishing, this paper presents the semantic publishing of the academic journals digital resources polymerization method in order to solve the contradictory of digital resources provided of academic journals and research information needs of the user. [Method/process] This paper proposes digital resource aggregation model of semantic publishing, and constructs the resource aggregation instance of “coronary heart disease”, and also verifies the feasibility of digital resource aggregation in the application of the digital publishing of academic journals. [Result/conclusion] The depth of polymerization of academic journals digital resources can achieve the semantic linkage between digital resources knowledge units and the external knowledge organization, and also can provide more efficient information resources services to scientific users.

**Keywords:** semantic publishing digital resource aggregation academic journals

《图书情报工作》2016 年增刊(1) 征订启事

《图书情报工作》2016 年增刊(1)已于 2016 年 6 月底出版,内容涉及理论研究、信息服务、资源建设、计量分析、情报研究、信息技术、队伍建设等诸多方面,有一定的参考和收藏价值。欢迎各图书馆、情报所和广大图书情报工作者订阅。定价:40 元。

地 址:北京中关村北四环西路 33 号 5D05 室 邮编:100190  
联系人:赵 芳 电 话:010-82623933 电子邮件:tsqbgz@vip.163.com