

*本文系中国科学院文献情报能力建设专项子项目“开放知识资源中心体系建设”（项目编号：Y6ZG291001）研究成果。

元数据语义化映射过程研究 ——以中科院机构名称规范控制库为例*

李慧佳 马建玲 张秀秀 王楠

摘 要 元数据语义化映射是实现数据库或系统语义互操作的基础。文章结合元数据语义描述标准和元数据语义化映射方法，对元数据的语义映射过程进行分阶段描述，主要包括元数据集识别、数据元素分组和语义映射三个阶段。文章以中科院机构名称规范控制库的元数据语义化映射进行实证研究，佐证标准化的元数据语义映射过程对今后特定领域的信息资源共享和交换具有指导意义。

关键词 元数据 语义化 映射 名称规范 机构名称

Research on the Process of Semantic Metadata Mapping ——The Case of Organization Name Control of Chinese Academy of Sciences

LI Hui-jia MA Jian-ling ZHANG Xiu-xiu WANG Nan

Abstract: Semantic metadata mapping is the basis of semantic database or system interoperability. This paper describes the stages of semantic metadata mapping processes, which is based on the semantic metadata description standards and semantic metadata mapping methods. It includes identifying metadata set, grouping data elements and semantic mapping in three stages. This article study on the case of name control system of the Organization of Chinese Academy of Science. So it has guiding significance in specific areas for future information resources sharing and exchange.

Keywords: Metadata; Semantic; Mapping; Name control; Organization name

0 引言

对信息资源质量最重要的评价方法是使用者对该信息资源理解是否充分。在传统意义上，机器可读的元数据是理解数据结构的最重要方式，在过去 15 年中，机器可读元数据是应用模型驱动技术实现数据整合的关键所在。主流的元数据格式定义仅包含一种机器可读元数据，也就是数据的语法元数据，这种元数据并不能帮助数据分析者决定两种不同格式的数据之间的映射关系，而且数据语义化定义的误区也会导致系统为了一些微小的错误而付出昂贵的代价，且不易被解决。另外，由于不同的领域存在多种元数据格式，当用不同元数据格式进行资源描述、检索和利用时，就需要解决多种元数据格式间的释读、转换问题。这些局限是导致目前数据整合工作仍为劳动密集型的关键所在，但机器可读的语义元数据可以适度的减少这些局限性。

1 元数据语义化映射

1.1 语义元数据的内涵

根据元数据的结构特征，语义元数据的内涵可分为弱语义性元数据和强语义性元

数据两类。弱语义性元数据只是给出其元素的语句性定义，并没有指定明确的内容提取规则（如都柏林核心）；强语义性元数据则明确规定内容的提取原则，不仅提供元素的语句性定义，更重要的是规定了一系列生成规则来规定数据元素含义，使得这些元素的语义非常明确和可操作。

1.1.1 语义元数据标准

要实现不同规范的元数据语义映射，就应有针对不同周期和内容的元数据语义描述标准。比如，元数据注册标准 ISO/IEC 11179 Metadata Registry (MDR) standard（简称 ISO/IEC 11179）是用于规范机构元数据的国际标准^[1]，描述了语义元数据映射程序（SMMP），包括元数据集的识别、元数据集分组和元数据语义映射，是一个可以在最大限度上实现元数据集语义互操作的程序。其中，ISO/IEC 11179 -3：注册元模型基本属性是关于元数据语义描述的标准化文档，其对如何实现元数据语义描述的流程做出了详细规定，包括对元数据语义标识、元数据语义内容、元数据语义关系和元数据语义值 4 方面的内容^[2]。

1.1.2 元数据语义化映射方法

语义映射有很多方法,可以分成基于结构的方法 (Structural-Based) 和基于本体的方法 (Ontology-Based) [3]。基于结构的方法重点从数据源的结构信息寻找语义映射关系,解决语法和结构异构对映射的影响;基于本体的方法则力图从数据源的内在语义来寻找语义映射关系[4]。这两种方法各有侧重,能够起到互补的作用。目前较为常用的元数据语义化映射方法有两种,一是基于结构的映射方法,首先对 XML 元素的名称进行分解、扩展等操作,然后对经过分解、扩展处理后的元素名称采用一种称为名称匹配 (Name Match) 的方法进行映射。其重点从分析 XML 的模式信息来寻找映射关系,主要解决 XML 数据语义映射过程中语法和结构上的异构问题,而缺乏对 XML 文档中语义信息的利用[5]。二是基于本体的映射方法,首先在应用需求的基础上建立基于专业领域的共享本体 (Share Ontology),除此之外还要对需要进行映射的 XML 数据源分别建立局部本体 (Local Ontology) 信息,然后根据本体所提供的对象间的关系解决 XML 文档之间的语义映射问题。近年在国际语义网会议的推动下,特别是举办 OAEI (the Ontology Alignment Evaluation Initiative) 竞赛以来,多种通用本体映射系统亮相 OAEI[6],促进了本体映射的发展。

1.2 元数据语义化映射过程

元数据是一个结构化的信息单元,通过对元数据的结构进行解析,可以清楚地得到元数据的层次结构:一个元数据是由内容规则层、元素集合层、载体层和字符编码层 4 个层次组成的。在元数据的层次结构中,每层之间并非相互独立,而是相互关联的,即元数据结构中的上层对下层是一种解释关系。具体而言,内容规则层给出元素的确切的和可操作的定义,是对元素集合层的解释。内容规则层主要是定义元素集合层的规范标准 (如都柏林核心),而每一个元数据规范又给出其元素的定义。在元数据的载体层,元数据表现为一组“实体—属性”对。这些“实体—属性”对元数据载体的予以含义进行界定和解释,因此元素集合层是载体层

的解释。元数据的载体语言将元数据以计算机能够理解的逻辑表达出来,通过字符编码规则转换成计算机代码。所以载体是一种计算机能够理解的语义表达,是字符编码层的解释。元数据载体层通过字符编码赋予了元数据机器可理解的意义。

元数据互操作是将一个元数据集映射到另一个元数据集的最常用方法。但目前介于大部分元数据都是弱语义性元数据,从而元数据互操作通常是基于数据元素之间的简单“一对一”映射关系。为了实现不同平台元数据之间更好的互操作和映射,需要每个元数据互操作时必须给出元数据语义内容,并兼容“一对一”及其他的映射关系的详细阐述。通常两个不同的系统直接不能共享同一个元数据模型。这是因为模型中的元数据类目并不是通过同一种方式分解的,这样的情况必然导致系统间元数据的互操作性受到抑制。例如,对于一个信息对象往往对应有两个或者更多的数据元素集合,如 DC (Dublin Core), MARC (MACHINE READABLE CATALOGUING) 和 MODS (METADATA OBJECT DESCRIPTION SCHEMA) 三种数据元素集合可以用来对书籍信息进行描述。一个信息对象的数据元素的命名是由数据库开发者自行决定的。因此,各数据库间的元数据交换就变得异常困难,甚至不可能实现。根据 ISO/IEC 11179 标准中所提供的框架可实现数据库间元数据的互操作,其中最重要的内容是进行元数据的语义化映射。元数据语义化映射主要分为 3 个阶段:识别元数据、分组数据元素和语义映射[7]。结合 ISO/IEC 11179-1: 框架和 ISO/IEC 11179-3: 注册元模型与基本属性标准对元数据语义化描述进行规范化表达,是实现元数据域语义化映射的前提条件。本文中有关元数据语义化映射的过程就是依据 ISO/IEC 11179-1: 2004 和 ISO/IEC 11179-3: 2003 标准创建公用数据元素概念 (DECs),从而实现不同系统间的元数据的语义化映射。

1.3.1 识别元数据集

识别元数据集的首要任务是将不同数据库中表示同一信息对象的所有有效数据元素集合进行采集,从中识别将被映射的候

选数据元素集合,目的是识别可被映射的数据元素,剔除不适合被映射的数据元素。在进行元数据集识别时,必须调查某一特定领域中需要实现互操作的数据库或系统以及元数据集的情况,主要包括 4 方面:① 调查数据库或系统中采用的元数据标准,国际上流行的元数据标准有 7 种: CDWA、DC、EAD、FGDC、GILS、TEI、VRA^[8]。其中 DC 是针对网络资源的元数据标准,目的是用于资源发现。简单的元素定义和设置便于著录,是 DC 获得广泛应用的重要原因。② 计算元数据元素集合的字段数量,目的是比较不同数据库的元数据完整性和全面性,是确定原始元数据集的重要依据。③ 检查数据库或系统中是否存在样本数据。④ 调查每个元数据集的授权机构。

由于待识别的元数据来源广泛,所采用元数据标准、编写习惯等不同,其元素和属性的名称不统一。因此,为了实现元数据语义化映射的自动化过程,必须通过模型化方式使得机器可以自动识别元数据集,通常的方法是通过构建元数据树匹配模型识别出符合标准的元数据,从而实现元数据的提取、分类、存储或检索^[9]。首先,制定网络资源的元数据标准,选择要识别的数据源;其次,根据元数据标准建立标准化的元数据模板,并将其转换为 XML 结构;第三,对不同数据源的元数据与标准化的元数据模板进行匹配,将数据源的元数据字段名称进行预处理,再进行元数据字段内容和结构的语义化匹配;最后,对标准化的不同数据源

元数据进行自动抽取和识别。

1.3.2 分组数据元素

第二阶段是从识别出来的数据元素集合中对各元素进行分组,主要是包括 4 个连续的过程:发现对象、根据对象对元素进行分组、发现属性、根据属性对元素进行分组。根据对象类对数据元素进行分组,并发现对象的属性,然后属性对数据元素进行再分组。为了方便起见,应在各数据库或系统的元数据集中选择一个原始元数据集,并通过原始元数据集将所有数据元素进行汇总。所谓原始数据元素就是所有源数据集中最简单或者级别最高的元数据集。在候选数据元素集合中所有数据元素都应按其属性进行汇总,其中重要性较小的数据元素,以及不能被分组的数据元素也会被剔除。在数据元素分组的过程中需要元数据专家和领域专家相互配合才能完成。

在没有可选择的最高级别元数据集作为原始数据元素的情况下,可以通过对多种数据源的数据进行汇总分析,构建符合要求的原始数据元素集合。以机构元数据元素为例,为了实现以机构为中心的知识组织体系,则需要构建具有语义关联的元数据容器和元数据元素(见图 1)。其中,元数据容器包括机构名称、机构关系、机构标签、地理信息、知识体系;每个元数据容器中又包含符合该容器定义的元数据元素。对机构元数据元素进行分组的目的是为了结构化机构元数据,从而为构建以机构为中心的知识组织体系奠定基础。

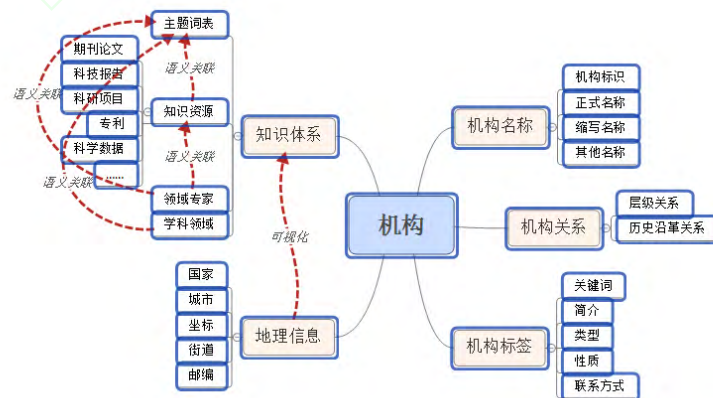


图 1 机构元数据元素分组示意图

1.3.3 语义映射

元数据语义化映射的最后一个阶段就是语义映射的实现。在这个阶段中，需要将所有的数据元素汇总成表，并对每个数据元素的匹配精确度进行注释。在元数据语义化映射过程中将会产生一组被推荐的元数据，并用于指导以后的元数据语义化映射标准。

元数据语义化映射首先要在第二阶段分组的数据元素中寻找公用数据元素概念

(DECs)。若领域本体或分类时已知的将会对构建公用数据元素概念非常有帮助。其次是所有候选数据元素将会按照公用数据元素概念进行分类汇总，分类汇总表中也会对语义元数据异构类型（见表1）进行描述，语义元数据的异构类型由6个类别组成，其中复杂项类别必须通过人工干预的方式才能得到解决，因此在进行元数据语义化映射时需要剔除复杂项类别的。

表1 语义元数据异构类型表

异构类型	子类	标记	例子
	映射关系		
同等级	同等级	—	—
	一对一映射		
分层级	一般	H/gen	Price
	特殊	H/spe	Retail price, Wholesale price
	一对一映射		
	合成	H/com	Name
分解	H/dec	Family name, Given name	
	一对多映射或多对一映射		
领域	领域	D	Summary : Synopsis
	一对一映射（如果需要）		
词法	同义词	L/syn	First name: Given name
	缩写	L/abb	Address: Addr
	首字母缩写	L/acr	Serial Number: SN
	大小写敏感度	L/cas	address: ADDRESS
	语言	L/lan	Name: 名字
	变形	L/var	Color: Colour
一对一映射			
句法	排序	S/ord	Family name: Name (Family)
	分隔符	S/del	Family-name: Family name
	缺失	S/mis	Author name: Author
	一对一映射		
复杂项	复杂项	C	
	不能映射		

2 中科院机构名称元数据语义化映射

2.1 中科院机构名称元数据识别

从不同数据库中对标识机构名称的所有有效元数据元素进行采集，“中科院机构名称规范库”（简称“名称规范库”）在对资源系统进行调研后，选择Web of Science、EI、中国知网（CNKI）和中国科学引文数据库（CSCD）作为机构名称有效源数据采集的来源。之所以选择以上4个数据库作为采集来源，是因为：（1）这些数据库涵盖自

然科学所有的学科领域；（2）这些数据库是在权威性、全面性、完整度、利用率和专业性等方面处于国内外领先地位，得到国内外各类科研机构的广泛认可；（3）经过多年发展，这些数据库的元数据规范不断更新和完善，可采集到较为规范的科研机构相关元数据元素。

本文以Web of Science为例，展现科研机构名称元数据元素集合的字段数量及字段内容（见表2和图2）。

表2 Web of Science 机构相关元数据元素集合

XML 标签名称				子标签名称		字段内容	取值示例
addresses	address_name	addr no	地址排序		1	
			organizations	organization 的值为机构名称缩写		Chinese Academy of Sciences	
				organization pref="Y"的值，机构名称全称			
			suborganizations	包含二级及以下级别机构名称		Xinjiang Inst Ecol & Geog (二级) State Key Lab Desert & Oasis Ecol (三级)	
			city	城市		Urumqi	
			country	国家		Peoples R China	
			zip	邮编		276826	

```

/records/REC/static_data/fullrecord_metadata
1 <fullrecord_metadata>
2   <languages count="1">
3     <language type="primary">English</language>
4   </languages>
5   <normalized_languages count="1">
6     <language type="primary">English</language>
7   </normalized_languages>
8   <normalized_doctypes count="1">
9     <doctype>Article</doctype>
10  </normalized_doctypes>
11  <refs count="55"/>
12  <addresses count="1">
13    <address_name>
14      <address_spec addr_no="1">
15        <full_address>Isfahan Univ Technol, Dept Soil Sci, Coll Agr, Esfahan 8415683111, Iran</full_address>
16        <organizations count="2">
17          <organization>Isfahan Univ Technol</organization>
18          <organization pref="Y">Isfahan University of Technology</organization>
19        </organizations>
20        <suborganizations count="2">
21          <suborganization>Dept Soil Sci</suborganization>
22          <suborganization>Coll Agr</suborganization>
23        </suborganizations>
24        <city>Esfahan</city>
25        <country>Iran</country>
26        <zip location="AC">8415683111</zip>
27      </address_spec>

```

图2 WOS 中机构相关字段 XML 格式

2.2 中科院机构名称元数据元素分组

通过对Web of Science、EI、CNKI和CSCD中机构相关元数据集进行比较,选择Web of Science的元数据集作为原始元数据集;而EI、CNKI和CSCD作为候选元数据集。之所以选择Web of

Science的元数据集作为原始元数据集,是因为在4个数据库中,Web of Science具有最为完整和多层级关系的元数据元素。确定了原始元数据集和候选元数据集后,根据对象的类和属性对元数据元素进行分组(见表3)。

表3 Web of Science 对象的类和属性

对象的类	对象的属性	标签
机构名称	机构名称(全称)	organization pref="Y"
	机构名称(缩写)	organization
	子机构名称(缩写)	suborganization
机构地理位置	国家	country
	城市	city
	邮编	zip

在进行元数据元素分组的时候有几点需要特别注意:(1)部分元数据字段内容需要通过其他方式进行补充,如Web of Science中的子机构名称的全称需要通过渠道采集;(2)部分数据库将多个对象类和属性表现在一个字段中,需要进行字段的切分,如CNKI中,将多个机构名称统一在“作者单位”字段中,将第一作者的机构地理位置统一在“第一作者单位”字段中;(3)机构之间的关系并没有体现在分组的对象类中,需要后期根据字段标签进行添加。

2.3 中科院机构名称元数据语义映射

(1)不同层级映射。一般与特殊关系的映射:一般类—机构名称全称为一对一映射关系,每个机构有且仅有一个正式的机构名称,选定来源于中科院机构网站上的正式名称作为映射关系的一般类;特殊类—机构名称缩写为多对一映射关系,每个机构有来自于不同来源的机构名称缩写,如来源

于WOS、EI、CNKI等数据库或机构网站。比如,中国科学院寒区旱区环境与工程研究所(来源于机构网站)是一般类;Clod & Arid Reg Environm & Engn Res Inst(来源于WOS),CAREERI,CAS(来源于机构网站),Cold And Arid Regions Environmental And Engineering Research Institute, Chinese Academy of Sciences(来源于EI),中科院寒旱所(来源于CNKI)均是特殊类。

合成与分解关系的映射:合成类—主机机构为一对一映射,每个机构有且仅有一个主机机构;分解类—子机构为多对一映射,一个主机机构可包含多个子机构。中科院寒区旱区环境与工程研究所(来源于机构网站)是合成类;中科院沙漠与沙漠化重点实验室(来源于机构网站),中科院寒旱所寒旱区遥感观测系统试验站(来源于CNKI),State Key Laboratory of Frozen Soil Engineering, Cold And Arid Regions Environmental And

Engineering Research Institute, Chinese (来源于EI), Key Lab. of Land Surface Process & Climate Change in Cold & Arid Regions, Cold & Arid Regions Environ. & Eng. Res. Inst. (来源于 WOS)均是分解类。

在合成和分解关系映射过程中需要注意,分解类同时包含一般类和特殊类,需要对每个条目再进行一般类和特殊类的映射。

(2) 学科领域映射。由于4个数据库中学科分类体系标准不同,学科分类体系的层级、数量以及名称都有出入,因此需要寻找

公用数据元素概念(DECs),再讲所有候选数据元素按照公用数据元素概念进行分类汇总。通过对WOS、EI、CNKI和CSCD的学科分类体系和分类标准进行调研和比较,最终确定将WOS的学科分类体系作为公用数据元素概念(DECs)(见表4);然后将来源于其他3个数据库的学科分类体系与WOS的学科分类体系进行映射。

(3) 词法与句法映射。词法映射包括同义词、缩写、首字母缩写、大小写敏感度、语种和变形(见表5)。

表4 WOS学科分类体系(部分)

一级学科领域	二级学科领域	
科学与技术: 自然科学	Astronomy & Astrophysics	Mining & Mineral Processing
	Chemistry	Oceanography
	Crystallography	Optics
	Electrochemistry	Physical Geography
	Geochemistry & Geophysics	Physics
	Geology	Polymer Science
	Mathematics	Thermodynamics
	Meteorology & Atmospheric Sciences	Water Resources
	Mineralogy	
	科学与技术: 生命科学与生物 医学	Agriculture
Allergy		Medical Laboratory Technology
Anatomy & Morphology		Microbiology
Anesthesiology		Mycology
Anthropology		Neurosciences & Neurology
Behavioral Sciences		Nursing
Biochemistry & Molecular Biology		Nutrition & Dietetics
Biodiversity & Conservation		Obstetrics & Gynecology
Biophysics		Oncology
Biotechnology & Applied Microbiology		Ophthalmology
Cardiovascular System & Cardiology		Orthopedics
Cell Biology		Otorhinolaryngology
Critical Care Medicine		Paleontology
Dentistry, Oral Surgery & Medicine		Parasitology
Dermatology		Pathology
Developmental Biology		Pediatrics
Emergency Medicine		Pharmacology & Pharmacy
Endocrinology & Metabolism		Physiology
Entomology		Plant Sciences
Environmental Sciences & Ecology		Psychiatry
Evolutionary Biology		Public, Environmental & Occupational Health
Fisheries		Radiology, Nuclear Medicine & Medical Imaging
Food Science & Technology		Rehabilitation
Forestry		Reproductive Biology
Gastroenterology & Hepatology		Research & Experimental Medicine
General & Internal Medicine		Respiratory System

	Genetics & Heredity	Rheumatology
	Geriatrics & Gerontology	Sport Sciences
	Health Care Sciences & Services	Substance Abuse
	Hematology	Surgery
	Immunology	Toxicology
	Infectious Diseases	Transplantation
	Integrative & Complementary Medicine	Tropical Medicine
	Legal Medicine	Urology & Nephrology
	Marine & Freshwater Biology	Veterinary Sciences
	Mathematical & Computational Biology	Virology
	Medical Ethics	Zoology
科学与技术: 应用科学	Acoustics	Mechanics
	Automation & Control Systems	Metallurgy & Metallurgical Engineering
	Computer Science	Microscopy
	Construction & Building Technology	Nuclear Science & Technology
	Energy & Fuels	Operations Research & Management Science
	Engineering	Remote Sensing
	Imaging Science & Photographic Technology	Robotics
	Information Science & Library Science	Spectroscopy
	Instruments & Instrumentation	Telecommunications
	Materials Science	Transportation
艺术与人文科学	(省略)	
社会科学	(省略)	

表5 词法映射示例

词法	示例
同义词	airplanes: aeroplanes; katathermometers: catathermometers; mineralization: mineralisation
缩写/首字母缩写	Biology/Biologie/Biologia: Biol ; Computer/Computing: Comp ; Physical/Physics/Physician/Physicians: Phys
大小写敏感度	Chinese Academy of Science: chinese academy of science
语种	Chinese Academy of Science: 中国科学院
变形	cite : citing/cites/cited/citation; defense : defense/defence

表6 句法映射示例

句法	示例
排序	Cold And Arid Regions Environmental And Engineering Research Institute, Chinese Academy of Science Chinese Academy of Science, Cold And Arid Regions Environmental And Engineering Research Institute
分隔符	State Key Laboratory of Frozen Soil Engineering, Cold And Arid Regions Environmental And Engineering Research Institute State Key Laboratory of Frozen Soil Engineering; Cold And Arid Regions Environmental And Engineering Research Institute
缺失	State Key Laboratory Of Frozen Soil Engineering, Cold And Arid Regions Environmental And Engineering Research Institute, Chinese State Key Laboratory Of Frozen Soil Engineering, Cold And Arid Regions Environmental And Engineering Research Institute, Chinese Academy of Science (缺失内容)

(4) 复杂项映射。通过4个数据库采集到的机构名称相关元数据元素集中无法对机构的历史变更关系进行映射。由于部分机构的历史变更关系复杂,而且没有统一入口进行批量采集,因此需要通过机构网站的调研进行人工干预。例如,中科院寒区旱区环境与工程研究所先后经历了合并 (Merged from), 拆分 (Decomposed from), 更名 (Changed from) 和前身 (Predecessor)^[10]。

这些复杂的机构历史变更关系需要从机构网站的“机构沿革”“关于我们”等栏目下通过对文字内容的梳理获得。由于不同机

构网站文本描述的风格不同,因此无法建立统一的规则进行机器学习来实现机构历史变更关系的语义化映射,而是需要先通过人工调研,发现通用事例(如大部分机构变更关系的描述都在导航栏“机构沿革”“关于我们”中),然后通过事例学习进行元数据映射。由于机器学习获得的信息往往是不完全的,所以机器学习所进行的推理并不完全是可卡的,必须通过对机器学习的执行效果加以验证,删除或修改不正确的规则,经过反复的学习,不断提高机器学习滤过,最终实现元数据复杂项的映射。

4 结语

即便是具有相同含义的数据元素也有可能具有不同的名称,这就有可能在数据共享或交换的时候产生数据不一致的问题。因此,语义元数据映射的出现可以在数据元素之间进行调解,以实现数据元素的共享或可互操作。元数据互通,即将一个数据元素映射到另一个数据元素上是实现元数据互操作最常用的方法。但是,传统的元数据映射具有差的语义性,这是因为传统的元数据映射仅仅在一对一映射时才有意义。因此,元数据语义化映射是实现元数据语义互操作

的基本方法,语义化映射不仅可以赋予元数据语义层面的意义,而且突破传统元数据映射仅适用于一对一映射的局限性。

本文描述的元数据语义化映射过程及其实现基于语义元数据映射程序(SMMP),该程序是在ISO/IEC 11179标准的指导下形成的,可以最大限度地提高数据元素之间的互操作性。本文结合元数据语义化描述标准和元数据语义化映射方法对元数据语义化映射的过程进行了分阶段描述,从而对今后在特定学科领域的不同数据库或系统间的元数据共享和交换具有指导性意义。

参考文献

- [1] ISO/IEC 11179 [S/OL].[2014-05-10].http://en.wikipedia.org/wiki/ISO/IEC_11179#Structure_of_the_ISO.2FIEC_11179_standard
- [2] ISO/IEC.《ISO/IEC 11179-3:2003 信息技术—元数据注册—第3部分:注册元模型与基本属性》[S].2003-02-15.
- [3] 周武,金远平. XML 数据语义映射方法研究[J]. 计算机工程与应用,2003(28):88-91.
- [4] 周武.XML 数据语义映射研究[D].南京:东南大学,2004.
- [5] 李跃龙. 基于本体的消防知识集成研究[D].大连:大连海事大学,2008.
- [6] 陶俊,孙坦.2009 年以来本体映射系统模型研究综述[J].图书情报工作,2011(12):101-105.
- [7] Sungjoon Lim, Taesul Seo, Changan Lee, Soungsoo Shin.Study on the International Standardization for the Semantic Metadata Mapping Procedure[C].DASFAA 2012, Part II, LNCS 7239,2012:243-249.
- [8] 冯项云,肖琬,廖三三,等. 国外常用源数据标准比较研究[J]. 大学图书馆学报,2001 (4): 15-21.
- [9] 冯秀珍,陈旒. 电子资源元数据的自动识别研究[J]. 情报杂志,2010 (4): 130-133.
- [10] 李慧佳,马建玲,张秀秀,等. 中文机构名称规范库建设的实践与分析——以“中科院机构名称规范库”建设为例[J]. 图书与情报,2016 (1): 133-139.

作者简介 李慧佳,中科院兰州文献情报中心馆员;马建玲,中科院兰州文献情报中心研究馆员;张秀秀,中科院兰州文献情报中心馆员;王楠,中科院兰州文献情报中心副研究馆员。

收稿日期 2017-04-16

(责任编辑:邝玥)